

# ENRICHED FUNCTIONAL LIMIT THEOREMS FOR DYNAMICAL SYSTEMS

ANA CRISTINA MOREIRA FREITAS, JORGE MILHAZES FREITAS, AND MIKE TODD

**ABSTRACT.** We prove functional limit theorems for dynamical systems in the presence of clusters of large values which, when summed and suitably normalised, get collapsed in a jump of the limiting process observed at the same time point. To keep track of the clustering information, which gets lost in the usual Skorohod topologies in the space of càdlàg functions, we introduce a new space which generalises the already more general spaces introduced by Whitt. Our main applications are to hyperbolic and non-uniformly expanding dynamical systems with heavy-tailed observable functions maximised at dynamically linked maximal sets (such as periodic points). We also study limits of extremal processes and record times point processes for observables not necessarily heavy tailed. The applications studied include hyperbolic systems such as Anosov diffeomorphisms, but also non-uniformly expanding maps such as maps with critical points of Benedicks-Carleson type or indifferent fixed points such as Pomeau-Manneville or Liverani-Saussol-Vaienti maps. The main tool is a limit theorem for point processes with decorations derived from a bi-infinite sequence called the transformed anchored tail process.

## CONTENTS

1. Introduction	2
1.1. The setting and $F'$	3
1.2. Point processes, weak mixing conditions and inducing	4
1.3. Organisation	5
2. Enriched functional limit theorems for non-uniformly hyperbolic dynamics	6
2.1. Dynamically defined stochastic processes	6
2.2. Applications to specific systems	7
2.3. The functional spaces	9
2.4. Rare events and point processes	12
2.5. Functional limit theorems for heavy tailed dynamical sums	14
2.6. Enriched extremal process dynamics in the presence of clustering	17
2.7. Record point processes	18
3. General complete convergence of multidimensional cluster point processes	19
3.1. Identifying clusters	20

---

*Date:* January 29, 2024.

*2010 Mathematics Subject Classification.* 37A25, 37A50, 37B20, 60F17, 60G55, 60G70.

*Key words and phrases.* Functional limit theorems, point processes, Lévy processes with decorations, extremal processes, records, clustering, hitting times.

ACMF, JMF were partially financed by Portuguese public funds through FCT – Fundação para a Ciência e a Tecnologia, I.P., in the framework of the projects PTDC/MAT-PUR/28177/2017, PTDC/MAT-PUR/4048/2021, 2022.07167.PTDC and CMUP’s project with reference UIDB/00144/2020. We would like to thank Ian Melbourne for insightful discussions on Skorohod spaces and for various suggestions for improvement of this text.

3.2. Dependence structure	20
3.3. Bookkeeping of clusters	22
3.4. Complete convergence of point processes	31
3.5. Applications to jointly regularly varying sequences	33
4. Proofs of the dynamical enriched functional limit theorems	35
4.1. Proof of Theorem 2.5	36
4.2. Proof of Theorem 2.10	38
4.3. Proof of Theorem 2.12	41
4.4. Point processes encompassing inducing	42
Appendix A. Completeness and separability of the space $F'$	47
Appendix B. Weak and weak <sup>#</sup> convergence	48
Appendix C. Convergence of point processes on non locally compact spaces	50
Appendix D. Computation of the transformed anchored tail processes for the dynamical examples	54
D.1. Applications to observables maximised at hyperbolic periodic points for general 1D systems	54
D.2. Particular application with overshooting	56
References	56

## 1. INTRODUCTION

The Donsker functional limit theorem gives an invariance principle for the sum of independent and identically distributed (i.i.d.) random variables with finite second moments. The functional limit is a Brownian motion which lives in the space  $C$  of continuous  $\mathbb{R}^d$ -valued functions defined on a subinterval of the real line, equipped with the uniform norm. Invariance principles such as this have been proved for a large class of dynamical systems with good mixing properties ([41, 22, 60, 27, 58, 42, 6, 59, 39]). Note that weak convergence to a Brownian motion implies the existence of a Central Limit Theorem (CLT) for the ergodic sums of such systems.

When i.i.d. random variables have infinite second moment, the heavy-tailed case, the sums of the variables can become dominated by a few very large values, leading to failure of the CLT and jumps/discontinuities in the functional limit, as well as a connection with maximal processes and Extreme Value Laws (the Type II/Fréchet case). If the tail of the distribution is sufficiently regular the CLT is replaced by an  $\alpha$ -stable law and Brownian motion is replaced by an  $\alpha$ -stable Lévy process. In the dynamical setting, in [38] Gouëzel showed that the same regime applies to ergodic sums of unbounded heavy tailed observables for the doubling map. Later, in [78], Tyran-Kamińska showed a functional limit theorem for heavy tailed ergodic sums of essentially uniformly expanding maps. The functional limit was an  $\alpha$ -stable Lévy process, which lives in the space  $D$  of càdlàg functions, *i.e.*, right continuous  $\mathbb{R}$ -valued functions with left limits defined, equipped with Skorohod's  $J_1$  topology.

The problem we address in this paper is when there are clusters of large  $\mathbb{R}^d$ -valued observations, so rather than a simple jump, the functional limit would see a sequence of jumps, possibly in different directions: this can be seen when the function is maximised at periodic points, eg [33, 34], or more generally, eg [4, 5, 29, 30, 68]. The  $J_1$  topology is not equipped

to handle convergence here. Indeed, even in quite simple cases, as we note further in an example below, Skorohod's  $M_1$  and  $M_2$  topologies also fail to give convergence. We also mention recent work [16], which gives a very flexible topology, allowing for convergence well beyond the Skorohod topologies, but does not give any information on what may happen in a jump. [9], which considered certain stochastic processes, gives a route to encoding clustering through point processes decorated with a bi-infinite sequence based on the tail process (introduced in [10]) and projection of these to generalised càdlàg functions which keep some of this information. Here, we obtain enriched functional limit theorems for dynamical systems by considering a new space of functions, denoted by  $F'$ , to capture the details of the clustering behaviour, which are recorded in a decoration device called the *transformed anchored tail process*, which is based on the tail process introduced in [10]. We remark that the design of the transformed anchored tail process is quite general encompassing not only the clustering information but also the extremal information observed 'during induced periods', which allows us to apply this theory to non-uniformly hyperbolic systems admitting a nice induced map, such as, the polynomially mixing, Pomeau-Manneville ([70]) or Liverani-Saussol-Vaienti ([55]) maps. We note that in context of this paper, we always assume that we have finite mean cluster size, which is identified by an Extremal Index strictly larger than 0. (We recall that the Extremal Index is a parameter between 0 and 1, which in most cases could be thought of as the reciprocal of the mean cluster size – see [1]). This assumption is vital for the transformed anchored tail process to be well defined.

Extreme Value Laws for heavy-tailed observables with clustering, and laws for ergodic averages of these have significant potential in applications to the theory of climate dynamics (see for example [74, 64, 63, 13]). Moreover, we also believe the space  $F'$  will be a very useful tool to obtain functional limits in other contexts. In fact, in  $F'$ , not only convergence can be proved in situations where, in the jumps appearing in the limit, there is collapsing of smaller jumps, oscillations or overshooting, as the information regarding whatever is happening in a jump is also recorded in the limit.

One possible set of future applications is a related, but different, set of problems in dynamical systems which have received attention recently (see [37, 78, 62, 61, 48, 47]): these are systems with bounded observables and modelled by Young towers ([80, 81]) with *return time functions* that are not square integrable (such as the Pomeau-Manneville or Liverani-Saussol-Vaienti maps or billiards with cusps). These correspond to a situation where the Extremal Index is 0 so that, although the observable is bounded, we get a stacking of numerous observations which add up to create a heavy tailed contribution for the ergodic sums. These problems are partly relevant to this paper since they present similar problems with finding appropriated Skorohod topologies for convergence, but we can not directly apply the theory here since, as noted above, the transformed anchored tail process is not well defined in these situations.

**1.1. The setting and  $F'$ .** We consider cases of vector-valued heavy tailed observables for systems modelled by Young towers, and in the presence of general clusters of extremal observations which get collapsed in the same jump of the limiting Lévy process. The limits that we will consider have discontinuous sample paths and therefore live in the space  $D$  of càdlàg functions. However, this space is not sufficiently rich to keep a record of the fluctuations occurring during the clusters of high observations. The height of the limiting jump accounts for the aggregate effect of all the cluster observations. However, the oscillations observed during the cluster may exceed the height of the jump (an 'overshoot'), for example. In counter to solve this loss of information, in [79], Whitt proposed a new space that he called  $E$ , which decorates each discontinuity of the limit process in  $D$  with an excursion corresponding to a connected set describing the maximum and minimum fluctuations observed. In fact, when  $d = 1$ , the excursion at the discontinuity time  $t^*$  of the limiting process  $V(t)$

is decorated with an interval bounded by the smallest and largest values achieved by the process during the collapsed cluster, which must contain  $V(t^{*-})$  and  $V(t^*)$ , where  $V(t^{*-})$  denotes the lefthand limit of the càdlàg function  $V$  at  $t^*$ .

Nevertheless, most of the information during the cluster is lost and only the maximum oscillations are recorded in  $E$ , while the intermediate fluctuations are completely disregarded. In [79], Whitt also proposed the space  $F$ , which keeps track in particular of the ordering in which points are visited within an excursion in  $E$ . However, the space  $F$  still disregards information because while it keeps track of all the changes of direction during the excursions, it does not keep record of the intermediate jumps observed in the same direction. One of our main goals is to consider a space where no information collapsed into a jump is lost. For that purpose we introduce a new space that we will call  $F'$ . We will endow it with a metric and discuss some of its properties. Then we will use this space to study sums of heavy tailed observables, general extremal processes and records, for which we will obtain enriched functional limit theorems, all of which will be carried out with very minimal loss of information. Another key goal is to build a theory flexible enough to handle a large class of non-uniformly hyperbolic systems modelled by Young towers.

Though we have not yet fully defined our space  $F'$ , here we sketch a very simple example (for more details, see Example 2.7) to show some of the features of our theory. Define  $T : [0, 1] \rightarrow [0, 1]$  by  $T(x) = 3x \bmod 1$  and let  $\mu = \text{Leb}|_{[0,1]}$  be our invariant measure. Set  $\psi(x) = |x - 1/8|^{-2} - |x - 3/8|^{-2}$  and define the one-dimensional stochastic process  $X_0, X_1, \dots$  by  $X_j = \psi \circ T^j(x)$ . Note that  $T(1/8) = 3/8$  and  $T(3/8) = 1/8$  and that  $\psi$  is regularly varying with index  $\alpha = 1/2$  at points  $1/8$  and  $3/8$ . Then the sequence, simulated in Figures 1.1 and 1.2,

$$S_n(t) = \frac{1}{16n^2} \sum_{i=1}^{\lfloor nt \rfloor} \psi \circ T^i$$

converges in our space  $F'$ , in such a way that the pattern of jumps seen when  $x$  is close to  $1/8$  and  $3/8$  are recorded by the limit point in  $F'$ . The way these jumps oscillate creates overshoots in the limit (most easily seen in Figure 1.3 and also in the form of the sum in (2.26)), which cannot be handled by the Skorohod spaces  $J_1, M_1$  or  $M_2$ . While convergence is possible in Whitt's spaces  $E$  and  $F$ , the patterns seen here would not be recorded in  $E$ , and the jump sizes would not be recorded in  $F$  (it would not distinguish between a single jump upwards and a string of multiple jumps upwards). The space  $F'$ , on the other hand, records the patterns: essentially it adds in something like the function seen in Figure 1.2 as a decoration around the corresponding jump.

**1.2. Point processes, weak mixing conditions and inducing.** Point processes have been used successfully to prove functional limit theorems in  $D$  ([72]). Here, since we need to keep control of the information during the excursions, we will use a new type of point processes introduced in [9], where the authors considered point processes in non-locally compact spaces, designed to maintain the ordering of the cluster observations collapsed into the same instant of time, and proved their convergence, under some conditions, for stationary jointly regularly varying sequences. In [9], the authors then applied this convergence of point processes to obtain convergence of sums of jointly regularly varying sequences of random variables in the space  $E$ . In order to prove convergence in  $F'$  we will first generalise their results to more general stationary processes and under weaker conditions, in particular, under a much weaker mixing condition, which is essential to apply to processes arising from dynamical systems. We then push this further using inducing methods, which essentially means that only the induced system needs to satisfy the mixing conditions. In this framework, each sequence for the induced system has an attached sequence which corresponds to what happens in the uninduced case. The point processes then have a  $\mathbb{Z}^2$  component

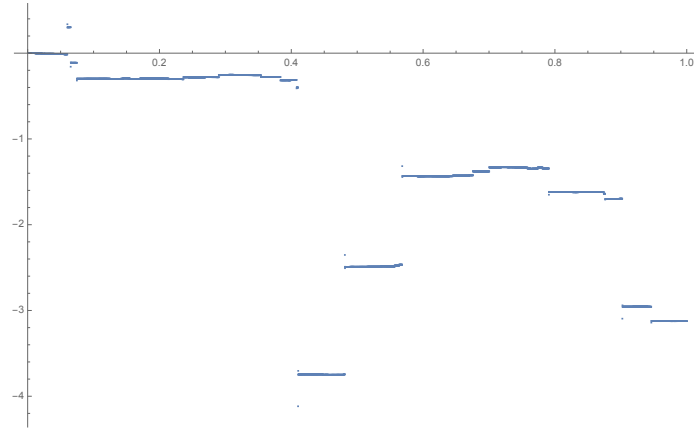


FIGURE 1.1. Plot of a finite sample simulation of  $S_n(t)$  with  $n = 5000$ , where  $X_j = \psi \circ T^j(x)$ , where  $T(x) = 3x \bmod 1$ ,  $\psi(x) = |x - 1/8|^{-2} - |x - 3/8|^{-2}$ .

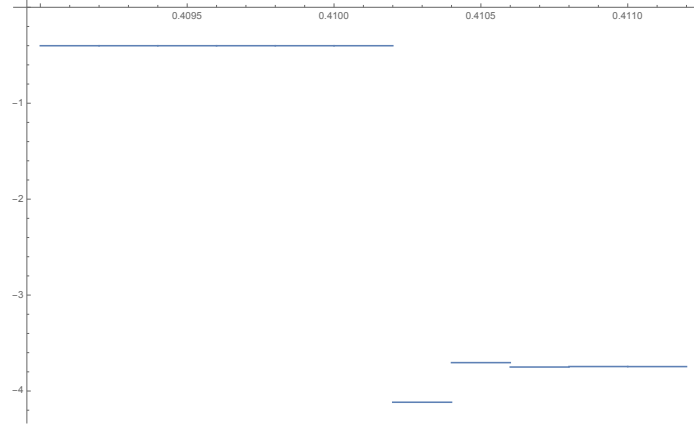


FIGURE 1.2. Blowup of the previous graph at the jump observed near 0.4: asymptotically the four jumps seen here happen instantaneously, necessitating an appropriate space for convergence.

at each  $t^*$ . These can then be projected to  $F'$ , incorporating all the information from the uninduced system into the limit functionals. We observe that the limit point processes in [68], in common with [30], do not record the order of observations in the cluster so our theory is more general than these previous results, even in the simple example given above. These generalisations may have an interest on its own, in the more classical probabilistic setting.

**1.3. Organisation.** This paper develops a large suite of new tools which apply in a wide variety of contexts. We have tried to put our main results and applications as early as practicable in the text in order to motivate the reader and give the theory a more clear context. This sometimes necessitates leaving full explanations and definitions for some of the results to later in the paper, but these are signposted.

In Section 2 we outline the theory from the point of view of dynamical systems, though the theory extends beyond that: we give the relevant observables, focussing on the heavy tailed cases, and briefly describe some of the basic dynamical systems models to which the theory applies. We also define our functional spaces, culminating in the new space  $F'$ , and are then able to state our first limit theorem on convergence to Lévy processes. We then put

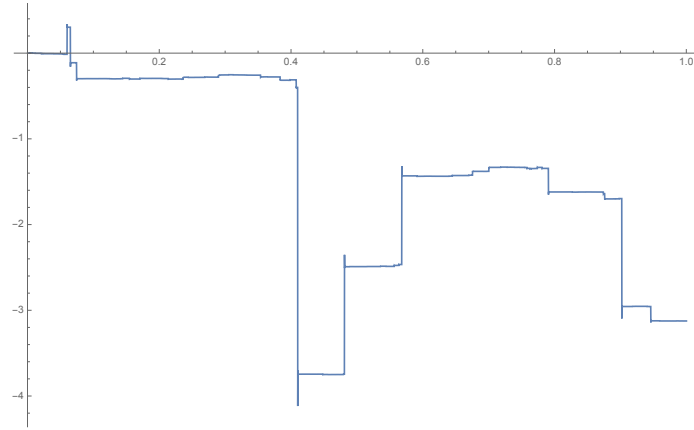


FIGURE 1.3. Completed graph version of the plot in Figure 1.1, to give a different view of the jumps.

the theory into the context of Extreme Value Theory and show convergence to the relevant extremal process. This naturally leads to a theorem on the convergence of record point processes.

In Section 3 we focus on the convergence of point processes for general stochastic processes and rather general observables (not only heavy tailed). Conditions  $\mathcal{A}_{q_n}$  and  $\mathcal{A}'_{q_n}$  are given and the appropriate sequence spaces to record our exceedances are defined. The transformed tail process is then defined which then leads to the definition of the transformed anchored tail process. With all of these conditions met and the transformed anchored tail process existing, we then prove complete convergence of the point process to a Poisson point process. We also show the connection to our setting of jointly regularly varying sequences.

In Section 4 we show how the results from the previous section can be applied in the dynamical context and then prove the functional limit theorems. We conclude by developing the theory for induced systems. In the appendices we cover some of the required background for the results here: Appendix A contains completely new theory for our space  $F'$ ; Appendices B and C give some adaptations of classical theory for sequence spaces and point processes; and Appendix D contains the remaining arguments to show that the theory in this paper applies to the dynamical systems models claimed.

## 2. ENRICHED FUNCTIONAL LIMIT THEOREMS FOR NON-UNIFORMLY HYPERBOLIC DYNAMICS

In this section we start by introducing the setting and in particular the new space  $F'$  and its properties. Then we state the main results regarding the enriched functional limits for sums of vector-valued heavy tailed observables, extremal processes and record point processes. We emphasise that the results in Section 3 hold well beyond the dynamical setting.

**2.1. Dynamically defined stochastic processes.** Let  $(\mathcal{X}, \mathcal{B}_{\mathcal{X}}, \mu, T)$  be a discrete time dynamical system, where  $\mathcal{X}$  is a compact manifold equipped with a norm  $\|\cdot\|$  (for definiteness, whenever it makes sense and unless specified otherwise, we assume that  $\|\cdot\|$  is the usual Euclidean norm),  $\mathcal{B}_{\mathcal{X}}$  is its Borel  $\sigma$ -algebra,  $T : \mathcal{X} \rightarrow \mathcal{X}$  is a measurable map and  $\mu$  is a  $T$ -invariant probability measure, *i.e.*,  $\mu(T^{-1}(B)) = \mu(B)$  for all  $B \in \mathcal{B}_{\mathcal{X}}$ . Let  $\Psi : \mathcal{X} \rightarrow \mathbb{R}^d$  be an observable (measurable) function and define the stochastic process  $\mathbf{X}_0, \mathbf{X}_1, \dots$  given by

$$\mathbf{X}_n = \Psi \circ T^n, \quad \text{for every } n \in \mathbb{N}_0. \quad (2.1)$$

High values of  $\|\Psi(\cdot)\|$  will correspond to entrances in a neighbourhood of a zero measure maximal set  $\mathcal{M}$ , which we express in the following way. Let  $\mathcal{M} \subset \mathcal{X}$  be such that  $\mu(\mathcal{M}) = 0$  and let  $g : [0, \infty) \rightarrow \mathbb{R} \cup \{+\infty\}$  be such that 0 is a global maximum, where we allow  $g(0) = +\infty$ , and  $g$  is a strictly decreasing bijection in a neighbourhood of 0. We assume that, on a neighbourhood of  $\mathcal{M}$ ,

$$\|\Psi(x)\| = g(\text{dist}(x, \mathcal{M})), \quad \text{where } \text{dist}(x, \mathcal{M}) = \inf\{\text{dist}(x, \zeta) : \zeta \in \mathcal{M}\}. \quad (2.2)$$

where  $g$  has one of the three types of behaviour:

Type  $g_1$ : there exists some strictly positive function<sup>1</sup>  $h : W \rightarrow \mathbb{R}$  such that for all  $y \in \mathbb{R}$

$$\lim_{s \rightarrow g_1(0)} \frac{g_1^{-1}(s + yh(s))}{g_1^{-1}(s)} = e^{-y}; \quad (2.3)$$

Type  $g_2$ :  $g_2(0) = +\infty$  and there exists  $\alpha > 0$  such that for all  $y > 0$

$$\lim_{s \rightarrow +\infty} \frac{g_2^{-1}(sy)}{g_2^{-1}(s)} = y^{-\alpha}; \quad (2.4)$$

Type  $g_3$ :  $g_3(0) = D < +\infty$  and there exists  $\gamma > 0$  such that for all  $y > 0$

$$\lim_{s \rightarrow 0} \frac{g_3^{-1}(D - sy)}{g_3^{-1}(D - s)} = y^\gamma. \quad (2.5)$$

Most of the results regarding hitting times and extreme values for dynamical systems were obtained when  $\mathcal{M}$  is reduced to a single point  $\zeta$ . Recent results have considered  $\mathcal{M}$  to be a countable set ([5]), submanifolds ([26, 14]) and fractal sets ([57, 31]). Our results can be applied to general maximal sets  $\mathcal{M}$  and general  $\Psi$ , under the assumption that the transformed anchored tail process is well defined, which will be verified and illustrated for the more common case where  $\mathcal{M} = \{\zeta\}$ , for some hyperbolic point  $\zeta \in \mathcal{X}$ , and where  $\Psi$ , on neighbourhood of  $\zeta$ , can be written as

$$\Psi(x) = g(\text{dist}(x, \zeta)) \frac{\Phi_\zeta^{-1}(x)}{\|\Phi_\zeta^{-1}(x)\|} \mathbb{1}_W(x), \quad (2.6)$$

where  $\Phi_\zeta : V \rightarrow W$  denotes a diffeomorphism, defined on an open ball  $V$  around zero in  $T_\zeta \mathcal{X}$ , the tangent space at  $\zeta$ , onto a neighbourhood  $W$  of  $\zeta$  in  $\mathcal{X}$ , such that  $\Phi_\zeta(E^{s,u} \cap V) = W^{s,u}(\zeta) \cap W$ .

*Remark 2.1.* If  $g$  is of type  $g_2$  (for example,  $g(x) = x^{-1/\alpha}$ ), the measure is sufficiently regular and the geometry of the maximal set  $\mathcal{M}$  is simple (for example,  $\mathcal{M} = \{\zeta\}$ , for some  $\zeta \in \mathcal{X}$ ), then the distribution of  $\mathbf{X}_0$  is regularly varying (see [56, Section 4], for example).

**2.2. Applications to specific systems.** The theory developed in this paper applies to general stochastic processes, but the applications we focus on are dynamical systems. In this subsection we give some preliminary examples of such applications ranging from systems with good exponential mixing properties to some with poor mixing behaviour, leaving further applications to future works. Since some aspects of the proofs of the facts used here require the establishment of some new tools, we postpone these to Section D (see also the discussion following Theorem 4.1).

<sup>1</sup>A possible choice for  $h$  is given in [56, Chapter 4.2.1].

### 2.2.1. Non-uniformly expanding systems.

**Uniformly expanding systems.** We first list some well-behaved dynamical systems which are essentially uniformly expanding, though they need not have Markov, or compactness, properties. More details about the required properties are at the beginning of Section 4.

- Uniformly expanding continuous maps of the interval/circle;
- Markov maps;
- Piecewise expanding maps of the interval with countably many branches such as Rychlik maps;
- Saussol's class of higher-dimensional expanding maps

Here we are always assuming that  $\mu$  is an absolutely continuous (with respect to Lebesgue) invariant probability measure or *acip* since these are a very natural class, though the theory extends beyond these.

Non-trivial examples of observables on these systems to which we can apply the theory include those maximised at a repelling periodic point, first studied in [33], but there is huge scope to study further clustering behaviour such as that shown in [4, 5, 31] for example. We require that the density  $\frac{d\mu}{\text{Leb}}$ , where Leb is Lebesgue, is bounded at the periodic point so that the conformal properties of  $T$  there are reflected in the measure  $\mu$  as well as Leb, but this is automatic in the Rychlik case.

In the heavy-tail applications we will restrict ourselves to  $0 < \alpha < 1$  for simplicity, but observe that techniques to prove (2.21), required in the  $1 \leq \alpha < 2$  case, are provided in [78]: these immediately apply to some of the simpler cases above. Conditions (2.22) and (2.23) are easy computations in the periodic case.

Finally for this introductory discussion on dynamical applications, we note that there is a condition on  $(Q_j)_j$  in Theorem 2.12 which is easily satisfied in all the dynamical examples, see Appendix D.

**Benedicks-Carleson quadratic maps.** Here we provide a class of maps which are far from uniformly expanding, indeed there are critical points.

Here we set  $I = [-1, 1]$  and for  $a \in (0, 2]$  define  $f_a : I \rightarrow I$  by  $f_a(x) = 1 - ax^2$ . This map satisfies the *Benedicks-Carleson* conditions (here  $a$  is close to 2) if there exists  $c, \gamma > 0$  ( $c$  should be close to  $\log 2$  and  $\gamma$  is small) such that

$$|Df_a^n(f(0))| \geq e^{cn} \text{ for } n \in \mathbb{N}_0 \text{ and } |f_a^n(0)| \geq e^{-\gamma\sqrt{n}} \text{ for } n \in \mathbb{N}.$$

It is known that there is a positive Lebesgue measure set  $BC$  of  $a$  such that  $f_a$  satisfies these conditions.

When for definiteness we consider that our observable is maximised at a periodic point of  $f_a$ , we note that the fact that the density at our periodic point is bounded is more delicate than above, but the existence of suitable periodic points for (a positive Lebesgue measure set of) maps in  $BC$  is shown in [34, Section 6].

**Manneville-Pomeau maps.** Our previous examples all have exponential decay of correlations, but our final interval map example shows that this is not necessary. This is what is often referred to as the Liverani-Saussol-Vaienti (LSV) version of the Manneville-Pomeau map: for  $\gamma \in (0, 1)$ , define  $T_\gamma : [0, 1] \rightarrow [0, 1]$  by

$$T_\gamma(x) := \begin{cases} x(1 + 2^\gamma x^\gamma) & \text{if } x \in [0, 1/2), \\ 2x - 1 & \text{if } x \in [1/2, 1]. \end{cases} \quad (2.7)$$

This has an acip  $\mu = \mu_\gamma$  and (sharp) polynomial decay of correlations for Hölder observables against observables in  $L^\infty(\mu)$ , but does not have decay of correlations for observables on some Banach space against all observables in  $L^1(\mu)$ . (We recall that, as seen in [3, Theorem B],



summable decay of correlations against all  $L^1(\mu)$  observables is a strong property which in particular would imply the existence of exponential decay of correlations). To handle this particular case, we introduce in Section 4.4 a new type of point processes encompassing the idea of inducing. This will allow us to obtain the same functional limit theorems in the full range  $\gamma \in (0, 1)$ , for observables of the type  $\varphi: [0, 1] \rightarrow \mathbb{R}$ , where  $\varphi(x) = g(\text{dist}(x, \mathcal{M}))$ ,  $g(x) = x^{-1/\alpha}$  and  $\mathcal{M}$  is finite set of points, possibly periodic. For simplicity, we consider again that  $0 < \alpha < 1$ .

**2.2.2. Invertible hyperbolic systems.** We also consider hyperbolic invertible systems consisting of Anosov linear diffeomorphism,  $T: \mathbb{T}^2 \rightarrow \mathbb{T}^2$ , defined on the flat torus  $\mathbb{T}^2 = \mathbb{R}^2/\mathbb{Z}^2$ . We can associate  $T$  to a  $2 \times 2$  matrix,  $L$ , with integer entries, determinant 1, and without eigenvalues of absolute value 1. As the determinant of  $L$  is equal to 1, the Riemannian structure induces a Lebesgue measure on  $\mathbb{T}^2$  which is invariant by  $T$ . These systems are Bernoulli and have exponential decay of correlations with respect to Hölder observables.

**2.3. The functional spaces.** Let  $D = D([0, 1], \mathbb{R}^d)$  be the space of càdlàg functions defined on  $[0, 1]$ . One can define several metrics in  $D$ . The most usual is the so-called  $J_1$  Skorohod's metric, which generalises the uniform metric by allowing a small deformation of the time scale. In this metric a jump in the limit function must be matched by a similar one in the approximating functions. In order to establish limits with unmatched jumps, Skorohod introduced the  $M_1$  and  $M_2$  topologies which use completed graphs of the functions. We will use a metric motivated by  $M_2$ , which considers the Hausdorff distance between compact sets. We refer to [79] for precise definitions and properties.

In order to keep some of the information collapsed in limit jumps, and to broaden the class of convergent sequences, Whitt introduced the space  $E = E([0, 1], \mathbb{R}^d)$  (see [79, Sections 15.4 and 15.5]), as the space of excursion triples

$$(x, S^x, \{I(s)\}_{s \in S^x}),$$

where  $x \in D$ ,  $S^x$  is a countable set containing the discontinuities of  $x$ , denoted by  $\text{disc}(x)$ , i.e.,  $\text{disc}(x) \subset S^x$ , and, for each  $s \in S^x$ ,  $I(s)$  is a compact connected subset of  $\mathbb{R}^d$  containing at least  $x(s^-)$  and  $x(s)$ . We may identify each element of  $E$  with the set-valued function

$$\hat{x}(t) = \begin{cases} I(t) & \text{if } t \in S^x \\ \{x(t)\} & \text{otherwise} \end{cases}, \quad (2.8)$$

and its graph  $\Gamma_{\hat{x}} = \{(t, z) \in [0, 1] \times \mathbb{R}^d: z \in \hat{x}(t)\}$ . Letting  $p_\ell: \mathbb{R}^d \rightarrow \mathbb{R}$  denote the projection onto the  $\ell$ -th coordinate for  $\ell = 1, \dots, d$ , we define  $\hat{x}^\ell(t) = p_\ell(\hat{x}(t))$  and  $\Gamma_{\hat{x}}^\ell = \{(t, z) \in [0, 1] \times \mathbb{R}: z \in \hat{x}^\ell(t)\}$ .

We embed  $D$  into  $E$ , in the following way. For  $a, b \in \mathbb{R}^d$ , we define the product segment

$$[[a, b]] := [a^1, b^1] \times \dots \times [a^d, b^d],$$

where the one-dimensional segment  $[a^\ell, b^\ell]$  coincides with the interval<sup>2</sup>  $[a^\ell \wedge b^\ell, a^\ell \vee b^\ell]$ . If we have  $a^\ell = b^\ell$ , then  $[a^\ell, b^\ell] = \{a^\ell\} = \{b^\ell\}$ . We identify  $x \in D$  with the element of  $E$

$$(x, \text{disc}(x), \{I(s)\}_{s \in \text{disc}(x)}), \quad \text{where } I(s) = [[x(s^-), x(s)]].$$

We use the Hausdorff metric to define a metric on  $E$ . Namely, recall that for compact sets  $A, B \subset \mathbb{R}^d$ , the Hausdorff distance between  $A$  and  $B$  is given as

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

<sup>2</sup>Recall the notation, used throughout this paper,  $x \wedge y = \min\{x, y\}$  and  $x \vee y = \max\{x, y\}$

For  $A \subset \mathbb{R}^d$ , let  $d(A) = \sup_{x,y \in A} \{\|x - y\|\}$  be the diameter of  $A$ . In order to use the Hausdorff metric, we assume that the elements of  $E$  satisfy the condition:

$$\text{for all } \epsilon > 0, \text{ there exist only finitely many } s \in S^x \text{ such that } d(I(s)) > \epsilon. \quad (2.9)$$

This guarantees that for each element of  $E$ , the associated graph  $\Gamma_{\hat{x}}$  is a compact set. This way, we endow  $E$  with the Hausdorff metric simply by establishing that

$$m_E(\hat{x}, \hat{y}) = \max_{\ell=1, \dots, d} m(\Gamma_{\hat{x}}^\ell, \Gamma_{\hat{y}}^\ell). \quad (2.10)$$

Endowed with this metric  $E$  is separable but not complete. Alternatively, we can define the stronger metric, called uniform metric given by

$$m_E^*(\hat{x}, \hat{y}) = \max_{\ell=1, \dots, d} \sup_{t \in [0,1]} m(\hat{x}^\ell(t), \hat{y}^\ell(t)). \quad (2.11)$$

When endowed with the metric  $m_E^*$ , the space  $E$  is complete but not separable. We refer to [79] for further details about  $E$  and its properties.

The space  $E$  only records the maximal oscillations when information collapses into a jump in the limit. In order to keep a closer track of the fluctuations during the excursions, Whitt introduced the space  $F$ , in [79, Section 15.6], which corresponds to the set of equivalence classes of the set of all the parametric representations of the graphs  $\Gamma_{\hat{x}}$  of elements  $\hat{x}$  of  $E$ , by setting that two parametrisations (continuous functions from  $[0, 1]$  into  $\Gamma_{\hat{x}}$ )  $(r_1, u_1)$  and  $(r_2, u_2)$  are equivalent if there exist continuous nondecreasing onto functions  $\lambda_1, \lambda_2 : [0, 1] \rightarrow [0, 1]$  such that  $(r_1, u_1) \circ \lambda_1 = (r_2, u_2) \circ \lambda_2$ . This means that, in particular, the two functions  $u_1, u_2$  in the two equivalent parametric representations of  $\Gamma_{\hat{x}}$  visit all the points of  $\Gamma_{\hat{x}}$  in each  $I(s)$  the same number of times and in the same order. However, note that  $F$  still misses some fluctuations such as intermediate jumps in the same direction which give rise to a big jump in the limit.

**2.3.1. The new functional space  $F'$  recording all fluctuations.** In order to keep track of all the fluctuations without missing information we introduce the space  $F' = F'([0, 1])$ . We start by considering  $\tilde{D} = \tilde{D}([0, 1], \mathbb{R}^d) = D([0, 1], \mathbb{R}^d)/\sim$  where  $x \sim y$  if there exists a reparametrisation  $\lambda : [0, 1] \rightarrow [0, 1]$ , i.e., a continuous strictly increasing bijection such that  $x \circ \lambda = y$ . Denote the equivalence class of  $x$  by  $[x]$ . We define

$$d_{\tilde{D}}([x], [y]) = \inf_{\lambda} \|x \circ \lambda - y\|,$$

where  $\|\cdot\|$  is the supremum norm and  $\lambda$  is the set of continuous strictly increasing bijections of  $[0, 1]$  to itself (this could be thought of as the induced metric from the  $J_1$  metric on  $\tilde{D}$ ).

We abuse notation within  $\tilde{D}$  by writing  $x$  to refer to both a representative of its equivalence class  $[x]$  and the equivalence class itself.

Now define

$$F' := \{\underline{x} = (x, S^x, \{e_x^s\}_{s \in S^x})\},$$

where  $x \in D([0, 1], \mathbb{R}^d)$ ,  $S^x \subset [0, 1]$  is an at most countable set containing  $\text{disc}(x)$ , the discontinuities of  $x$  and  $e_x^s \in \tilde{D}([0, 1], \mathbb{R}^d)$  is the excursion at  $s \in S^x$ , which is such that  $e_x^s(0) = x(s^-)$  and  $e_x^s(1) = x(s)$ .

We can embed  $D$  into  $F'$ , by associating to  $x \in D([0, 1], \mathbb{R}^d)$  the element

$$\underline{x} = (x, \text{disc}(x), \{e_x^s\}_{s \in \text{disc}(x)}), \quad (2.12)$$

where  $e_x^s \in \tilde{D}([0, 1], \mathbb{R}^d)$  is the equivalence class represented, for example, by

$$e_x^s(t) = x(s^-) + (x(s) - x(s^-)) \mathbb{1}_{[1/2, 1]}(t), \quad t \in [0, 1].$$

We project  $F'$  into Whitt's space  $E$  and into  $\tilde{D}$ , which will give us a metric and a space with more information than Whitt's space  $F$ .

Let  $\pi_E(\underline{x}) = \underline{x}^E := (x, S^x, \{I(s)\}_{s \in S^x})$  where

$$I(s) = \left[ \inf_{t \in [0,1]} e_x^{s,1}(t), \sup_{t \in [0,1]} e_x^{s,1}(t) \right] \times \cdots \times \left[ \inf_{t \in [0,1]} e_x^{s,d}(t), \sup_{t \in [0,1]} e_x^{s,d}(t) \right],$$

with  $e_x^{s,\ell}(t) = p_\ell(e_x^s(t))$ . We project  $F'$  into  $\tilde{D}$  as follows. Suppose that  $S^x$  is countable and write  $S = \{s_i\}_{i=1}^\infty$ , since the finite case follows more straightforwardly. Let  $0 = a_1 < a_2 < \cdots < 1$  be such that  $a_i \rightarrow 1$  as  $i \rightarrow \infty$  (the choice of  $(a_n)_n$  really is arbitrary). We insert the intervals  $[a_i, a_{i+1}]$  at the points  $s \in S^x$ . This is a simple idea, but since  $S$  may be complicated, we need some notation. Define for  $i \in \mathbb{N}$ ,

$$\bar{a}_i := \sum_{s_j \leq s_i} (a_{j+1} - a_j), \quad c_i := s_i + \bar{a}_i - (a_{i+1} - a_i), \quad d_i := s_i + \bar{a}_i, \quad \bar{t} := \sup\{\bar{a}_i : s_i < t\}.$$

Thus  $[c_i, d_i]$  is the interval corresponding to  $s_i$  and  $\bar{t}$  is the accumulated length of the intervals inserted before  $t$ , so we can think of  $t$  being shunted to  $t + \bar{t}$ . Note also that our time line is now of length 2, so we will need to rescale back to a length 1 interval. We then define a representative of the equivalence class  $\tilde{\pi}(\underline{x})$  by

$$\underline{x}^{\tilde{D}}(t) = \begin{cases} x(2t - \bar{t}) & \text{if } 2t \notin \cup_i [c_i, d_i], \\ e_x^{s_i} \left( \frac{2t - c_i}{d_i - c_i} \right) & \text{if } 2t \in [c_i, d_i]. \end{cases} \quad (2.13)$$

Define

$$d_{F'}(\underline{x}, \underline{y}) = d_E(\pi_E(\underline{x}), \pi_E(\underline{y})) + d_{\tilde{D}}(\tilde{\pi}(\underline{x}), \tilde{\pi}(\underline{y})), \quad (2.14)$$

where  $d_E$  denotes  $m_E$  in (2.10). Note that we could have chosen  $m_E^*$  defined in (2.11) rather than  $m_E$  here (see Proposition A.3 and discussions in [79, Section 5.4] and [9, Section 4.1]), but we fix  $m_E$  for definiteness and a more direct comparison with [9].

Note that we could also project into Whitt's  $F$  space here (using  $\pi_E$ , but with  $\{e_x^s\}_{s \in S^x}$  to give the order of the parametrisation in  $I(s)$ ), and that convergence in  $F'$  implies convergence in  $F$ . In  $F'$  we keep the information of the displacement in  $\mathbb{R}^d$  of all (including intermediate) jumps in the discontinuities, while  $F$  only keeps the order and information on the 'local range', *i.e.*, it only captures local extrema. To see this in a 1d example note that the excursions denoted

$$e_x(t) = \begin{cases} 0 & \text{if } 0 \leq t < \frac{1}{3}, \\ \frac{1}{2} & \text{if } \frac{1}{3} \leq t < \frac{2}{3}, \\ 1 & \text{if } \frac{2}{3} \leq t \leq 1, \end{cases} \quad e_y(t) = \begin{cases} 0 & \text{if } 0 \leq t < \frac{1}{3}, \\ 1 & \text{if } \frac{1}{3} \leq t \leq 1, \end{cases}$$

yield the same representations as part of  $F$ : namely the line  $[0, 1]$  with any parametrisation which is an orientation preserving homeomorphism. However, as components of  $F'$  they are distinct with  $d_{\tilde{D}}(e_x, e_y) = \frac{1}{2}$ . Indeed, if  $\underline{x}$  and  $\underline{y}$  differ only by having a discontinuity  $s$  having  $e_x$  and  $e_y$  as the corresponding excursion respectively, then  $d_{F'}(\underline{x}, \underline{y}) = \frac{1}{2}$ .

*Remark 2.2.* The definition of the space and topology can be generalised trivially to other compact time domains such as  $[t_1, t_2]$ , with  $0 \leq t_1 < t_2$ . In order to consider a notion of convergence on non-compact domains such as  $F'((0, \infty), \mathbb{R}^d)$ , we say that  $\underline{x}_n \rightarrow \underline{x}$ , in  $F'((0, \infty), \mathbb{R}^d)$  if the same holds for the respective restrictions to  $F'([t_1, t_2], \mathbb{R}^d)$ , for all  $0 \leq t_1 < t_2$  such that  $t_1, t_2 \notin S^x$ .

We illustrate schematic versions of a sequence of elements in  $F'$  in Figure 2.1 which converge to the element of  $F'$  in Figure 2.2. The jumps stack up to one big jump, but the jumping behaviour is recorded in  $F'$ . Figures 1.1, 1.2 show more complex behaviours in simulations.

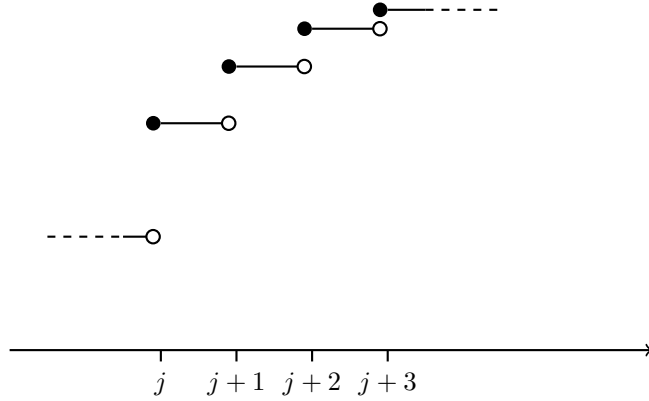


FIGURE 2.1. A piece of  $S_n(t)$  (see (2.20)) in the Gouëzel example [38] where there is a close approach to zero at time  $1 \leq j \leq n-3$ .

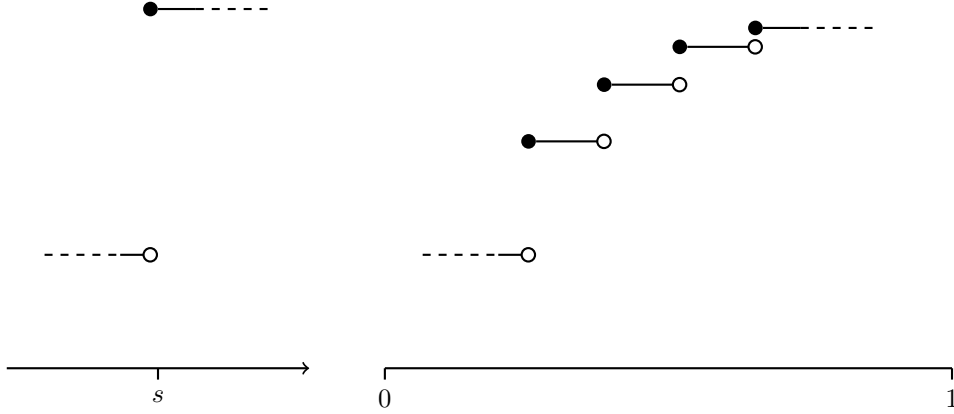


FIGURE 2.2. A sequence of jumps as in the previous figure, after rescaling in time and space, may converge to a part of  $(x, S, \{e^s\}_{s \in S}) \in F'$ . We sketch the  $x$  part on the left and a representative of the corresponding  $e^s$  on the right (this belongs to an equivalence class up to time rescaling). Clearly this convergence cannot take place in  $J_1$ , and, while convergence does hold in  $M_1$  or  $M_2$  (and indeed in  $E$  and  $F$ ), the height of the individual (rather than cumulative) jumps will not be recorded.

We discuss some properties such as completeness and separability of the space  $F'$  in Appendix A.

**2.4. Rare events and point processes.** Since [54, 19], it has been known that the behaviour of the mean of heavy tailed processes is determined by the extremal observations. Hence, as in the context of extremal processes and records, we are lead to the study of rare events corresponding to the occurrence of abnormally large observations. In particular, this means that we need to impose some regularity of the tails of the distributions.

**2.4.1. Normalising threshold functions.** We assume that the stationary sequence of random variables  $\mathbf{X}_0, \mathbf{X}_1, \dots$  has proper tails, in the sense that there exists a normalising sequence of threshold functions  $(u_n)_{n \in \mathbb{N}} : \mathbb{R}^+ \rightarrow \mathbb{R}^+$ , where  $\mathbb{R}^+ = (0, +\infty)$ , satisfying the following properties (see [45]):

- (1) For each  $n$ , the function  $u_n$  is nonincreasing, left continuous and such that

$$\lim_{\tau_1 \rightarrow 0, \tau_2 \rightarrow \infty} \mathbb{P}(u_n(\tau_2) < \|\mathbf{X}_0\| < u_n(\tau_1)) = 1;$$

- (2) For each,  $\tau \in \mathbb{R}^+$ ,

$$\lim_{n \rightarrow \infty} n\mathbb{P}(\|\mathbf{X}_0\| > u_n(\tau)) = \tau. \quad (2.15)$$

We observe that (2.15) requires that the average number of exceedances of  $u_n(\tau)$ , *i.e.*, events of the type  $\|\mathbf{X}_j\| \geq u_n(\tau)$ , for  $j = 0, \dots, n-1$ , is asymptotically constant and equal to  $\tau > 0$ , which can be interpreted as the asymptotic frequency of exceedances. The nonincreasing nature of  $u_n$  reflects the fact that the higher the frequency  $\tau$  of observed exceedances, the lower the corresponding threshold  $u_n(\tau)$  should be.

For every  $z \in \mathbb{R}^+$ , we define

$$u_n^{-1}(z) = \sup\{\tau > 0: z \leq u_n(\tau)\} \quad (2.16)$$

Observe that for each value  $z$  in the range of the r.v.  $\|\mathbf{X}_0\|$ , the function  $u_n^{-1}$  returns the asymptotic frequency  $\tau = u_n^{-1}(z)$  that corresponds to the average number of exceedances of a threshold placed at the value  $z$ , among the  $n$  i.i.d. observations of the r.v.  $\|\mathbf{X}_0\|$ .

Also note that for all  $\tau, s \in \mathbb{R}^+$ ,

$$u_n^{-1}(s) < \tau \quad \text{if and only if} \quad s > u_n(\tau). \quad (2.17)$$

**2.4.2. Point processes of rare events.** Multidimensional point processes are a powerful tool to record information regarding rare events (see for example [69, 72]), which can then be used to study record times ([71, 72]), extremal processes [23, 51, 69, 71, 44], sums of heavy tailed random variables [19, 20, 78, 77, 7, 9]. In particular, they are very useful to keep track of the information within the clusters [65, 45, 20, 66, 7, 9, 30, 68]. More specifically, the Mori-Hsing characterisation tells us that in nice situations the limiting process can essentially be described as having two components: a Poisson process determining the occurrences of clusters and an “orthogonal” point process describing the clustering structure. The description of the clustering component can be accomplished in different ways. In [66] a natural generalisation of a compound Poisson point process is used. In [30], the authors use an outer measure to describe the piling of observations at the base cluster points of the Poisson process, which in the dynamical systems setting can be computed based on the action of the derivative of the map generating the dynamics. In [10], the authors introduced the *tail process*, which is a mechanism to describe the clusters and was subsequently used in [7, 9, 50], for example.

Here, we are going to use a description provided by the *transformed anchored tail process*, which is an adaptation of the tail process introduced in [10] (see Section 3.5 for the relation between the two). Devices of this kind are very well understood in the probabilistic literature, and we refer in particular to a recent work [8] (and references therein) where the authors consider anchored tail processes in a more general framework of processes indexed over integer lattices (in our setting the index is restricted to  $\mathbb{Z}$ ). We will show how the transformed anchored tail process relates with the outer measure of [30] and compute it in the context of dynamical systems.<sup>3</sup>

Although we defer the formal definitions of point processes and point processes of rare events designed to keep all clustering information to Section 3.4 and Appendix C, we give here a brief description of the latter, which again has two components. The first is an underlying

<sup>3</sup>Due to this piling phenomenon, in an earlier preprint version of this paper, we also referred to the anchored tail process as ‘piling process’.

Poisson process on  $\mathbb{R}_0^+ \times \mathbb{R}_0^+$  with intensity measure  $\text{Leb} \times \theta \text{Leb}$ , which can be represented by

$$M = \sum_{i=1}^{\infty} \delta_{(T_i, U_i)}, \quad \text{where } \delta_x \text{ denotes the Dirac measure at } x \in \mathbb{R}_0^+ \times \mathbb{R}_0^+,$$

so that for any measurable disjoint sets  $A, A_1, \dots, A_\ell \subset \mathbb{R}_0^+ \times \mathbb{R}_0^+$ , we have that  $M(A)$  is a Poisson distributed random variable of intensity  $\text{Leb} \times \theta \text{Leb}(A)$  and  $M(A_1), \dots, M(A_\ell)$  are mutually independent. The parameter  $\theta$  is defined formally in Section 3.3.2 and can be thought of as the reciprocal of the average number of exceedances in a cluster. The sequence  $(T_i)_i$  in the first component are points in time and the second component is the angular component of the spectral decomposition of the transformed anchored tail process, which will be defined in Section 3.3.4. For each time  $T_i$  there is a bi-infinite sequence  $(Q_{i,j})_{j \in \mathbb{Z}}$  which will decorate the second coordinate of the mass point of  $M$  at time  $T_i$  and is such that  $\|Q_{i,j}\| \rightarrow \infty$ , as  $|j| \rightarrow \infty$ ,  $\min_{j \in \mathbb{Z}} \|Q_{i,j}\| = 1$ . For a given  $i$ , the sequence  $(Q_j)_{j \in \mathbb{Z}} = (Q_{i,j})_{j \in \mathbb{Z}}$  is given in (3.25). The sequences  $(Q_{i,j})_{i \in \mathbb{N}, j \in \mathbb{Z}}$  are mutually independent and also independent of the sequences  $(T_i)_i$  and  $(U_i)_i$ . The distribution of each transformed anchored tail process, which corresponds to a sequence  $(U_i Q_{i,j})_{j \in \mathbb{Z}}$  at  $T_i$ , is designed to capture the behaviour of the observations within a cluster of exceedances, which was initiated at ('vertical') time  $j = 0$  and whose most severe exceedance has a corresponding asymptotic frequency given by  $U_i$ , in the sense of the interpretation we provided for the  $u_n^{-1}$  function given in (2.16) (recall that the larger the exceedance, the smaller the corresponding asymptotic frequency).

*Remark 2.3.* In order to have some intuition regarding the sequence  $(Q_j)_{j \in \mathbb{Z}}$ , we mention that in the case  $\Psi$  has the form (2.6), with  $\mathcal{M}$  reduced to a repelling fixed point  $\zeta$  where the invariant density is sufficiently regular, then, in the non-invertible case,  $Q_0$  has a uniform distribution on  $\mathbb{S}^{d-1}$ , the unit sphere in  $\mathbb{R}^d$ , and for all  $j \in \mathbb{N}$ ,

$$Q_j = \|(DT_\zeta)^j(Q_0)\|^d \frac{(DT_\zeta)^j(Q_0)}{\|(DT_\zeta)^j(Q_0)\|}, \quad (2.18)$$

where  $DT_\zeta$  denotes the derivative of  $T$  at  $\zeta$ ,  $(DT_\zeta)^j$  its  $j$ -fold product and the norm is just the usual Euclidean norm in  $\mathbb{R}^d$ . For all negative  $j$  we have  $Q_j = \infty$  a.s. (note that here ' $\infty$ ' can be thought of as any point in the completion of  $\mathbb{R}^d$  which is not contained in  $\mathbb{R}^d$ : this has infinite norm).

*Remark 2.4.* We observe that in line with [9] we have placed the  $\theta$  in the second component of the intensity measure  $(\text{Leb} \times \theta \text{Leb})$  of the point process  $M$ . However, as can be seen, for example, from Corollary 3.16, we could have put it in the first coordinate which would be in line with [30, Equation (2.9)] or [50, Remark 7.3.2], for example. One could also express it as  $\theta$  times bidimensional Lebesgue measure as in [45, Corollary 3.7].

**2.5. Functional limit theorems for heavy tailed dynamical sums.** Throughout this section we assume that the process  $\mathbf{X}_0, \mathbf{X}_1, \dots$  is obtained from a system as described in (2.1) and (2.2), where  $g$  is of type  $g_2$ , which together with some regular behaviour of the invariant measure  $\mu$  in the vicinity of the maximal set  $\mathcal{M}$ , guarantees that there exists a sequence of positive real numbers  $(a_n)_{n \in \mathbb{N}}$ , such that

$$\lim_{n \rightarrow \infty} n \mathbb{P}(\|\mathbf{X}_0\| > y a_n) = y^{-\alpha}. \quad (2.19)$$

We refer to [56, Chapter 4] on how (2.19) can be verified and to [33, 40, 43, 34, 15, 4, 31, 14] for several examples of particular dynamical systems and maximal sets satisfying the regularity conditions. Hence, taking  $\tau = x^{-\alpha}$  and  $u_n(\tau) = a_n \tau^{-\frac{1}{\alpha}}$ , equation (2.15) holds.

We are also going to assume that the transformed anchored tail process given in Definition 3.8 exists and is well defined, which implies that existence of the sequences  $(Q_j)_j$  as above. Section 2.2 provides examples of systems satisfying all our requirements.

The Lévy-Itô representation gives a nice way to describe the Lévy process as a functional of Poisson point process, whose intensity measure gives the Lévy measure that determines the process (see [73]). In the case of an  $\alpha$ -stable Lévy process, it is usually identified through a limit of a Poisson integral of a Poisson point process  $M_\alpha = \sum_{i=1}^{\infty} \delta_{(T_i, P_i)}$  with intensity measure  $\text{Leb} \times \nu_\alpha$ , where the Lévy measure  $\nu_\alpha$  is such that  $\nu_\alpha(\{x : \|x\| > y\}) = y^{-\alpha}$ . Namely, when there is no clustering, for example, the limiting Lévy process can be written as

$$V(t) = \lim_{\varepsilon \rightarrow 0} \left( \sum_{T_i \leq t} P_i \mathbb{1}_{\{\|P_i\| > \varepsilon\}} - \int_{\varepsilon < \|x\| \leq 1} x \, d\nu_\alpha(x) \right).$$

Hence, in this case, as explained in more detail in Section 3.5, we consider a transformed version of the general rare events point processes mentioned earlier so that the limit has a Poisson component which can be written as  $M_\alpha = \sum_{i=1}^{\infty} \delta_{(T_i, P_i)} = \sum_{i=1}^{\infty} \delta_{(T_i, U_i^{-1/\alpha})}$ , with intensity measure  $\text{Leb} \times \theta \nu_\alpha$ , with  $\nu_\alpha(y) = d(-y^{-\alpha})$ , while the decorations, which we denote by  $(Q_j)_{j \in \mathbb{Z}}$ , in this case, are related to the  $(Q_j)_{j \in \mathbb{Z}}$  above through (3.41), in Section 3.5.

Consider the partial sum process in  $D([0, 1], \mathbb{R}^d)$  defined by:

$$S_n(t) = \sum_{i=0}^{\lfloor nt \rfloor - 1} \frac{1}{a_n} \mathbf{X}_i - tc_n, \quad t \in [0, 1], \quad (2.20)$$

where the sequence  $(c_n)_{n \in \mathbb{N}}$  is such that  $c_n = 0$  if  $0 < \alpha < 1$  and

$$c_n = \frac{n}{a_n} \mathbb{E}(\mathbf{X}_0 \mathbb{1}_{\|\mathbf{X}_0\| \leq a_n}), \quad \text{for } 1 \leq \alpha < 2.$$

Our main goal in this section is to establish an invariance principle for  $S_n$  which keeps record of all the fluctuations during a cluster of high values which are responsible for a jump of  $S_n$ . As usual, for  $1 \leq \alpha < 2$ , we need that the small contributions for the sum are close to the respective expectation, namely, for all  $\delta > 0$

$$\lim_{\varepsilon \rightarrow 0} \limsup_{n \rightarrow \infty} \mathbb{P} \left( \max_{1 \leq k \leq n} \left\| \sum_{j=1}^k (\mathbf{X}_j \mathbb{1}_{\|\mathbf{X}_j\| \leq \varepsilon a_n}) - \mathbb{E}(\mathbf{X}_j \mathbb{1}_{\|\mathbf{X}_j\| \leq \varepsilon a_n}) \right\| \geq \delta a_n \right) = 0. \quad (2.21)$$

In order to describe the limit, we assume the existence of the transformed anchored tail process, as in Definition 3.8. For  $1 < \alpha < 2$ , we will also need to assume that the sequence  $(Q_j)_{j \in \mathbb{Z}}$ , obtained from the spectral decomposition of the transformed anchored tail process, satisfies the assumption

$$\mathbb{E} \left( \left( \sum_{j \in \mathbb{Z}} \|Q_j\| \right)^\alpha \right) < \infty, \quad (2.22)$$

which in the case  $\alpha = 1$  should be replaced by

$$\mathbb{E} \left( \sum_{j \in \mathbb{Z}} \|Q_j\| \log \left( \|Q_j\|^{-1} \sum_{i \in \mathbb{Z}} \|Q_i\| \right) \right) < \infty. \quad (2.23)$$

In order to describe the excursions during the clusters, we will need an orientation preserving bijection from  $[0, 1]$  to  $[-\infty, \infty]$ , continuous on  $(0, 1)$ . For definiteness, we take:

$$v(t) := \tan \left( \pi \left( t - \frac{1}{2} \right) \right) \quad (2.24)$$

We can now state our main theorem regarding the behaviour of sums of heavy tailed observables.

**Theorem 2.5.** *Let  $T : \mathcal{X} \rightarrow \mathcal{X}$  be a dynamical system as described in Section 2.2. Let  $\mathbf{X}_0, \mathbf{X}_1, \dots$  be obtained from such a system as described in (2.1) and (2.2), where  $g$  is of type  $g_2$  and condition (2.19) holds. Assume also that the transformed anchored tail process given in Definition 3.8 is well defined. Consider the continuous time process  $S_n$  given by (2.20). For  $1 \leq \alpha < 2$  assume further that condition (2.21) holds and, for  $1 < \alpha < 2$ , assume also (2.22), while for  $\alpha = 1$  assume (2.23), instead. Then  $S_n$  converges in  $F'$  to  $\underline{V} := (V, \text{disc}(V), \{e_V^s\}_{s \in \text{disc}(V)})$ , where  $V$  is an  $\alpha$ -stable Lévy process on  $[0, 1]$  which can be written as*

$$V(t) = \sum_{T_i \leq t} \sum_{j \in \mathbb{Z}} U_i^{-\frac{1}{\alpha}} \mathcal{Q}_{i,j}, \quad (2.25)$$

for  $0 < \alpha < 1$  and

$$V(t) = \lim_{\varepsilon \rightarrow 0} \left( \sum_{T_i \leq t} \sum_{j \in \mathbb{Z}} U_i^{-\frac{1}{\alpha}} \mathcal{Q}_{i,j} \mathbb{1}_{\{\|U_i^{-\frac{1}{\alpha}} \mathcal{Q}_{i,j}\| > \varepsilon\}} - t\theta \int_0^{+\infty} \mathbb{E} \left( y \sum_{j \in \mathbb{Z}} \mathcal{Q}_j \mathbb{1}_{\{\varepsilon < y\|\mathcal{Q}_j\| \leq 1\}} \right) d(-y^{-\alpha}) \right)$$

for  $1 \leq \alpha < 2$ ; and the excursions can be represented by

$$e_V^{T_i}(t) = V(T_i^-) + U_i^{-\frac{1}{\alpha}} \sum_{j \leq \lfloor v(t) \rfloor} \mathcal{Q}_{i,j}, \quad t \in [0, 1],$$

where  $(T_i)_{i \in \mathbb{N}}$ ,  $(U_i)_{i \in \mathbb{N}}$  are as described above (see also (3.34)),  $(\mathcal{Q}_{i,j})_{i \in \mathbb{N}, j \in \mathbb{Z}}$  is such that  $\mathcal{Q}_{i,j} = \xi(Q_{i,j})$ , where  $(Q_{i,j})_{i \in \mathbb{N}, j \in \mathbb{Z}}$  are as in (3.25) and  $\xi$  as in (3.37).

*Remark 2.6.* Recall that when  $\Psi$  is as in (2.6) and  $\mathcal{M}$  is reduced to an hyperbolic periodic point, the transformed anchored tail process is well defined and the  $(Q_{i,j})_{i \in \mathbb{N}, j \in \mathbb{Z}}$  are as in Remark 2.3. Also note that when  $\mathcal{M}$  is reduced to a generic point, we have no clustering and then the result holds with the trivial  $Q_j = \infty$ , for all  $j \in \mathbb{Z} \setminus \{0\}$  and  $Q_0$  as before.

**Example 2.7.** We illustrate the theorem with a concrete application which was mentioned in the introduction. We consider the system  $T : [0, 1] \rightarrow [0, 1]$ , given by  $T(x) = 3x \bmod 1$ , which is a uniformly expanding systems as those described in Section 2.2.1. The probability measure  $\mu = \text{Leb}|_{[0,1]}$  is invariant and  $T$  has exponential decay of correlations against  $L^1(\mu)$  (see Definition 4.2, below). We consider the set  $\mathcal{M} = \{1/8, 3/8\}$  which maximises the observable  $\psi(x) = |x - 1/8|^{-2} - |x - 3/8|^{-2}$  and the one-dimensional stochastic process  $X_0, X_1, \dots$  given by  $X_j = \psi \circ T^j(x)$ . Note that  $T(1/8) = 3/8$  and  $T(3/8) = 1/8$ .

We observe that  $\psi$  is regularly varying with index  $\alpha = 1/2$ . In fact, we take  $a_n = 16n^2$ , then

$$\lim_{n \rightarrow \infty} n\mu(\{x \in [0, 1] : |\psi(x)| > a_n u\}) \sim u^{-1/2}.$$

We are now ready to describe the limit of  $S_n(t) = \sum_{i=0}^{\lfloor nt \rfloor - 1} X_i / a_n$ , with  $t \in [0, 1]$  in  $F'$ . For that purpose, let  $\theta = 2/3$  and  $M = \sum_{i=1}^{\infty} \delta_{(T_i, U_i)}$  be a Poisson point process defined on  $\mathbb{R}_0^+ \times \mathbb{R}_0^+$ , with intensity measure  $\text{Leb} \times \theta \text{Leb}$ . Let  $E_1, E_2, \dots$  be a sequence of i.i.d. discrete random variables independent of  $(T_i)_i$  and  $(U_i)_i$  and such that  $\mathbb{P}(E_1 = 1) = 1/2 = \mathbb{P}(E_1 = -1)$ . For each  $i \in \mathbb{N}$ , we set  $Q_{i,j} = (-3)^j E_i$ , for all  $j \in \mathbb{N}_0$  and  $Q_{i,j} = \infty$  for all  $j \in \mathbb{Z} \setminus \mathbb{N}_0$ . We recall that  $\mathcal{Q}_{i,j} = \xi(Q_{i,j}) = (-9)^{-j} E_i$ , for all  $j \in \mathbb{N}_0$  and  $\mathcal{Q}_{i,j} = \xi(Q_{i,j}) = 0$ , for all  $j \in \mathbb{Z} \setminus \mathbb{N}_0$ .



Hence,  $S_n(t)$  converges in  $F'$  to  $(V, \text{disc}(V), e_{V \in \text{disc}(V)}^s)$ , where  $V$  is given in (2.25) and the excursions can be easily written as:

$$e_V^{T_i}(t) = V(T_i^-) + U_i^{-2} E_i \sum_{0 \leq j \leq \lfloor v(t) \rfloor} (-9)^{-j}, \quad t \in [0, 1]. \quad (2.26)$$

Numerical simulations of the finite sample behaviour of  $S_n$ , for large  $n$ , are given in Figures 1.1 and 1.2. The latter blow-up shows how the stacking of jumps can occur in  $S_n$ , and the requirement for convergence in  $F'$  in the limit. As noted in the introduction, here we have overshooting in the discontinuities (namely,  $\inf_{t \in [0, 1]} e_V^s(t) < \min\{V(s^-), V(s)\} < \sup_{t \in [0, 1]} e_V^s(t)$ ), which means that convergence of  $S_n(t)$  in  $D$  would be precluded in any of the Skorohod's topologies.

We leave the proof of the form of the transformed anchored tail process of this example to Appendix D.

**2.6. Enriched extremal process dynamics in the presence of clustering.** Extremal processes are a very useful tool to study the stochastic behaviour of maxima and records (see [72]). We define the partial maxima associated to the sequence  $\mathbf{X}_0, \mathbf{X}_1, \dots$  by

$$M_n := \max\{\|\mathbf{X}_0\|, \dots, \|\mathbf{X}_{n-1}\|\} = \bigvee_{i=0}^{n-1} \|\mathbf{X}_i\|. \quad (2.27)$$

Finding a distributional limit for  $M_n$  is one of the first goals in Extreme Value Theory (see for example [53, 24, 11, 21, 25]).

**Definition 2.8.** We say that we have an *Extreme Value Law* (EVL) for  $M_n$  if there is a non-degenerate d.f.  $H : \mathbb{R}_0^+ \rightarrow [0, 1]$  with  $H(0) = 0$  and, for every  $\tau > 0$ , there exists a sequence of thresholds  $u_n(\tau)$ ,  $n = 1, 2, \dots$ , satisfying equation (2.15) and for which the following holds:

$$\mathbb{P}(M_n \leq u_n(\tau)) \rightarrow \bar{H}(\tau), \quad \text{as } n \rightarrow \infty. \quad (2.28)$$

where  $\bar{H}(\tau) := 1 - H(\tau)$  and the convergence is meant at the continuity points of  $H(\tau)$ .

It turns out that the limit  $H$  allows us to describe the functional limit for associated extremal processes.

In this context, we now consider the continuous time process  $\{Z_n(t) : 0 \leq t \leq 1\}$  defined by

$$Z_n(t) := u_n^{-1}(M_{\lfloor nt \rfloor}), \quad t \in (0, \infty) \quad (2.29)$$

Recall that  $u_n^{-1}(z)$  gives the asymptotic frequency of exceedances of a threshold placed at  $z$  and therefore  $Z_n$  is non-increasing.

For each  $n \geq 1$ ,  $Z_n(t)$  is a random graph with values in  $D((0, \infty), \mathbb{R})$ , which can be embedded into  $F'$ , as in (2.12). The process  $Z_n$  will be shown to converge, in  $F'$ , to the process  $\underline{Z}$ , whose first component in  $D$  is  $Z_H$ , which can be described by the finite-dimensional distributions:

$$\mathbb{P}(Z_H(t_1) \geq y_1, \dots, Z_H(t_k) \geq y_k) = \bar{H}^{t_1} \left( \bigvee_{i=1}^k \{y_i\} \right) \bar{H}^{t_2 - t_1} \left( \bigvee_{i=2}^k \{y_i\} \right) \dots \bar{H}^{t_k - t_{k-1}}(y_k), \quad (2.30)$$

with  $0 \leq t_1 < t_2 < \dots < t_k \leq 1$ . By the Kolmogorov extension theorem such a process is well defined and we call it an *extremal process*, although, strictly speaking, this is a transformed version of the original extremal processes studied by Resnick [72]. The relation between the two is obtained through the connection between the levels  $u_n(\tau)$  and the more

classical linear normalising sequences  $(a_n)_{n \in \mathbb{N}} \subset \mathbb{R}^+$  and  $(b_n)_{n \in \mathbb{N}} \subset \mathbb{R}$  such that we can write  $u_n = y/a_n + b_n$  and

$$n\mathbb{P}(X_0 > u_n) = n\mathbb{P}(a_n(X_0 - b_n) > y) \rightarrow \tau$$

with  $\tau = f(y)$  for some homeomorphism  $f$ , then, as in (2.27),

$$\mathbb{P}(M_n \leq u_n) = \mathbb{P}(a_n(M_n - b_n) \leq y) \rightarrow G(y)$$

where  $G = \bar{H} \circ f$ . Then, if  $Y_n(t) = a_n(M_{\lfloor nt \rfloor + 1} - b_n)$  and  $Y_G$  denotes the respective extremal process obtained in [72], we have that  $Z_n(t) = f(Y_n(t))$  and  $Z_H(t) = f(Y_G(t))$ .

*Remark 2.9.* Depending on the type of limit law that applies,  $f(y)$  is of one of the following three types:  $f_1(y) = e^{-y}$  for  $y \in \mathbb{R}$ ,  $f_2(y) = y^{-\alpha}$  for  $y > 0$ , and  $f_3(y) = (-y)^\alpha$  for  $y \leq 0$ .

**Theorem 2.10.** *Let  $T : \mathcal{X} \rightarrow \mathcal{X}$  be a dynamical system as described in Section 2.2. Let  $\mathbf{X}_0, \mathbf{X}_1, \dots$  be obtained from such a system as described in (2.1) and assume that the transformed anchored tail process given in Definition 3.8 is well defined. Consider the continuous time process  $Z_n$  defined by (2.29). Then  $Z_n$  converges in  $F'$  to  $(Z_H, \text{disc}(Z_H), \{e_{Z_H}^s\}_{s \in \text{disc}(Z_H)})$ , where  $Z_H$  is defined as in (2.30), with  $\bar{H}(\tau) = e^{-\theta\tau}$ , and the excursions can be represented by*

$$e_{Z_H}^s(t) = \min \left\{ Z_H(s^-), \inf_{j \leq \lfloor v(t) \rfloor} Z_H(s) \cdot Q_j^s \right\}, \quad t \in [0, 1]$$

where each sequence  $(Q_j^s)_{j \in \mathbb{N}}$  is independent of  $Z_H(s^-)$  and with common distribution given by (3.25). Moreover,  $Z_H$  can be seen as a Markov jump process with

$$\mathbb{P}(Z_H(t+s) \geq y \mid Z_H(s) = z) = \begin{cases} e^{-\theta ty} & \text{if } y < z \\ 0 & \text{if } y \geq z \end{cases}, \quad \text{for } t, s > 0.$$

The parameter of the exponential holding time in state  $z$  is  $\theta z$  and given that a jump is due to occur the process jumps from  $z$  to  $[0, y)$  with probability

$$\Pi(z, [0, y)) = \begin{cases} \frac{y}{z} & \text{if } y < z \\ 1 & \text{if } y \geq z \end{cases}.$$

*Remark 2.11.* Alternatively and similarly to extremal processes that can be described as a projection into  $D$  of a point process (see [72, Equation (4.20)]), we can describe the limit of  $Z_n$  as the projection into  $F'$  of the point process  $N = \sum_{i=1}^{\infty} \delta_{(T_i, U_i \bar{\mathbf{Q}}_i)}$ , given in (3.34). Namely,  $Z_H(t) = \inf\{U_i : T_i \leq t\}$ ,  $\text{disc}(Z_H) = \{T_i : i \in \mathbb{N}\}$  and

$$e_{Z_H}^{T_i}(t) = \min \left\{ \inf\{U_j : T_j < T_i\}, \inf_{j \leq \lfloor v(t) \rfloor} U_i \cdot Q_{i,j} \right\}, \quad t \in [0, 1].$$

**2.7. Record point processes.** The study of record times of observational data has important applications in the study of natural phenomena. Consider the original sequences  $(\mathbf{X}_n)_{n \in \mathbb{N}_0}$  and  $(M_n)_{n \in \mathbb{N}}$  given in (2.27), and let  $t_1 = 0$ . Define the strictly increasing sequence  $(t_k)_{k \in \mathbb{N}}$ :

$$t_k := \inf\{j > t_{k-1} : \|\mathbf{X}_j\| > M_j\}. \quad (2.31)$$

This sequence  $(t_k)_k$  corresponds to the *record times* associated to  $M_n$ , namely the times where  $M_n$  jumps. In the presence of clustering, record times may collapse in the limit and it is important to keep track of possible increments on the number of records occurring during the clusters. Two possible approaches are to consider the enriched limits in  $F'$  of the extremal processes  $Z_H$  or simply to project directly from the point processes  $N$ , as done in [9]. Since following the former approach, to be able to handle the possibility of having vanishing jump points in the limit we would need to consider more restrictive subspaces to obtain continuity and then apply the Continuous Mapping Theorem (CMT), we will use

the latter approach so that we can benefit from the work already carried in [9] and reduce the length of the exposition.

We follow [9, Section 5] closely, although we make some adjustments, in particular, due to the fact that we are using a transformed version of processes associated to the asymptotic frequencies given by the normalisation by  $u_n^{-1}$ . The main advantage here is that we obtain the convergence of the record times point processes for stationary vector-valued sequences with much more general distributions rather than regularly varying sequences (or sequences that could be monotonically transformed into regularly varying ones) as in [9]. The notation and notions of convergence for point processes used here are detailed in Appendix C.

In order to count the number of records of the process  $\mathbf{X}_0, \mathbf{X}_1, \dots$ , we introduce the record point process

$$\mathfrak{R}_n = \sum_{i=0}^{\infty} \delta_{\frac{i}{n}} \mathbb{1}_{\{\|\mathbf{X}_i\| > M_i\}}. \quad (2.32)$$

**Theorem 2.12.** *Let  $T : \mathcal{X} \rightarrow \mathcal{X}$  be a dynamical system as described in Section 2.2. Let  $\mathbf{X}_0, \mathbf{X}_1, \dots$  be obtained from such a system as described in (2.1) and assume that the transformed anchored tail process given in Definition 3.8 is well defined and that*

$$\mathbb{P}(\text{all finite } Q_j \text{'s are mutually different}) = 1.$$

*Then  $\mathfrak{R}_n$  converges weakly to  $\mathfrak{R}$ , in  $\mathcal{N}_{(0,+\infty)}^\#$ . The limiting process is a compound Poisson process which can be represented as*

$$\mathfrak{R} = \sum_{i \in \mathbb{Z}} \delta_{\tau_i} \kappa_i, \quad (2.33)$$

*where  $\sum_{i \in \mathbb{Z}} \delta_{\tau_i}$  is a Poisson point process on  $(0, \infty)$  with intensity measure  $x^{-1} dx$ . Here  $(\kappa_i)_i$  is a sequence of i.i.d. random variables independent of  $(\tau_i)_i$  with distribution corresponding to the number of record lower values observed in the sequence  $\mathbf{Q} = (Q_j)_j$ , which beat (dropped below) the threshold  $U^{-1}$ , where  $U$  is a uniformly distributed random variable independent of  $\mathbf{Q}$ .*

*Remark 2.13.* In the periodic point case, the condition regarding the fact that all finite  $Q_j$ 's are a.s. mutually different is trivially satisfied.

### 3. GENERAL COMPLETE CONVERGENCE OF MULTIDIMENSIONAL CLUSTER POINT PROCESSES

In this section we present our technical tools and various results in the context of general stochastic processes: the shift  $\sigma$  marries this with a dynamical point of view, but the work here holds in wide generality.

Let  $\mathcal{V} = \mathbb{R}^d$ , for some  $d \in \mathbb{N}$ , where we consider a norm which we denote by  $\|\cdot\|$ . For definiteness, we may consider the usual Euclidean norm. We will be considering the spaces of one-sided and two-sided  $\mathcal{V}$ -valued sequences, which we will denote, respectively, by  $\mathcal{V}^{\mathbb{N}_0}$  and  $\mathcal{V}^{\mathbb{Z}}$ , where we consider the one-sided and two-sided shift operators defined by  $\sigma : \mathcal{V}^{\mathbb{N}_0, \mathbb{Z}} \rightarrow \mathcal{V}^{\mathbb{N}_0, \mathbb{Z}}$ , where

$$\sigma((x)_i) = ((x)_{i+1}). \quad (3.1)$$

Consider a stationary sequence of random vectors  $\mathbf{X}_0, \mathbf{X}_1, \dots$ , taking values on  $\mathcal{V} = \mathbb{R}^d$ , which we will identify with the respective coordinate-variable process on  $(\mathcal{V}^{\mathbb{N}_0}, \mathcal{B}^{\mathbb{N}_0}, \mathbb{P})$ , given by Kolmogorov's existence theorem, where  $\mathcal{B}^{\mathbb{N}_0}$  is the  $\sigma$ -field generated by the coordinate functions  $Z_n : \mathcal{V}^{\mathbb{N}_0} \rightarrow \mathcal{V}$ , with  $Z_n(x_0, x_1, \dots) = x_n$ , for  $n \in \mathbb{N}_0$ . Note that, under these identifications, we can write:

$$Z_{i-1} \circ \sigma = Z_i, \quad \text{for all } i \in \mathbb{N}.$$

Since, we assume that the process is stationary, then  $\mathbb{P}$  is  $\sigma$ -invariant. Note that  $Z_i = Z_0 \circ \sigma^i$ , for all  $i \in \mathbb{N}_0$ , where  $\sigma^i$  denotes the  $i$ -fold composition of  $\sigma$ , with the convention that  $\sigma^0$  denotes the identity map on  $\mathcal{V}^{\mathbb{N}_0, \mathbb{Z}}$ .

In what follows, for every  $A \in \mathcal{B}^{\mathbb{N}_0, \mathbb{Z}}$ , we denote the complement of  $A$  as  $A^c := \mathcal{V}^{\mathbb{N}_0, \mathbb{Z}} \setminus A$ .

**3.1. Identifying clusters.** Our goal is to study the impact of clustering on the convergence of general multidimensional point processes. As mentioned earlier, information regarding the observations within the same cluster gets collapsed at the same time point, which makes it hard to recover it from the limiting process. In order to keep track of that information, we need to start by identifying clusters.

There are two main approaches to identify clusters, which are commonly referred to as declustering procedures. One is the blocking method and the other is the runs declustering procedure.

We start with the blocking procedure, which, in fact, serves two purposes. Namely, not only will it separate clusters, it will also introduce time gaps between the blocks in order to restore some independence between them. The size of the blocks must be sensitively tuned so that the blocks are neither too long so that they do not separate clusters, nor too small so that they do not split clusters apart. The same applies to time gaps between the blocks, which are created by disregarding observations. Following the classical scheme [53], considering a finite sample of size  $n$ , we split the data into  $k_n \in \mathbb{N}$  blocks of size  $r_n := \lfloor n/k_n \rfloor$  and take time gaps of size  $t_n \in \mathbb{N}$ . This way, we define sequences  $(k_n)_{n \in \mathbb{N}}$ ,  $(r_n)_{n \in \mathbb{N}}$ ,  $(t_n)_{n \in \mathbb{N}}$ , which we assume to be such that

$$k_n, r_n, t_n \xrightarrow{n \rightarrow \infty} \infty \quad \text{and} \quad k_n t_n = o(n). \quad (3.2)$$

Regarding the runs declustering, for a finite sample of size  $n$ , we set the run length  $q_n \in \mathbb{N}$  with the aim that all abnormal observations occurring within a time difference of at most  $q_n$  units between each other belong to the same cluster. The sequence  $(q_n)_{n \in \mathbb{N}}$  must be chosen so that

$$q_n = o(r_n), \quad (3.3)$$

and also so that it satisfies conditions  $\mathcal{D}_{q_n}$ ,  $\mathcal{D}'_{q_n}$ , below. Note that  $q_n = q$  for all  $n \in \mathbb{N}$  and some  $q \in \mathbb{N}$  is a possibility here. Namely, when applying to periodic points  $q_n$  can be taken as  $q$ , the period of the point (see [33] and [1, Section 2.1]).

To summarise, we will define objects motivated by a runs declustering scheme (see (3.5) below), the dependence conditions that we introduce stem from a blocking procedure, which will eventually determine the identification of clusters when we introduce the point processes of clusters in Section 3.4. The connection between the two approaches is essentially provided by condition  $\mathcal{D}'_{q_n}$ . See also Remarks 3.1 and 3.2.

**3.2. Dependence structure.** In order to prove the main convergence results we need to introduce some conditions on the dependence structure of the stationary processes and therefore introduce the following objects. We follow more or less the notation used in [29, 30].

Let  $A \in \mathcal{B}^{\mathbb{N}_0, \mathbb{Z}}$  be an event, let  $J$  be an interval contained in  $[0, \infty)$ . We define

$$\mathcal{W}_J(A) := \bigcap_{i \in J \cap \mathbb{N}_0} \sigma^{-i}(A^c), \quad \mathcal{W}_J^c(A) := (\mathcal{W}_J(A))^c = \bigcup_{i \in J \cap \mathbb{N}_0} \sigma^{-i}(A). \quad (3.4)$$

Let  $i, j \in \mathbb{N}$  and set  $\mathcal{W}_j(A) := \mathcal{W}_{[0, j)}(A)$ ,  $\mathcal{W}_{i, j}(A) := \mathcal{W}_{[i, j)}(A)$ , when  $i \leq j - 1$ ,  $\mathcal{W}_{i, j}(A) := \mathcal{V}^{\mathbb{N}_0, \mathbb{Z}}$ , when  $i \geq j$ . Moreover,  $\mathcal{W}_j^c(A) = (\mathcal{W}_j(A))^c$  and  $\mathcal{W}_{i, j}^c(A) = (\mathcal{W}_{i, j}(A))^c$ .

For the event  $A \in \mathcal{B}^{\mathbb{N}_0, \mathbb{Z}}$  and  $j \in \mathbb{N}$ ,

$$A^{(j)} := A \cap \sigma^{-1}(A^c) \cap \dots \cap \sigma^{-j}(A^c), \quad (3.5)$$

and for  $j = 0$  we simply define  $A^{(0)} := A$ .

In what follows, for some  $a > 0$ ,  $y \in \mathbb{R}^d$  and a set  $A \subset \mathbb{R}^d$ , we set  $aA = \{ax : x \in A\}$  and  $A + y = \{x + y : x \in A\}$ . We define the class of sets

$$\mathcal{F} = \left\{ \{(x_j)_j \in \mathcal{V}^{\mathbb{N}_0, \mathbb{Z}} : x_j \in H_j, j = 0, \dots, m\} : H_j \in \mathcal{F}_{\mathcal{V}}, j = 0, \dots, m, m \in \mathbb{N} \right\}, \quad (3.6)$$

where  $\mathcal{F}_{\mathcal{V}}$  denotes the field generated by the rectangles of  $\mathcal{V}$  of the form  $[e_1, f_1) \times \dots \times [e_d, f_d)$ . Note that  $\mathcal{F}$  is a field.

For each  $\ell = 1, \dots, m \in \mathbb{N}$  suppose

$$A_\ell \in \mathcal{F} \text{ and } J_\ell = [a_\ell, b_\ell), \quad (3.7)$$

where  $0 \leq a_1 < b_1 \leq a_2 < b_2 \leq \dots \leq a_m < b_m \leq 1$ . Then for each  $n \in \mathbb{N}$  define

$$\tilde{J}_{k_n, \ell} := k_n J_\ell, \quad J_{n, \ell} := [(\lceil k_n a_\ell \rceil - 1)r_n, (\lfloor k_n b_\ell \rfloor + 1)r_n), \quad (3.8)$$

$$A_{n, \ell} = \left\{ \left( u_n^{-1}(\|\mathbf{X}_j\|) \frac{\mathbf{X}_j}{\|\mathbf{X}_j\|} \right)_j \in \mathcal{V}^{\mathbb{N}_0} : \left( u_n^{-1}(\|\mathbf{X}_j\|) \frac{\mathbf{X}_j}{\|\mathbf{X}_j\|} \right)_j \in A_\ell \right\}, \quad (3.9)$$

where  $u_n$  is defined as in (2.15),  $k_n$  is as in (3.2). We discuss this normalisation further in (3.13) and (3.14).

We introduce a mixing condition which is specially designed for the application to the dynamical setting.

**Condition** ( $\mathcal{D}_{q_n}$ ). We say that  $\mathcal{D}_{q_n}$  holds for the sequence  $\mathbf{X}_0, \mathbf{X}_1, \dots$  if there exist sequences  $(k_n)_{n \in \mathbb{N}}$ ,  $(r_n)_{n \in \mathbb{N}}$ ,  $(t_n)_{n \in \mathbb{N}}$  and  $(q_n)_{n \in \mathbb{N}}$  satisfying (3.2) and (3.3), such that for every  $m, t, n \in \mathbb{N}$  and every  $J_\ell$  and  $A_\ell$ , with  $\ell = 1, \dots, m$ , chosen as in (3.7), we have

$$\left| \mathbb{P} \left( A_{n, \ell}^{(q_n)} \cap \bigcap_{i=\ell}^m \mathcal{W}_{J_{n, i}} \left( A_{n, i}^{(q_n)} \right) \right) - \mathbb{P} \left( A_{n, \ell}^{(q_n)} \right) \mathbb{P} \left( \bigcap_{i=\ell}^m \mathcal{W}_{J_{n, i}} \left( A_{n, i}^{(q_n)} \right) \right) \right| \leq \gamma(n, t), \quad (3.10)$$

where  $\min\{J_{n, \ell} \cap \mathbb{N}_0\} \geq t$  and  $\gamma(n, t)$  is decreasing in  $t$  for each  $n$  and  $\lim_{n \rightarrow \infty} n\gamma(n, t_n) = 0$ , where  $A_{n, \ell}^{(q_n)}$  is given by (3.9) and (3.5).

This mixing condition is much milder than similar conditions used in the literature and is particularly suited for applications to dynamical systems, since it is easily verified for systems with sufficiently fast decay of correlations, see for example the discussion in Section 2.2.

**Condition** ( $\mathcal{D}'_{q_n}$ ). We say that  $\mathcal{D}'_{q_n}$  holds for the sequence  $\mathbf{X}_0, \mathbf{X}_1, \mathbf{X}_2, \dots$  if there exist sequences  $(k_n)_{n \in \mathbb{N}}$ ,  $(r_n)_{n \in \mathbb{N}}$  and  $(q_n)_{n \in \mathbb{N}}$  satisfying (3.2) and (3.3), such that for every  $A_1 \in \mathcal{F}$ , we have

$$\lim_{n \rightarrow \infty} n \mathbb{P} \left( A_{n, 1}^{(q_n)} \cap \mathcal{W}_{[q_n+1, r_n)}^c(A_{n, 1}) \right) = 0,$$

where  $A_{n, 1}^{(q_n)}$  is given by (3.9) and (3.5).

*Remark 3.1.* Condition  $\mathcal{D}'_{q_n}$  forbids the appearance of new abnormal observations (the occurrence of  $A_{n, \ell}$ ), within the same block, once a run of  $q_n$  consecutive non-abnormal observations has been realised. This means that  $\mathcal{D}'_{q_n}$  establishes a connection between the two declustering procedures and, in particular, requires that no more than one cluster should occur within one block.

*Remark 3.2.* Suppose  $(q_n)_{n \in \mathbb{N}}$  and  $(\tilde{q}_n)_{n \in \mathbb{N}}$  are two sequences satisfying (3.3) and  $q_n \leq \tilde{q}_n$ , for all  $n \in \mathbb{N}$ . Then  $\tilde{\mathcal{A}}'_{q_n}$  implies  $\tilde{\mathcal{A}}'_{\tilde{q}_n}$ . As in Section 2.2, for nice dynamical systems when clustering is created by a periodic point of prime period  $q$ , then condition  $\mathcal{A}'_q$  holds, in the sense that  $q_n = q$ , for all  $n$ .

Next we state a stronger version of  $\mathcal{A}'_{q_n}$ , which, in some cases may be easier to check. We set

$$U(\tau) = \left\{ (x_j)_j \in \mathcal{V}^{\mathbb{N}_0, \mathbb{Z}} : x_0 \in B(0, \tau) \right\} \quad (3.11)$$

where, for  $x \in \mathcal{V}$  and  $\epsilon > 0$ , we denote by  $B(x, \epsilon)$ , the ball centred at  $x$  of radius  $\epsilon$ . Then, following (3.9) and (3.5), we define:

$$U_n(\tau) = \{ \|\mathbf{X}_0\| > u_n(\tau) \} \\ U_n^{(q_n)}(\tau) = \{ \|\mathbf{X}_0\| > u_n(\tau), \|\mathbf{X}_1\| \leq u_n(\tau), \dots, \|\mathbf{X}_{q_n}\| \leq u_n(\tau) \}.$$

**Condition  $(\tilde{\mathcal{A}}'_{q_n})$ .** We say that  $\tilde{\mathcal{A}}'_{q_n}$  holds for the sequence  $\mathbf{X}_0, \mathbf{X}_1, \mathbf{X}_2, \dots$  if there exist sequences  $(k_n)_{n \in \mathbb{N}}$ ,  $(r_n)_{n \in \mathbb{N}}$ ,  $(t_n)_{n \in \mathbb{N}}$  and  $(q_n)_{n \in \mathbb{N}}$  satisfying (3.2) and (3.3), such that for every  $\tau > 0$ , we have

$$\lim_{n \rightarrow \infty} n\mathbb{P} \left( U_n(\tau) \cap \mathcal{W}_{[q_n+1, r_n]}^c(U_n(\tau)) \right) = 0. \quad (3.12)$$

For  $A_\ell$ , as in (3.7), let  $\tau^* > \sup\{\|x\| : x \in H_0\}$ . Then, by (2.17), we have

$$\left\| u_n^{-1}(\|\mathbf{X}_0\|) \frac{\mathbf{X}_0}{\|\mathbf{X}_0\|} \right\| < \tau^* \Leftrightarrow \|\mathbf{X}_0\| > u_n(\tau^*).$$

Therefore, it is clear that  $A_{n,\ell} \subset U_n(\tau^*)$ , so if  $\tilde{\mathcal{A}}'_{q_n}$  holds then so does  $\mathcal{A}'_{q_n}$ .

*Remark 3.3.* Condition  $\tilde{\mathcal{A}}'_{q_n}$  is already weaker than [9, Assumption 1.1], which had been used in previous papers (see, for example, [20, Equation (2.8)], [75, Equation (3)], [10, Condition 4.1], [7, Condition 2.1]) and was introduced in [76]. We also remark that all these conditions allow for the appearance of clustering which already makes them weaker than conditions  $D'$  from [19] or  $\text{LD}(\phi_0)$  from [77], which imply  $\mathcal{A}'_{q_n}$  with  $q_n = 1$  for all  $n \in \mathbb{N}$ .

**3.3. Bookkeeping of clusters.** In this section, we introduce a device called the *transformed anchored tail process*, which is designed to keep track of the clustering oscillations. It is an adaptation of the *tail process*, introduced in [10], to a tool more applicable in the dynamical setting. In [10] and subsequent papers (for example, [7, 9]), the tail process was always defined under the assumption that the original process  $\mathbf{X}_0, \mathbf{X}_1, \dots$  is *jointly regularly varying* (see Definition 3.21), which is not natural to assume *a priori* in the dynamical systems setting. This assumption (joint regular variation) together with an assumption on the dependence structure stronger than  $\tilde{\mathcal{A}}'_{q_n}$  ([10, Condition 4.1]) allowed the authors there to prove the existence of the tail process and several very useful properties about it. Motivated by the applications to dynamical systems, here, we will not assume, *a priori*, joint regular variation and since  $\mathcal{A}'_{q_n}$  is even weaker than  $\tilde{\mathcal{A}}'_{q_n}$ , some of the properties of the tail process will be required as adapted assumptions in the definition of the transformed anchored tail process. One of the advantages is that we obtain very general enriched functional limits for extremal processes and record point processes, for example, without assuming regularly varying tails.

**3.3.1. The normalisation of the blocks.** We recall that under assumption  $\mathcal{A}'_{q_n}$  the information regarding to the structure of the clusters is kept in each block of size  $r_n$ , which we are going to normalise in the following way, by defining for each  $i < j \in \{0, \dots, n\}$

$$\mathbb{X}_n^{i,j} = \left( u_n^{-1}(\|\mathbf{X}_i\|) \frac{\mathbf{X}_i}{\|\mathbf{X}_i\|}, \dots, u_n^{-1}(\|\mathbf{X}_{j-1}\|) \frac{\mathbf{X}_{j-1}}{\|\mathbf{X}_{j-1}\|} \right), \quad \mathbb{X}_{n,i} := \mathbb{X}_n^{(i-1)r_n, ir_n} \quad (3.13)$$

so that  $\mathbb{X}_{n,i}$  denotes the  $i$ -th normalised block. We also use the notation  $\mathbb{X}_n^j = \mathbb{X}_n^{j,j+1}$ , for all  $j = 0, \dots, n$ .

Note that the normalisation used is such that the norm of each normalised variable in each block is equal to the asymptotic frequency  $\tau = u_n^{-1}(\|\mathbf{X}_j\|)$  corresponding to the mean number of exceedances of a threshold placed at the value  $\|\mathbf{X}_j\|$ , among the first  $n$  observations of the process.

In particular, observe that by (2.17), for all  $\tau > 0$ , we have

$$\|\mathbb{X}_n^j\| = \left\| u_n^{-1}(\|\mathbf{X}_j\|) \frac{\mathbf{X}_j}{\|\mathbf{X}_j\|} \right\| < \tau \Leftrightarrow u_n^{-1}(\|\mathbf{X}_j\|) < \tau \Leftrightarrow \|\mathbf{X}_j\| > u_n(\tau). \quad (3.14)$$

**3.3.2. The Extremal Index.** Before we characterise the transformed anchored tail process, we define the Extremal Index (EI), denoted by  $\theta \in [0, 1]$ , which was formally introduced by Leadbetter in [52] and measures the degree of clustering of exceedances. When  $\theta = 1$  we have no clustering and a small  $\theta$  means intense clustering. A common interpretation for the EI is that it is reciprocal of the average cluster size (see [1]). We define the EI following O'Brien's formula ([67]) and assume that, for all  $\tau > 0$ , we have

$$\theta = \lim_{n \rightarrow \infty} \frac{\mathbb{P}(U_n^{(q_n)}(\tau))}{\mathbb{P}(U_n(\tau))}. \quad (3.15)$$

**3.3.3. Underlying spaces.** Let  $\dot{\mathcal{V}} = \overline{\mathbb{R}^d} \setminus \{0\} = (\mathbb{R}^d \cup \{\infty\}) \setminus \{0\}$ ,  $\mathcal{V} = \mathbb{R}^d$  and recall that  $\infty \in \dot{\mathcal{V}}$  can be thought of as any point in the completion of  $\mathbb{R}^d$  which is not contained in  $\mathbb{R}^d$ : this has infinite norm. Define

$$l_\infty = \left\{ \mathbf{x} = (x_j)_j \in \dot{\mathcal{V}}^{\mathbb{Z}} : \lim_{|j| \rightarrow \infty} \|x_j\| = \infty \right\} \quad l_0 = \left\{ \mathbf{x} = (x_j)_j \in \mathcal{V}^{\mathbb{Z}} : \lim_{|j| \rightarrow \infty} \|x_j\| = 0 \right\},$$

where for definiteness we are taking the usual Euclidean norm  $\|\cdot\|$  in  $\mathbb{R}^d$ . The transformed anchored tail process will be defined to take values in  $l_\infty$ , while the tail process lives in  $l_0$ . The space  $l_\infty$  will borrow the metric structure of  $l_0$  by means of the map  $P: l_\infty \rightarrow l_0$  given by  $P((x_j)_j) = (p(x_j))_j$ , where

$$p: \dot{\mathcal{V}} \rightarrow \mathcal{V} \\ x \mapsto \begin{cases} \frac{x}{\|x\|^2}, & \text{if } x \neq \infty, \\ 0, & \text{otherwise.} \end{cases}$$

**Lemma 3.4.** *The map  $P$  is invertible.*

*Proof.* By definition of  $P$ , we only need to show that  $p$  is invertible. First note that  $p(x) = 0$  if and only if  $x = \infty$ . Let  $x, y \in \dot{\mathcal{V}} \setminus \{\infty\}$ . Then  $p(x) = p(y)$  implies that

$$x = \frac{\|x\|^2}{\|y\|^2} y. \quad (3.16)$$

Let  $c = \frac{\|x\|^2}{\|y\|^2}$ . Then  $x = cy$  which means that  $\|x\| = c\|y\|$ . Substituting back in (3.16) we obtain  $x = c^2y$ . Hence, we must have  $cy = c^2y$ , which implies that  $c = 1$  and, therefore,  $x = y$ . Hence  $P$  is one-to-one.

To see that  $P$  is onto, we let  $\mathbf{y} \in l_0$  and show that there exists  $\mathbf{x} \in l_\infty$  such that  $P(\mathbf{x}) = \mathbf{y}$ . For all the  $j \in \mathbb{Z}$  such that  $y_j = 0$ , we set  $x_j = \infty$ . For all the other  $j \in \mathbb{Z}$ , we require  $x_j/\|x_j\|^2 = y_j$ , which implies that  $\|y_j\| = 1/\|x_j\|$ . But then,  $x_j = \|x_j\|^2 y_j = y_j/\|y_j\|^2$ , which means that by setting  $x_j = p(y_j)$ , for all such  $j \in \mathbb{Z}$ , we have defined the desired  $\mathbf{x} \in l_\infty$ .  $\square$

As in [9], in  $l_0$ , we consider the supremum norm given by

$$\|\mathbf{x}\|_\infty = \sup_{j \in \mathbb{Z}} \|x_j\|,$$

and the complete metric defined on  $l_0 \setminus \{\mathbf{0}\}$ , where  $\mathbf{0} = (0)_j$ ,

$$d'(\mathbf{x}, \mathbf{y}) = (\|\mathbf{x} - \mathbf{y}\|_\infty \wedge 1) \vee \left| \frac{1}{\|\mathbf{x}\|_\infty} - \frac{1}{\|\mathbf{y}\|_\infty} \right|. \quad (3.17)$$

Now, we consider the metric  $d$  defined on  $l_\infty \setminus \{(\infty)_j\}$  given by

$$d(\mathbf{x}, \mathbf{y}) = d'(P(\mathbf{x}), P(\mathbf{y})). \quad (3.18)$$

Recall that  $l_0 \setminus \{\mathbf{0}\}$  equipped with the metric  $d'$  is a complete separable metric space. Since  $P$  is invertible and componentwise continuous, one can easily show that  $l_\infty \setminus \{(\infty)_j\}$  equipped with the metric  $d$  is a complete separable metric space.

Note that we can embed  $\cup_{n \in \mathbb{N}} \dot{\mathcal{V}}^n$  ( $\cup_{n \in \mathbb{N}} \mathcal{V}^n$ ) into  $l_\infty$  ( $l_0$ ) simply by adding a sequence of  $\infty$  (0) before and after the  $n$  entrances of any element of  $\dot{\mathcal{V}}^n$  ( $\mathcal{V}^n$ ). For example,  $\mathbb{X}_n^{i,j}$  can be seen as an element of  $l_\infty$  by identifying it with

$$\left( \dots, \infty, \infty, u_n^{-1}(\|\mathbf{X}_i\|) \frac{\mathbf{X}_i}{\|\mathbf{X}_i\|}, \dots, u_n^{-1}(\|\mathbf{X}_{j-1}\|) \frac{\mathbf{X}_{j-1}}{\|\mathbf{X}_{j-1}\|}, \infty, \infty, \dots \right). \quad (3.19)$$

We define the quotient spaces  $\tilde{l}_\infty = l_\infty / \sim$  and  $\tilde{l}_0 = l_0 / \sim$ , where  $\sim$  is the equivalence relation defined on both  $l_\infty$  and  $l_0$  by  $\mathbf{x} \sim \mathbf{y}$  if and only if there exists  $k \in \mathbb{Z}$  such that  $\sigma^k(\mathbf{x}) = \mathbf{y}$ , where  $\sigma$  is the shift operator defined in (3.1). Also let  $\tilde{\pi}$  denote the natural projection from  $l_\infty$  ( $l_0$ ) to  $\tilde{l}_\infty$  ( $\tilde{l}_0$ ), which assigns to each element  $\mathbf{x}$  of  $l_\infty$  ( $l_0$ ) the corresponding equivalence class  $\tilde{\pi}(\mathbf{x}) = \tilde{\mathbf{x}}$  in  $\tilde{l}_\infty$  ( $\tilde{l}_0$ ). Given any vector  $v$  of  $\dot{\mathcal{V}}^m$  ( $\mathcal{V}^m$ ), for some  $m \in \mathbb{N}$ , we write  $\tilde{\pi}(v)$  for the projection of the natural embedding of  $v$  into  $l_\infty$  ( $l_0$ ) to the quotient space  $\tilde{l}_\infty$  ( $\tilde{l}_0$ ). Namely,

$$\tilde{\pi}(\mathbb{X}_n^{i,j}) = \tilde{\pi} \left( \left( \dots, \infty, u_n^{-1}(\|\mathbf{X}_i\|) \frac{\mathbf{X}_i}{\|\mathbf{X}_i\|}, \dots, u_n^{-1}(\|\mathbf{X}_{j-1}\|) \frac{\mathbf{X}_{j-1}}{\|\mathbf{X}_{j-1}\|}, \infty, \dots \right) \right). \quad (3.20)$$

Observe that, since  $P$  is invertible, we may define  $\tilde{P} : \tilde{l}_\infty \rightarrow \tilde{l}_0$  so that  $\tilde{P}(\tilde{\pi}(\mathbf{x})) = \tilde{\pi}(P(\mathbf{x}))$ .

Consider the metric  $\tilde{d}'$  in  $\tilde{l}_0 \setminus \{\tilde{\mathbf{0}}\}$  given by

$$\tilde{d}'(\tilde{\mathbf{x}}, \tilde{\mathbf{y}}) = \inf \{d'(\mathbf{x}', \mathbf{y}') : \mathbf{x}' \in \pi^{-1}(\tilde{\mathbf{x}}), \mathbf{y}' \in \pi^{-1}(\tilde{\mathbf{y}})\} = \inf \{d'(\sigma^k(\mathbf{x}), \sigma^m(\mathbf{y})) : k, m \in \mathbb{Z}\}. \quad (3.21)$$

This metric makes  $\tilde{l}_0 \setminus \{\tilde{\mathbf{0}}\}$  a complete separable metric space. (See Lemma 2.1 and Lemma 6.1 of [9]). Accordingly, on  $\tilde{l}_\infty \setminus \{\tilde{\infty}\}$ , where  $\tilde{\infty} = (\infty)_j$ , we define the metric

$$\tilde{d}(\tilde{\mathbf{x}}, \tilde{\mathbf{y}}) = \tilde{d}'(\tilde{P}(\tilde{\mathbf{x}}), \tilde{P}(\tilde{\mathbf{y}})),$$

which also gives a complete separable metric space.

*Remark 3.5.* The choice of the metric implies that a set  $A \subset \tilde{l}_\infty \setminus \{\tilde{\infty}\}$  is bounded if and only if there exists  $\varepsilon > 0$ , such that for all  $\tilde{\mathbf{x}} \in A$  we have  $\|\tilde{P}(\tilde{\mathbf{x}})\|_\infty > \varepsilon$  or, equivalently, that  $\inf_{j \in \mathbb{Z}} \|x_j\| < 1/\varepsilon$ .

For  $A \in \mathcal{F}$ , as in (3.6), we define

$$\tilde{A} = \{\tilde{\mathbf{x}} \in \tilde{l}_\infty : \tilde{\pi}^{-1}(\tilde{\mathbf{x}}) \cap A \neq \emptyset\}; \quad \tilde{\mathcal{F}} = \{\tilde{A} : A \in \mathcal{F}\}. \quad (3.22)$$

Using that  $\mathcal{F}$  is a field, one can show that the class of subsets  $\tilde{\mathcal{F}}$  is closed for unions. Indeed, this follows easily by observing that

$$\tilde{A} \cup \tilde{B} = \{\tilde{\mathbf{x}} \in \tilde{l}_\infty : \tilde{\pi}^{-1}(\tilde{\mathbf{x}}) \cap (A \cup B) \neq \emptyset\}.$$



Let  $\tilde{\mathcal{H}}$  denote the class of subsets of  $\tilde{l}_\infty, \tilde{l}_0$  corresponding to the ring generated by  $\tilde{\mathcal{J}}$ , which is actually a field because by definition of  $\mathcal{F}$ , we have that  $l_\infty, l_0 \subset \mathcal{F}$ .

For  $A \in \mathcal{F}$ , let

$$\mathbb{A} = \{\mathbf{x} \in l_\infty : \tilde{\pi}(\mathbf{x}) \in \tilde{A}\} = \left\{ \mathbf{x} \in l_\infty, l_0 : \mathbf{x} \in \bigcup_{j \in \mathbb{Z}} \sigma^{-j}(A) \right\}, \quad (3.23)$$

$$\mathcal{J} = \{\mathbb{A} : A \in \mathcal{F}\} \text{ and } \mathcal{R} \text{ be the ring generated by } \mathcal{J}. \quad (3.24)$$

If we start with some  $A_\ell \in \mathcal{F}$  here, we correspondingly write  $\mathbb{A}_\ell$ . Note that  $\sigma^{-1}(\mathbb{A}) = \mathbb{A}$ , which means that both  $\mathcal{J}$  and  $\mathcal{R}$  are  $\sigma$ -invariant classes of subsets of  $l_\infty, l_0$ . Also observe that  $\mathcal{R} = \tilde{\pi}^{-1}(\tilde{\mathcal{R}})$  and  $\mathcal{J} = \tilde{\pi}^{-1}(\tilde{\mathcal{J}})$ , which is also closed for unions.

**3.3.4. The transformed anchored tail process.** We can now define the transformed tail process, which presupposes the existence of a process  $(Y_j)_{j \in \mathbb{Z}} \in l_\infty$  satisfying the following assumptions:

- (1)  $\mathcal{L} \left( \frac{1}{\tau} \mathbb{X}_n^{r_n+s, r_n+t} \mid \|\mathbf{X}_{r_n}\| > u_n(\tau) \right) \xrightarrow{n \rightarrow \infty} \mathcal{L}((Y_j)_{j=s, \dots, t})$ , for all  $s < t \in \mathbb{Z}$  and all  $\tau > 0$ ;
- (2) the process  $(\Theta_j)_{j \in \mathbb{Z}}$  given by  $\Theta_j = \frac{Y_j}{\|Y_0\|}$  is independent of  $\|Y_0\|$ ;
- (3)  $\lim_{|j| \rightarrow \infty} \|Y_j\| = \infty$  a.s.;
- (4)  $\mathbb{P}(\inf_{j \leq -1} \|Y_j\| \geq 1) > 0$ .

Here  $(r_n)_n$  is assumed to satisfy (3.2): in our applications it is the sequence appearing in  $\Pi_{q_n}$  and  $\Pi'_{q_n}$ .

*Remark 3.6.* Most of the applications given here are to non-invertible discrete dynamical systems, which means that the sequence  $X_0, X_1, \dots$  is one-sided and therefore we needed to recentre by  $r_n$  so that we can obtain a bi-infinite sequence which includes the past. The particular role of  $r_n$  is not important as long as its is asymptotically larger than  $q_n$ . Alternatively, we could have considered the natural extension of the system to obtain a two-sided sequence  $\dots, X_{-1}, X_0, X_1, \dots$  and then condition on  $\|\mathbf{X}_0\| > u_n(\tau)$ , instead.

*Remark 3.7.* We remark that in the setting of heavy tailed distributions, the process  $(Y_j)_{j \in \mathbb{Z}}$  defined here (which lives in  $l_\infty$ ) is a transformed version of the tail process introduced in [10], which takes values in  $l_0$  and was used later in [7, 9], for example. The existence of such a sequence for stationary heavy tailed stochastic processes was proved to be equivalent to joint regular variation, which we define in Section 3.5, where further details on the relations with the tail process are also given.

We finally define the transformed anchored tail process by considering the canonical anchor used in [10, 9], which corresponds to conditioning on the fact that  $Y_0$  is marking the beginning of a new cluster. For more general anchors we refer to [8].

**Definition 3.8.** Assuming the existence of a sequence  $(Y_j)_{j \in \mathbb{Z}}$  satisfying conditions (1)–(4), we define the *transformed anchored tail process*  $(Z_j)_{j \in \mathbb{Z}}$  as a sequence of random vectors satisfying

$$\mathcal{L}((Z_j)_{j \in \mathbb{Z}}) = \mathcal{L} \left( (Y_j)_{j \in \mathbb{Z}} \mid \inf_{j \leq -1} \|Y_j\| \geq 1 \right).$$

We consider a polar decomposition of the transformed anchored tail process by defining the random variable  $L_Z$  and the process  $(Q_j)_{j \in \mathbb{Z}}$  by

$$L_Z = \inf_{j \in \mathbb{Z}} \|Z_j\| \quad Q_j = \frac{Z_j}{L_Z}. \quad (3.25)$$

We carry this polar decomposition to  $\tilde{l}_\infty$  by letting  $\mathbb{S} = \{\tilde{\mathbf{x}} \in \tilde{l}_\infty : \|\tilde{P}(\tilde{\mathbf{x}})\|_\infty = 1\}$  and defining the map

$$\begin{aligned} \psi: \tilde{l}_\infty \setminus \{\tilde{\infty}\} &\longrightarrow \mathbb{R}^+ \times \mathbb{S} \\ \tilde{\mathbf{x}} &\mapsto \left( \frac{1}{\|\tilde{P}(\tilde{\mathbf{x}})\|_\infty}, \frac{\tilde{x}}{(\|\tilde{P}(\tilde{\mathbf{x}})\|_\infty)^{-1}} \right). \end{aligned} \quad (3.26)$$

We define  $\tilde{\mathbf{Q}} := \tilde{\pi}((Q_j)_{j \in \mathbb{Z}})$  and observe that  $\psi(\tilde{\pi}((Z_j)_{j \in \mathbb{Z}})) = (L_Z, \tilde{\mathbf{Q}})$ .

In order to illustrate the advantage of considering the transformed version of tail process rather than the original version given by [10, equation (1.1)] (or [9, equation (1.6)], [50, equation (5.2.3)]), we consider a concrete dynamical system with an observable which commonly arises in the study of extremal dynamics.

**Example 3.9.** Let  $T: [0, 1] \rightarrow [0, 1]$  be the doubling map:  $T(x) = 2x \bmod 1$ . Let  $\zeta_1 = 1/3$  and  $\zeta_2 = T(\zeta_1) = 2/3$  denote the period two orbit of  $T$ . Define also the observable function  $\psi: [0, 1] \rightarrow \mathbb{R}$  by:

$$\psi(x) = -\log|x - \zeta_1| + \log|x - \zeta_2|$$

Consider now the stochastic process  $X_0, X_1, \dots$  given by  $X_n = \psi \circ T_\gamma^n$ .

For this stochastic process, the tail process,  $(\tilde{Y}_j)_{j \in \mathbb{Z}}$ , given by [10, equation (1.1)] is ill defined. In fact, it is easy to observe that  $\mathbb{P}(|\tilde{Y}_0| > y) = 0$  for all  $y > 1$ , which contrasts with formula [50, equation (5.2.4)] which establishes that  $\mathbb{P}(|\tilde{Y}_0| > y) = (y \vee 1)^{-\alpha}$ . In contrast, the transformed tail process is well defined and we easily obtain that the transformed anchored tail process is equal to (see Appendix D.1):

$$(\dots, \infty, \infty, U \cdot E, U \cdot E(-1)2, U \cdot E(-1)^2 2^2, \dots, U \cdot E(-1)^k 2^k, \dots)$$

where  $U$  is a uniformly distributed random variable on  $[0, 1]$ , the independent random variable  $E$  is such that  $\mathbb{P}(E = 1) = \frac{1}{2} = \mathbb{P}(E = -1)$ .

Note that due to assumption (3) both  $(Y_j)_{j \in \mathbb{Z}}$  and the transformed anchored tail process  $(Z_j)_{j \in \mathbb{Z}}$  take values in  $l_\infty$ .

**3.3.5. Properties of the transformed anchored tail process.** The following lemma is a nice consequence for  $\|Y_0\|$  of our transformed anchored tail process being based in  $l_\infty$ .

**Lemma 3.10.** *The random variable  $\|Y_0\|$  is uniformly distributed.*

*Proof.* Using stationarity (2.15) and (2.17), it follows that for all  $v \in [0, 1]$ ,

$$\begin{aligned} \mathbb{P}(\|Y_0\| < v) &= \lim_{n \rightarrow \infty} \mathbb{P} \left( \left\| \frac{u_n^{-1}(\|\mathbf{X}_{r_n}\|)}{\tau} \frac{\mathbf{X}_{r_n}}{\|\mathbf{X}_{r_n}\|} \right\| < v \mid \|\mathbf{X}_{r_n}\| > u_n(\tau) \right) \\ &= \lim_{n \rightarrow \infty} \mathbb{P}(\|\mathbf{X}_{r_n}\| > u_n(\tau v) \mid \|\mathbf{X}_{r_n}\| > u_n(\tau)) = \frac{\tau v}{\tau} = v. \end{aligned}$$

□

Next we show a relation that will provide a connection between the transformed anchored tail process and the outer measure used in [30]: this describes clustering by splitting the events into annuli of different cluster lengths.

**Proposition 3.11.** *Let  $A_1 \in \mathcal{F}$ . Under  $\mathcal{A}'_{q_n}$ , for  $\mathbb{X}_{n,1}$  as in (3.13),*

$$\lim_{n \rightarrow \infty} \left| k_n \mathbb{P}(\mathbb{X}_{n,1} \in A_1) - n \mathbb{P}(A_{n,1}^{(q_n)}) \right| = \lim_{n \rightarrow \infty} \left| k_n \mathbb{P}(\mathcal{W}_{r_n}^c(A_{n,1})) - n \mathbb{P}(A_{n,1}^{(q_n)}) \right| = 0,$$

where  $A_{n,1}$  is defined in (3.9) and  $A_1$  is from (3.23) applied to  $A_1$ .

*Proof.* We start by estimating  $\mathbb{P}(\mathbb{X}_{n,1} \in \mathbb{A}_1) = \mathbb{P}(\mathcal{W}_{r_n}^c(A_{n,1}))$ , which we do by decomposing  $\mathcal{W}_{r_n}^c(A_{n,1})$  according to the last occurrence of event  $A_{n,1}$ . Namely,

$$\mathbb{P}(\mathcal{W}_{r_n}^c(A_{n,1})) = \sum_{j=0}^{r_n-1} \mathbb{P}(\sigma^{-j}(A_{n,1})) - \sum_{j=0}^{r_n-1} \mathbb{P}(\sigma^{-j}(A_{n,1}), \mathcal{W}_{j+1, r_n}^c(A_{n,1})).$$

It follows that

$$\begin{aligned} & \left| \mathbb{P}(\mathcal{W}_{r_n}^c(A_{n,1})) - \left( \sum_{j=0}^{r_n-q_n-2} (\mathbb{P}(\sigma^{-j}(A_{n,1})) - \mathbb{P}(\sigma^{-j}(A_{n,1}), \mathcal{W}_{j+1, r_n}^c(A_{n,1}))) \right) \right| \\ & \leq \sum_{j=r_n-q_n-1}^{r_n-1} \mathbb{P}(\sigma^{-j}(A_{n,1}), \mathcal{W}_{j+1, r_n}^c(A_{n,1})) \leq (q_n + 1)\mathbb{P}(A_{n,1}) \end{aligned} \quad (3.27)$$

Using stationarity,

$$\sum_{j=0}^{r_n-q_n-2} \mathbb{P}(\sigma^{-j}(A_{n,1}), \mathcal{W}_{j+1, r_n}^c(A_{n,1})) = \sum_{s=q_n+2}^{r_n} \mathbb{P}(A_{n,1}, \mathcal{W}_{1, s}^c(A_{n,1})).$$

For  $s \geq q_n + 2$ ,

$$\mathbb{P}(A_{n,1}, \mathcal{W}_{1, s}^c(A_{n,1})) - \mathbb{P}(A_{n,1}, \mathcal{W}_{1, q_n+1}^c(A_{n,1})) = \mathbb{P}(A_{n,1}^{(q_n)}, \mathcal{W}_{q_n+1, s}^c(A_{n,1}))$$

and therefore

$$\begin{aligned} & \left| \sum_{s=q_n+2}^{r_n} \mathbb{P}(A_{n,1}, \mathcal{W}_{1, s}^c(A_{n,1})) - \sum_{s=q_n+2}^{r_n} \mathbb{P}(A_{n,1}, \mathcal{W}_{1, q_n+1}^c(A_{n,1})) \right| \\ & \leq \sum_{s=q_n+2}^{r_n} \mathbb{P}(A_{n,1}^{(q_n)}, \mathcal{W}_{q_n+1, s}^c(A_{n,1})) \leq r_n \mathbb{P}(A_{n,1}^{(q_n)}, \mathcal{W}_{q_n+1, r_n}^c(A_{n,1})). \end{aligned} \quad (3.28)$$

Combining (3.27) and (3.28) and using stationarity, we obtain

$$\begin{aligned} & \left| \mathbb{P}(\mathcal{W}_{r_n}^c(A_{n,1})) - \left( \sum_{s=q_n+2}^{r_n} \mathbb{P}(A_{n,1}) - \sum_{s=q_n+2}^{r_n} \mathbb{P}(A_{n,1}, \mathcal{W}_{1, q_n+1}^c(A_{n,1})) \right) \right| \\ & \leq (q_n + 1)\mathbb{P}(A_{n,1}) + r_n \mathbb{P}(A_{n,1}^{(q_n)}, \mathcal{W}_{q_n+1, r_n}^c(A_{n,1})) \end{aligned}$$

Noting that the term between big brackets is equal to

$(r_n - q_n + 1)\mathbb{P}(A_{n,1}, \mathcal{W}_{1, q_n+1}^c(A_{n,1})) = (r_n - q_n + 1)\mathbb{P}(A_{n,1}^{(q_n)})$  then multiplying by  $k_n$  we obtain

$$\left| k_n \mathbb{P}(\mathbb{X}_{n,1} \in \mathbb{A}_1) - n \mathbb{P}(A_{n,1}^{(q_n)}) \right| \leq 2q_n k_n \mathbb{P}(A_{n,1}^{(q_n)}) + n \mathbb{P}(A_{n,1}^{(q_n)}, \mathcal{W}_{q_n+1, r_n}^c(A_{n,1})).$$

The second term on the right vanishes by  $\mathcal{A}'_{q_n}$ . Since by definition of  $A_{n,1}$ , we have that  $A_{n,1} \subset \{\|X_0\| > u_n(h_0)\}$ , where  $h_0 = \inf\{\|x\| : x \in H_0\} \geq 0$ , then  $n\mathbb{P}(A_{n,1}) \leq n\mathbb{P}(\|X_0\| > u_n(h_0)) \xrightarrow{n \rightarrow \infty} h_0$ . Recalling that  $q_n = o(r_n)$ , it follows that the first term on right also vanishes.  $\square$

**Corollary 3.12.** Under  $\mathcal{A}'_{q_n}$ ,

$$\lim_{n \rightarrow \infty} k_n \mathbb{P}(\mathcal{W}_{r_n}^c(U_n(\tau))) = \theta \tau \quad \text{and} \quad \lim_{n \rightarrow \infty} \frac{\mathbb{P}(\mathcal{W}_{r_n}^c(U_n(\tau)))}{r_n \mathbb{P}(U_n(\tau))} = \theta.$$

*Proof.* By Proposition 3.11, (2.15) and (3.15),

$$\lim_{n \rightarrow \infty} k_n \mathbb{P}(\mathcal{W}_{r_n}^c(U_n(\tau))) = \lim_{n \rightarrow \infty} n \mathbb{P}(U_n^{q_n}(\tau)) = \lim_{n \rightarrow \infty} n \mathbb{P}(U_n(\tau)) \frac{\mathbb{P}(U_n^{q_n}(\tau))}{\mathbb{P}(U_n(\tau))} = \tau \theta.$$

It follows that

$$\lim_{n \rightarrow \infty} \frac{\mathbb{P}(\mathcal{W}_{r_n}^c(U_n(\tau)))}{r_n \mathbb{P}(U_n(\tau))} = \frac{k_n \mathbb{P}(\mathcal{W}_{r_n}^c(U_n(\tau)))}{n \mathbb{P}(U_n(\tau))} = \frac{\tau \theta}{\tau} = \theta.$$

□

**Lemma 3.13.** *Recalling the definition of the EI given in (3.15),*

$$\theta = \mathbb{P} \left( \inf_{j \geq 1} \|Y_j\| \geq 1 \right) = \mathbb{P} \left( \inf_{j \leq -1} \|Y_j\| \geq 1 \right).$$

*Proof.* By (3.15), stationarity and definition of the process  $(Y_j)_{j \in \mathbb{Z}}$ , given in assumption (1), we may write

$$\begin{aligned} \theta &= \lim_{n \rightarrow \infty} \mathbb{P} \left( \frac{1}{\tau} \mathbb{X}_n^{r_n, r_n + q_n} \in \mathcal{W}_{r_n + 1, r_n + q_n + 1}(U(1)) \mid \|\mathbf{X}_{r_n}\| > u_n(\tau) \right) \\ &= \mathbb{P} \left( \inf_{j \geq 1} \|Y_j\| \geq 1 \right). \end{aligned}$$

The second equality in the statement of the lemma follows easily from stationarity and standard arguments. □

The next result is instrumental because it shows how the transformed anchored tail process can be used to encode the information regarding the clustering. Essentially, it says that the joint distribution of the random variables in a block where an exceedance is observed (which makes it a cluster) is given by the transformed anchored tail process. Recall that by (3.19),  $\mathbb{X}_{n,i}$  can be thought of as lying in  $l_\infty$ .

**Proposition 3.14.** *Under the assumptions used to define the transformed anchored tail process and condition  $\mathcal{A}'_{q_n}$ , for every  $\tau > 0$  and for  $\mathbb{X}_{n,1}$  as in (3.13),*

$$\mathcal{L} \left( \tilde{\pi} \left( \frac{1}{\tau} \mathbb{X}_{n,1} \right) \mid \mathbb{X}_{n,1} \in \mathcal{W}_{r_n}^c(U(\tau)) \right) \longrightarrow \mathcal{L}(\tilde{\pi}((Z_j)_{j \in \mathbb{Z}})).$$

*Proof.* In what follows we write  $U^\tau$  for  $U(\tau)$ . As in Appendix B,  $\tilde{\mathcal{J}}$  is a convergence determining class and therefore we need to show that for all  $A \in \mathcal{F}$  and corresponding  $\tilde{A} \in \tilde{\mathcal{J}}$ , such that  $\mathbb{P}(\tilde{\pi}((Z_j)_{j \in \mathbb{Z}}) \in \partial \tilde{A}) = 0$ , we have

$$\mathbb{P} \left( \tilde{\pi} \left( \frac{1}{\tau} \mathbb{X}_{n,1} \right) \in \tilde{A} \mid \mathbb{X}_{n,1} \in \mathcal{W}_{r_n}^c(U^\tau) \right) \longrightarrow \mathbb{P}(\tilde{\pi}((Z_j)_{j \in \mathbb{Z}}) \in \tilde{A}),$$

which will follow if we show that

$$\mathbb{P} \left( \frac{1}{\tau} \mathbb{X}_{n,1} \in \mathbb{A} \mid \mathbb{X}_{n,1} \in \mathcal{W}_{r_n}^c(U^\tau) \right) \longrightarrow \mathbb{P} \left( (Y_j)_{j \in \mathbb{Z}} \in \mathbb{A} \mid \inf_{j \leq -1} \|Y_j\| \geq 1 \right),$$

for all  $\mathbb{A} \in \mathcal{J}$ , such that  $\mathbb{P}((Y_j)_{j \in \mathbb{Z}} \in \partial \mathbb{A}) = 0$ . Recall that  $\mathbb{A}$  is  $\sigma$ -invariant, i.e.,  $\sigma^{-1}(\mathbb{A}) = \mathbb{A}$ .

We start estimating  $\mathbb{P}(\mathbb{X}_{n,1} \in \tau \mathbb{A} \cap \mathcal{W}_{r_n}^c(U^\tau))$  by decomposing the event on the right (which essentially says that at least one exceedance of  $u_n(\tau)$  has occurred up to time  $r_n - 1$ ) with respect to the first time,  $i = 0, \dots, r_n - 1$ , when that exceedance occurs, i.e.,  $\{\mathbb{X}_{n,1} \in \sigma^{-i}(U^\tau)\} = \{\|\mathbf{X}_i\| > u_n(\tau)\}$ :

$$\mathbb{P}(\mathbb{X}_{n,1} \in \tau \mathbb{A} \cap \mathcal{W}_{r_n}^c(U^\tau)) = \sum_{i=0}^{r_n-1} \mathbb{P}(\mathbb{X}_{n,1} \in \tau \mathbb{A} \cap \mathcal{W}_i(U^\tau) \cap \sigma^{-i}(U^\tau))$$

Since  $B_{n,i} := \mathbb{P}(\mathbb{X}_{n,1} \in \tau\mathbb{A} \cap \mathscr{W}_i(U^\tau) \cap \sigma^{-i}(U^\tau)) \leq \mathbb{P}(\|\mathbf{X}_i\| > u_n(\tau))$ ,

$$\left| \mathbb{P}(\mathbb{X}_{n,1} \in \tau\mathbb{A} \cap \mathscr{W}_{r_n}^c(U^\tau)) - \sum_{i=q_n}^{r_n-1} B_{n,i} \right| \leq q_n \mathbb{P}(U_n(\tau)) =: I(n) \quad (3.29)$$

For  $i \geq q_n$ , we use  $D_{n,i} := \mathbb{P}(\mathbb{X}_{n,1} \in \tau\mathbb{A} \cap \mathscr{W}_{i-q_n,i}(U^\tau) \cap \sigma^{-i}(U^\tau))$  to estimate  $B_{n,i}$ . Namely,

$$|B_{n,i} - D_{n,i}| \leq \mathbb{P}(\mathbb{X}_{n,1} \in U^\tau \cap \mathscr{W}_{1,q_n+1}(U^\tau) \cap \mathscr{W}_{q_n+1,i+1}^c(U^\tau))$$

Therefore,

$$\left| \sum_{i=q_n}^{r_n-1} B_{n,i} - \sum_{i=q_n}^{r_n-1} D_{n,i} \right| \leq (r_n - q_n) \mathbb{P}(U_n^{(q_n)}(\tau) \cap \mathscr{W}_{q_n+1,r_n}^c(U_n(\tau))) =: II(n) \quad (3.30)$$

By stationarity and because  $\mathbb{A}$  is  $\sigma$ -invariant, for all  $i = q_n, \dots, r_n - 1$ ,

$$\begin{aligned} D_{n,i} &= \mathbb{P}(\mathbb{X}_{n,2} \in \tau\mathbb{A} \cap \mathscr{W}_{r_n+i-q_n,r_n+i}(U^\tau) \cap \sigma^{-(r_n+i)}(U^\tau)) \\ &= \mathbb{P}(\mathbb{X}_n^{r_n-q_n,2r_n} \in \tau\mathbb{A} \cap \mathscr{W}_{r_n-q_n,r_n}(U^\tau) \cap \sigma^{-r_n}(U^\tau)) =: D_n \end{aligned}$$

Then using estimates (3.29) and (3.30), we obtain

$$\left| \mathbb{P}(\mathbb{X}_{n,1} \in \tau\mathbb{A} \cap \mathscr{W}_{r_n}^c(U^\tau)) - r_n \mathbb{P}(D_n) \right| \leq 2I(n) + II(n).$$

Hence, letting  $P_n := \frac{r_n \mathbb{P}(U_n(\tau))}{\mathbb{P}(\mathbb{X}_{n,1} \in \mathscr{W}_{r_n}^c(U^\tau))} \mathbb{P}(D_n \mid \|\mathbf{X}_{r_n}\| > u_n(\tau))$ , we can write

$$\left| \mathbb{P}\left(\frac{1}{\tau} \mathbb{X}_{n,1} \in \mathbb{A} \mid \mathbb{X}_{n,1} \in \mathscr{W}_{r_n}^c(U^\tau)\right) - P_n \right| \leq (2I(n) + II(n)) \frac{1}{\mathbb{P}(\mathbb{X}_{n,1} \in \mathscr{W}_{r_n}^c(U^\tau))} =: E(n).$$

Since, by Corollary 3.12,  $\lim_{n \rightarrow \infty} \frac{r_n \mathbb{P}(U_n(\tau))}{\mathbb{P}(\mathbb{X}_{n,1} \in \mathscr{W}_{r_n}^c(U^\tau))} = \theta^{-1}$  and, by definition of  $U_n(\tau)$  we have  $\lim_{n \rightarrow \infty} k_n r_n \mathbb{P}(U_n(\tau)) = \lim_{n \rightarrow \infty} n \mathbb{P}(U_n(\tau)) = \tau$ , it follows that for some  $C > 0$ , we have

$$E(n) \leq C(k_n I(n) + k_n II(n)).$$

By definition of the sequences  $(k_n)_n$  and  $(q_n)_n$ , we have that  $\lim_{n \rightarrow \infty} k_n q_n \mathbb{P}(U_n(\tau)) = 0$ , which means that  $\lim_{n \rightarrow \infty} k_n I(n) = 0$ . Observe also that  $\mathcal{A}'_{q_n}$  implies that  $\lim_{n \rightarrow \infty} k_n II(n) = 0$  and therefore  $\lim_{n \rightarrow \infty} E(n) = 0$ .

In order to get the result we need to check that

$$\lim_{n \rightarrow \infty} P_n = \mathbb{P}\left((Y_j)_{j \in \mathbb{N}} \in \mathbb{A} \mid \inf_{j \leq -1} \|Y_j\| \geq 1\right) := P.$$

Since by Corollary 3.12 and Lemma 3.13, we have  $\lim_{n \rightarrow \infty} \frac{\mathbb{P}(\mathbb{X}_{n,1} \in \mathscr{W}_{r_n}^c(U^\tau))}{r_n \mathbb{P}(U_n(\tau))} = \theta$  and  $\theta = \mathbb{P}(\inf_{j \leq -1} \|Y_j\| \geq 1)$ , then we need to show that

$$\lim_{n \rightarrow \infty} \mathbb{P}(\mathbb{X}_n^{r_n-q_n,2r_n} \in \tau\mathbb{A} \cap \mathscr{W}_{r_n-q_n,r_n}(U^\tau) \mid \|\mathbf{X}_{r_n}\| > u_n(\tau)) = P$$

Note that  $\{(Y_j)_{j \in \mathbb{N}} \in \partial \mathscr{W}_{-\infty,0}(U^1)\} \subset \cup_{j \leq 1} \{\|Y_j\| = 1\}$ . Since, by assumption (2), for every  $j \in \mathbb{Z}$ , we have  $Y_j = \|Y_0\| \Theta_j$ , with  $\Theta_j$  independent of  $\|Y_0\|$  and since the latter is uniformly distributed by Lemma 3.10, then  $\mathbb{P}(\|Y_j\| = 1) = 0$ , for all  $j \in \mathbb{Z}$ . Then the desired limit follows by definition of the sequence  $(Y_j)_{j \in \mathbb{N}}$  given in assumption (1).  $\square$

**Corollary 3.15.** *Under the assumptions of Proposition 3.14, we have that  $L_Z$  and the process  $(Q_j)_{j \in \mathbb{Z}}$  defined in (3.25) satisfy*

- (1)  $L_Z$  is uniformly distributed on  $[0, 1]$ ;
- (2)  $L_Z$  and  $\tilde{\pi}((Q_j)_{j \in \mathbb{Z}})$  are independent.

*Proof.* For part (1), note that  $L_Z \leq \|Z_0\| \leq \|Y_0\| \leq 1$  a.s. Let  $0 \leq v \leq 1$ . By Proposition 3.14, (3.14) and Corollary 3.12, we may write

$$\begin{aligned} \mathbb{P}(L_Z < v) &= \lim_{n \rightarrow \infty} \mathbb{P}\left(\frac{1}{\tau} \mathbb{X}_{n,1} \in \mathcal{W}_{r_n}^c(U(v)) \mid \mathbb{X}_{n,1} \in \mathcal{W}_{r_n}^c(U(\tau))\right) \\ &= \lim_{n \rightarrow \infty} \mathbb{P}\left(\mathcal{W}_{r_n}^c(U_n(\tau v)) \mid \mathcal{W}_{r_n}^c(U_n(\tau))\right) = \lim_{n \rightarrow \infty} \frac{\mathbb{P}(\mathcal{W}_{r_n}^c(U_n(\tau v)))}{\mathbb{P}(\mathcal{W}_{r_n}^c(U_n(\tau)))} = \frac{\tau v \theta}{\tau \theta} = v. \end{aligned}$$

To prove (2), we start by observing that  $\frac{1}{\|P(\tilde{\mathbf{x}})\|_\infty} = \inf_{j \in \mathbb{Z}} \|x_j\|$  and that the map  $\tilde{\mathbf{x}} \mapsto \left(\tilde{\mathbf{x}}, \frac{1}{\|\tilde{P}(\tilde{\mathbf{x}})\|_\infty}\right)$  is continuous on  $\tilde{l}_\infty \setminus \{\tilde{\infty}\}$ . Then, by Proposition 3.14 and the CMT

$$\mathcal{L}\left(\tilde{\pi}\left(\frac{1}{\tau} \mathbb{X}_{n,1}\right), \frac{1}{\|\tilde{P}(\tilde{\pi}(\frac{1}{\tau} \mathbb{X}_{n,1}))\|_\infty} \mid \mathbb{X}_{n,1} \in \mathcal{W}_{r_n}^c(U(\tau))\right) \longrightarrow \mathcal{L}(\tilde{\pi}((Z_j)_{j \in \mathbb{Z}}), L_Z)$$

Since the map  $(\tilde{\mathbf{x}}, a) \mapsto (\frac{\tilde{\mathbf{x}}}{b}, a)$  is continuous on  $\tilde{l}_\infty \setminus \{\tilde{\infty}\} \times (0, \infty)$ , then

$$\begin{aligned} \mathcal{L}\left(\left\|\tilde{P}\left(\tilde{\pi}\left(\frac{1}{\tau} \mathbb{X}_{n,1}\right)\right)\right\|_\infty, \tilde{\pi}\left(\frac{1}{\tau} \mathbb{X}_{n,1}\right), \frac{1}{\|\tilde{P}(\tilde{\pi}(\frac{1}{\tau} \mathbb{X}_{n,1}))\|_\infty} \mid \mathbb{X}_{n,1} \in \mathcal{W}_{r_n}^c(U(\tau))\right) \\ \longrightarrow \mathcal{L}(\tilde{\pi}((Q_j)_{j \in \mathbb{Z}}), L_Z) \end{aligned} \quad (3.31)$$

Since  $\tilde{\mathcal{J}}$  is a convergence determining class (see Appendix B), the result will follow if we show that for all  $A \in \tilde{\mathcal{F}}$  and corresponding  $\tilde{A} \in \tilde{\mathcal{J}}$ , such that  $\mathbb{P}(\tilde{\pi}((Q_j)_{j \in \mathbb{Z}}) \in \partial \tilde{A}) = 0$ , and all  $v \in [0, 1]$ , we have

$$\mathbb{P}\left(\tilde{\pi}((Q_j)_{j \in \mathbb{Z}}) \in \tilde{A}, L_Z < v\right) = \mathbb{P}\left(\tilde{\pi}((Q_j)_{j \in \mathbb{Z}}) \in \tilde{A}\right) \cdot \mathbb{P}(L_Z < v) \quad (3.32)$$

Letting  $\mathbb{A} \in \mathcal{J}$  be such that  $\mathbb{A} = \tilde{\pi}^{-1}(\tilde{A})$  and  $m_{r_n} = \min\{\|\mathbb{X}_n^0\|, \dots, \|\mathbb{X}_n^{r_n-1}\|\}$ , by (3.31), we can write that

$$\begin{aligned} \mathbb{P}\left((Q_j)_{j \in \mathbb{Z}} \in \mathbb{A}, L_Z < v\right) &= \lim_{n \rightarrow \infty} \mathbb{P}\left(\frac{1}{m_{r_n}} \mathbb{X}_{n,1} \in \mathbb{A} \mid \mathbb{X}_{n,1} \in \mathcal{W}_{r_n}^c(U(\tau v))\right) \\ &\cdot \frac{\mathbb{P}(\mathcal{W}_{r_n}^c(U_n(\tau v)))}{\mathbb{P}(\mathcal{W}_{r_n}^c(U_n(\tau)))} = \mathbb{P}\left(\tilde{\pi}((Q_j)_{j \in \mathbb{Z}}) \in \tilde{A}\right) \cdot v = \mathbb{P}\left(\tilde{\pi}((Q_j)_{j \in \mathbb{Z}}) \in \tilde{A}\right) \cdot \mathbb{P}(L_Z < v). \end{aligned}$$

□

The next result, an analogue of [9, Lemma 3.3] in our setting, formally establishes the convergence of the intensity measures of the cluster point processes we introduce later. We refer to Appendix C for the definitions of weak<sup>#</sup> convergence and boundedly finite measures.

**Corollary 3.16.** *Under the assumptions used to define the transformed anchored tail process and condition  $\mathcal{A}'_{q_n}$ , the sequence of boundedly finite measures  $\eta_n = k_n \mathbb{P}(\tilde{\pi}(\mathbb{X}_{n,1}) \in \cdot)$  in  $\mathcal{M}_{\tilde{l}_\infty \setminus \{\tilde{\infty}\}}^\#$  converges in the  $w^\#$  topology to  $\eta = \theta(\text{Leb} \times \mathbb{P}_{\tilde{\mathbf{Q}}}) \circ \psi$ , where  $\mathbb{P}_{\tilde{\mathbf{Q}}}$  is the distribution of  $\tilde{\pi}((Q_j)_{j \in \mathbb{Z}})$ .*

*Proof.* By Lemma B.4, we only need to check the convergence for all bounded  $\tilde{A} \in \tilde{\mathcal{J}}$  such that  $\mu(\partial \tilde{A}) = 0$ . By Remark 3.5, since  $\tilde{A}$  is bounded, there exists  $\tau > 0$  such that for all  $\tilde{\mathbf{x}} \in \tilde{A}$ , we have  $\inf_{j \in \mathbb{Z}} \|x_j\| < \tau$ . Hence, by 3.14, if  $\tilde{\pi}(\mathbb{X}_{n,1}) \in \tilde{A}$ , then  $\|\mathbf{X}_j\| > u_n(\tau)$  for  $j = 0, \dots, r_n - 1$  and hence  $\mathbb{X}_{n,1} \in \mathcal{W}_{r_n}^c(U(\tau))$ . Let, as before,  $\mathbb{A} = \tilde{\pi}^{-1}(\tilde{A})$ . Then

$$\begin{aligned} k_n \mathbb{P}(\tilde{\pi}(\mathbb{X}_{n,1}) \in \tilde{A}) &= k_n \mathbb{P}(\mathbb{X}_{n,1} \in \mathbb{A}) = k_n \mathbb{P}(\mathbb{X}_{n,1} \in \mathbb{A} \cap \mathcal{W}_{r_n}^c(U(\tau))) \\ &= n \mathbb{P}(\|\mathbf{X}_0\| > u_n(\tau)) \frac{\mathbb{P}(\mathbb{X}_{n,1} \in \mathcal{W}_{r_n}^c(U(\tau)))}{r_n \mathbb{P}(\|\mathbf{X}_0\| > u_n(\tau))} \mathbb{P}\left(\frac{\mathbb{X}_{n,1}}{\tau} \in \tau^{-1} \mathbb{A} \mid \mathbb{X}_{n,1} \in \mathcal{W}_{r_n}^c(U(\tau))\right) \end{aligned}$$

By (2.15) and Corollary 3.12, we have that the first term converges to  $\tau$  and the second to  $\theta$ , as  $n \rightarrow \infty$ . By Proposition 3.11, the third term goes to  $\mathbb{P}((Z_j)_{j \in \mathbb{Z}} \in \tau^{-1}\mathbb{A})$ . We now use Corollary 3.15 in order to finish the proof.

$$\begin{aligned} \mathbb{P}((Z_j)_{j \in \mathbb{Z}} \in \tau^{-1}\mathbb{A}) &= \mathbb{P}(\tau(Z_j)_{j \in \mathbb{Z}} \in \mathbb{A}) = \int_0^1 \mathbb{P}(\tau(Z_j)_{j \in \mathbb{Z}} \in \mathbb{A} \mid L_Z = v) dv \\ &= \int_0^1 \mathbb{P}(\tau v(Q_j)_{j \in \mathbb{Z}} \in \mathbb{A} \mid L_Z = v) dv = \frac{1}{\tau} \int_0^\tau \mathbb{P}(s(Q_j)_{j \in \mathbb{Z}} \in \mathbb{A}) ds = \frac{\eta(\tilde{A})}{\tau\theta}. \end{aligned}$$

□

**3.4. Complete convergence of point processes.** We define and prove the weak convergence of the point processes that keep all the cluster information. We refer to Appendix C for the precise definition of point processes and their weak convergence. Essentially, we consider a random element on the space  $\mathcal{N}_{\mathbb{R}_0^+ \times \tilde{l}_\infty \setminus \{\tilde{\infty}\}}^\#$  of boundedly finite point measures on  $\mathbb{R}_0^+ \times \tilde{l}_\infty \setminus \{\tilde{\infty}\}$ . Namely, similarly to [9], we define the point processes of clusters by

$$N_n = \sum_{i=1}^{\infty} \delta_{(i/k_n, \tilde{\pi}(\mathbb{X}_{n,i}))}. \quad (3.33)$$

Now, we define the point process that will appear as the limit of the cluster point process. Let  $(T_i)_{i \in \mathbb{N}}$  and  $(U_i)_{i \in \mathbb{N}}$  be such that  $\sum_{i=1}^{\infty} \delta_{(T_i, U_i)}$  is a bidimensional Poisson point process on  $\mathbb{R}_0^+ \times \mathbb{R}_0^+$  with intensity measure  $\text{Leb} \times \theta \text{Leb}$ . Also let  $(\tilde{\mathbf{Q}}_i)_{i \in \mathbb{N}}$  be an i.i.d. sequence of random elements in  $\mathbb{S}$  such that each  $\tilde{\mathbf{Q}}_i$  has a distribution given by (3.25). We assume that the sequences  $(T_i)_{i \in \mathbb{N}}$ ,  $(U_i)_{i \in \mathbb{N}}$  and  $(\tilde{\mathbf{Q}}_i)_{i \in \mathbb{N}}$  are mutually independent. We define

$$N = \sum_{i=1}^{\infty} \delta_{(T_i, U_i, \tilde{\mathbf{Q}}_i)}. \quad (3.34)$$

*Remark 3.17.* Note that  $N$  above is a Poisson point process on  $\mathbb{R}_0^+ \times \tilde{l}_\infty \setminus \{\tilde{\infty}\}$  with intensity  $\text{Leb} \times \eta$ , where  $\eta$  is as in Corollary 3.16. Here  $\eta$  describes the distribution of the transformed anchored tail process  $\tilde{\pi}((Z_j)_{j \in \mathbb{Z}})$ , which is characterised by means of the spectral decomposition  $\psi$  given in (3.26), which allows us to identify the contribution from each component, namely, the  $\theta \text{Leb}$  part associated to  $L_Z$  (related to the second coordinate of the bidimensional Poisson point process  $\sum_{i=1}^{\infty} \delta_{(T_i, U_i)}$ ) and the  $\mathbb{P}_{\tilde{\mathbf{Q}}}$  part which is the distribution of  $\tilde{\pi}((Q_j)_{j \in \mathbb{Z}})$ .

We are now ready to state a general complete convergence result.

**Theorem 3.18.** *Let  $\mathbf{X}_0, \mathbf{X}_1, \dots$  be a stationary process of random vectors in  $\mathbb{R}^d$  with proper tails, in the sense given in Section 2.4.1. Assume that the transformed anchored tail process given in Definition 3.8 is well defined and conditions  $\mathbb{A}_{q_n}$  and  $\mathbb{A}'_{q_n}$  hold. Then point process  $N_n$ , given in (3.33) converges weakly in  $\mathcal{N}_{\mathbb{R}_0^+ \times \tilde{l}_\infty \setminus \{\tilde{\infty}\}}^\#$  to the Poisson point process  $N$  given by (3.34).*

*Remark 3.19.* We observe that the point processes convergence stated here corresponds essentially to a transformed version of similar statements in [9, Theorem 3.6] and [50, Corollary 7.4.2], but we emphasise that the dependence conditions assumed here are both weaker, which is of crucial importance for the application to stochastic processes arising from dynamical systems. We refer to Corollary 3.23, in Section 3.5, where our point processes, and convergence properties, are transformed back to provide an easier comparison with previous results and also formulae on how to relate the various objects involved.

In order to prove the theorem, we need essentially to show two things. The first is a sort of *independent increments* property together with convergence in distribution of joint random variables using the avoidance function, *i.e.*, by computing the probability of having no extremal occurrences (see the definition of the avoidance function in Appendix C). This is done in Proposition 3.20, whose complete proof is rather lengthy because we are using the very weak mixing assumption  $\mathbb{D}_{q_n}$ , though a lot of the work for that has been done in previous papers by the authors. Then, in the second step, we need to show that the intensity measures of the processes converge to the right intensity measure. This has actually already been done in Corollary 3.16. Finally we need to join the pieces using the theory of weak<sup>#</sup> convergence on non locally compact spaces developed in [17, 18]. In fact, we needed to redo one of the results to correct a typo and improve it in order to be able to use the convergence of the intensity measures (see Appendix C).

In [30] the existence of a  $\sigma$ -finite outer measure  $\nu$  on  $\mathcal{V}^{\mathbb{N}_0, \mathbb{Z}}$  was assumed, so that the following limit exists

$$\lim_{n \rightarrow \infty} n\mathbb{P}(A_{n,\ell}^{(q_n)}) = \nu(A_\ell), \quad (3.35)$$

for all  $A_\ell \in \mathcal{F}$  and  $A_{n,\ell}^{(q_n)}$  given by equation (3.9). This outer measure described the piling of points on the multidimensional point processes created by clustering in [30]. Note that if we associate  $\tilde{A}_\ell, \mathbb{A}_\ell$  to  $A_\ell \in \mathcal{F}$  as in (3.22) and (3.23), respectively, then using Proposition 3.11 and Corollary 3.16, it follows that when we have the existence of a transformed anchored tail process and condition  $\mathbb{D}'_{q_n}$  then (3.35) holds and

$$\nu(A_\ell) = \mu(\tilde{A}_\ell). \quad (3.36)$$

We state now the main result that provides independence of disjoint time pieces and convergence of joint distributions by use of the avoidance function.

**Proposition 3.20.** *Let  $m \in \mathbb{N}$  and for each  $\ell = 1, \dots, m$  let  $J_\ell, A_\ell$  be given, as in (3.7). For  $n \in \mathbb{N}$ , consider the respective versions  $J_{n,\ell}, A_{n,\ell}$  given in (3.8) and (3.9). Assume that  $\mathbb{D}_{q_n}$  and  $\mathbb{D}'_{q_n}$  hold. Also assume that there exists a  $\sigma$ -finite outer measure  $\nu$  such that (3.35) holds. Then*

$$\lim_{n \rightarrow \infty} \mathbb{P} \left( \bigcap_{\ell=1}^m \mathcal{W}_{J_{n,\ell}}(A_{n,\ell}) \right) = \lim_{n \rightarrow \infty} \mathbb{P} \left( \bigcap_{\ell=1}^m \mathcal{W}_{J_{n,\ell}}(A_{n,\ell}^{(q_n)}) \right) = \prod_{\ell=1}^m e^{-\nu(A_\ell)|J_\ell|}.$$

The first equality follows from the fact that the non-occurrence of the asymptotically rare event  $A_{n,\ell}$  can be replaced by the non-occurrence of the event  $A_{n,\ell}^{(q_n)}$ , up to an asymptotically negligible error. This idea goes back to [33, Proposition 1] and was further developed in [35, Proposition 2.7]. We refer to [30, Proposition 3.2] for a proof. The second equality in Proposition 3.20 follows from minor adjustments to the argument used to prove [30, Theorem 3.3].

We are now ready to prove the weak convergence of the cluster point processes.

*Proof of Theorem 3.18.* By Proposition C.2 and Lemma B.2, we first need to check that for all bounded sets  $B \in \mathcal{I}$ , defined in (C.6) we have  $\lim_{n \rightarrow \infty} \mathbb{P}(N_n(B) = 0) = \mathbb{P}(N(B) = 0)$ . Let  $B = \bigcup_{\ell=1}^m J_\ell \times \tilde{A}_\ell$ , where for each  $\ell = 1, \dots, m$ , we have  $J_\ell = [a_\ell, b_\ell]$  and  $\tilde{A}_\ell$  is associated to some  $A_\ell \in \mathcal{F}$  as in (3.22).

Since  $\tilde{\pi}(\mathbb{X}_{n,i}) \in \tilde{A}_\ell$  corresponds to the event  $\mathbb{X}_{n,i} \in \mathbb{A}_\ell = \bigcup_{i \in \mathbb{Z}} \sigma^{-i}(A_\ell)$ , then, by definition of the sets  $J_{n,\ell}$  and  $A_{n,\ell}$  given in (3.8) and (3.9), we have

$$\mathbb{P}(N_n(B) = 0) = \mathbb{P} \left( \bigcap_{\ell=1}^m \{N_n(J_\ell \times \tilde{A}_\ell) = 0\} \right) = \mathbb{P} \left( \bigcap_{\ell=1}^m \mathcal{W}_{J_{n,\ell}}(A_{n,\ell}) \right).$$



Now, by definition of the Poisson point process  $N$ , we have  $\mathbb{P}(N(B) = 0) = \prod_{\ell=1}^m e^{-|J_\ell|\mu(\tilde{A}_\ell)}$ . Hence, condition (i) of Proposition C.2 follows from Proposition 3.20 and (3.36).

In order to check condition (II) of Proposition C.2, we observe that by stationarity and Corollary 3.16 we have

$$\begin{aligned} \mathbb{E}(N_n(B)) &= \mathbb{E} \left( \sum_{\ell=1}^m \sum_{i=\lceil k_n a_\ell \rceil}^{\lceil k_n b_\ell \rceil - 1} \mathbb{1}_{\{\tilde{\pi}(\mathbb{X}_{n,i}) \in \tilde{A}_\ell\}} \right) \sim \sum_{\ell=1}^m |J_\ell| k_n \mathbb{P}(\tilde{\pi}(\mathbb{X}_{n,i}) \in \tilde{A}_\ell) \\ &\xrightarrow{n \rightarrow \infty} \sum_{\ell=1}^m |J_\ell| \mu(\tilde{A}_\ell) = \mathbb{E}(N(B)). \end{aligned}$$

□

**3.5. Applications to jointly regularly varying sequences.** In statistics of extremes there is a special interest in the shape of the tail of the distributions and, particularly, heavy tails play a significant role. In this setting, the study of the mean and of the extremes is linked and point processes and regularly varying measures have revealed to be very useful tools (see [72], for example). In this heavy tail context, the information regarding clustering is particularly well captured by the tail process introduced in [10] and, in fact, the process  $(Y_j)_{j \in \mathbb{Z}}$  used to define the transformed anchored tail process, in Section 3.3.4, can be identified as a transformed version in  $l_\infty$  of the original tail process, which lives in  $l_0$ .

In the study of rare events for dynamical systems, one is not so interested in different tail behaviours but rather on the statistical properties of the system, which are intimately related with quantitative recurrence properties (see [32]). For this reason the particular homeomorphism  $f$  establishing the shape of  $\tau$  (see Remark 2.9) is not as relevant as proving the existence of a limiting law and therefore we tried to establish the definitions and devices in a more general setting so that they are more amenable for application to dynamical systems. The universality of the uniform distribution we get, for example, in Lemma 3.10 (with no parameter involved as opposed to the  $\alpha$  appearing in the tail process version) and the fact that the limiting point process has a Poisson component with  $\text{Leb} \times \theta \text{Leb}$  intensity measure partly motivated introducing the transformed anchored tail process, which is related to the piles of clustering points obtained in the limiting rare events point processes studied in [30].

In the stationary heavy tail setting, the existence of the tail process is equivalent to joint regular variation of the process  $(\mathbf{X}_i)_{i \in \mathbb{Z}}$  (see [10, Theorem 2.1]), a notion that we define next.

**Definition 3.21.** A  $k$ -dimensional random vector  $\mathbf{X}$  is said to be *jointly regularly varying*, with index  $\alpha > 0$ , if there exists a sequence of constants  $(a_n)_{n \in \mathbb{N}}$  and a random vector  $\Theta$  with  $\mathbb{P}(\|\Theta\| = 1) = 1$ , such that

$$n\mathbb{P}(\|\mathbf{X}\| > xa_n, \mathbf{X}/\|\mathbf{X}\| \in \cdot) \xrightarrow[n \rightarrow \infty]{w} x^{-\alpha} \mathbb{P}(\Theta \in \cdot).$$

where we are considering weak convergence of measures on  $\mathbb{S}^{k-1}$ , the unit sphere in  $\mathbb{R}^k$ . An  $\mathbb{R}^d$ -valued sequence  $\mathbf{X}_0, \mathbf{X}_1, \dots$  is said to be *jointly regularly varying*, with index  $\alpha > 0$ , if all the finite-dimensional vectors  $(\mathbf{X}_k, \dots, \mathbf{X}_\ell)$ ,  $k \leq \ell \in \mathbb{N}_0$ , are jointly regularly varying, with index  $\alpha > 0$ .

In [10, 9, 50], for example, the original stochastic process is assumed to be jointly regularly varying. In what follows, we show that, in this setting, we recover the convergence of the point process of clusters stated in these works and give the particular relation between the tail process and the transformed anchored tail process.

For a jointly regularly varying sequence  $\mathbf{X}_0, \mathbf{X}_1, \dots$ , in particular, condition (2.19) holds. Hence, taking  $\tau = x^{-\alpha}$  and  $u_n(\tau) = a_n \tau^{-\frac{1}{\alpha}}$ , equation (2.15) holds. We also have that  $u_n^{-1}(z) = z^{-\alpha} a_n^\alpha$ .

*Remark 3.22.* Observe that joint regular variation of  $(\mathbf{X}_j)_{j \in \mathbb{Z}}$  implies the existence of the sequence  $(Y_j)_{j \in \mathbb{Z}}$  assumed in (1) (see [10, Theorem 2.1]) and also the independence of the respective polar decomposition assumed in (2) (see [10, Theorem 3.1]). Moreover, If, instead of  $\mathcal{A}'_{q_n}$ , we assume the much stronger conditions considered in [9, Assumption 1.1] and other previous works, going back to [76], one can show that both (3) and (4) also hold. See [10, Proposition 4.2].

In order to make the connection between the tail and the transformed tail processes, we start by defining the map:

$$\begin{aligned} \xi: (\mathbb{R}^d \cup \{\infty\}) \setminus \{0\} &\longrightarrow \mathbb{R}^d \\ x &\mapsto \begin{cases} (\|x\|)^{-\frac{1}{\alpha}} \frac{x}{\|x\|}, & \text{if } x \neq \infty \\ 0, & \text{otherwise} \end{cases}, \end{aligned} \quad (3.37)$$

Then, we define  $\Xi: l_\infty \longrightarrow l_0$  given by  $\Xi((x_j)_j) = (\xi(x_j))_j$ . Observe that, like  $P$  (see Lemma 3.4), the function  $\Xi$  is invertible and we may define  $\tilde{\Xi}: \tilde{l}_\infty \rightarrow \tilde{l}_0$  so that  $\tilde{\Xi}(\tilde{\pi}(\mathbf{x})) = \tilde{\pi}(\Xi(\mathbf{x}))$ . Note that

$$\xi \left( u_n^{-1}(\|\mathbf{X}_j\|) \frac{\mathbf{X}_j}{\|\mathbf{X}_j\|} \right) = \frac{\mathbf{X}_j}{a_n}. \quad (3.38)$$

**Corollary 3.23.** *Let  $\mathbf{X}_0, \mathbf{X}_1, \dots$  be a stationary  $\mathbb{R}^d$ -valued jointly regularly varying sequence, with tail index  $\alpha > 0$ , satisfying conditions  $\mathcal{A}_{q_n}$  and  $\mathcal{A}'_{q_n}$  and for which the transformed anchored tail process given in Definition 3.8 is well defined. Then the point process*

$$N'_n = \sum_{i=1}^{\infty} \delta_{(i/k_n, \tilde{\Xi}(\tilde{\pi}(\mathbb{X}_{n,i})))} = \sum_{i=1}^{\infty} \delta_{\left( \frac{i}{k_n}, \tilde{\pi} \left( \frac{\mathbf{x}_{(i-1)r_n}}{a_n}, \dots, \frac{\mathbf{x}_{ir_n-1}}{a_n} \right) \right)} \quad (3.39)$$

converges weakly in  $\mathcal{N}_{\mathbb{R}_0^+ \times \tilde{l}_0 \setminus \{0\}}^\#$  to the Poisson point process  $N'$  given by

$$N' = \sum_{i=1}^{\infty} \delta_{(T_i, U_i^{-\frac{1}{\alpha}} \tilde{\Xi}(\tilde{\mathbf{Q}}_i))}, \quad (3.40)$$

where  $(T_i)_{i \in \mathbb{N}}$ ,  $(U_i)_{i \in \mathbb{N}}$  and  $(\tilde{\mathbf{Q}}_i)_{i \in \mathbb{N}}$  are as in (3.34).

This corollary of Theorem 3.18 follows from a direct application of the CMT for the map  $\Xi^\# : \mathcal{N}_{\mathbb{R}_0^+ \times \tilde{l}_\infty \setminus \{\infty\}}^\# \rightarrow \mathcal{N}_{\mathbb{R}_0^+ \times \tilde{l}_0 \setminus \{0\}}^\#$  defined by

$$\Xi^\# \left( \sum_{i=1}^{\infty} \delta_{(t_i, \tilde{\mathbf{x}}_i)} \right) = \sum_{i=1}^{\infty} \delta_{(t_i, \tilde{\Xi}(\tilde{\mathbf{x}}_i))}.$$

Observe that the Poisson component of the process  $N'$  can be written as  $M_\alpha = \sum_{i=1}^{\infty} \delta_{(T_i, P_i)} = \sum_{i=1}^{\infty} \delta_{(T_i, U_i^{-1/\alpha})}$ , with intensity measure  $\text{Leb} \times \theta \nu_\alpha$ , with  $\nu_\alpha(y) = d(-y^{-\alpha})$ , while the angular component associated to the tail process, which we denoted earlier by  $(\mathbf{Q}_j)_{j \in \mathbb{Z}}$ , is given by the equation:

$$\mathbf{Q}_j = \xi(\mathbf{Q}_j). \quad (3.41)$$

Recall that as observed in Remark 3.22, the existence of transformed anchored tail process and the respective properties are guaranteed by the joint regular variation of the process  $\mathbf{X}_0, \mathbf{X}_1, \dots$  and [9, Assumption 1.1] (assumption  $\mathcal{AC}(r_n, c_n)$  in [50]).

## 4. PROOFS OF THE DYNAMICAL ENRICHED FUNCTIONAL LIMIT THEOREMS

The major step to obtain the invariance principles stated in Theorems 2.5, 2.10 and 2.12 for dynamically defined stochastic processes is the complete convergence of point processes stated as follows

**Theorem 4.1.** *Let  $T : \mathcal{X} \rightarrow \mathcal{X}$  be a dynamical system as described in Section 2.2 and  $\mathbf{X}_0, \mathbf{X}_1, \dots$  be obtained from such a system as described in (2.1) and assume that the transformed anchored tail process given in Definition 3.8 is well defined. Consider the point process  $N_n$  defined as in (3.33). Then  $N_n$  converges weakly in  $\mathcal{N}_{\mathbb{R}_0^+ \times \tilde{I}_\infty \setminus \{\tilde{\infty}\}}^\#$  to the Poisson point process  $N$  given by (3.34).*

It is sufficient to show that the systems and the observables that we consider give rise to stochastic processes for which  $\mathbb{A}_{q_n}$  and  $\mathbb{A}'_{q_n}$  hold, since then the conclusion follows immediately by Theorem 3.18. A property we use to prove these conditions is in the next definition

**Definition 4.2.** Let  $\mathcal{C}_1$  and  $\mathcal{C}_2$  be Banach spaces of real-valued measurable functions on  $\mathcal{X}$ . Define the *correlation* of non-zero functions  $\phi \in \mathcal{C}_1$  and  $\psi \in \mathcal{C}_2$  with respect to  $\mu$  at time  $n \in \mathbb{N}$  by

$$\text{Cor}_\mu(\phi, \psi, n) := \frac{1}{\|\phi\|_{\mathcal{C}_1} \|\psi\|_{\mathcal{C}_2}} \left| \int \phi(\psi \circ T^n) d\mu - \int \phi d\mu \int \psi d\mu \right|.$$

Then say that the system has *decay of correlations*, with respect to  $\mu$ , for observables in  $\mathcal{C}_1$  against observables in  $\mathcal{C}_2$  if there exists a rate function  $\rho : \mathbb{N} \rightarrow [0, \infty)$  with

$$\lim_{n \rightarrow \infty} \rho(n) = 0,$$

and for every  $\phi \in \mathcal{C}_1$ ,  $\psi \in \mathcal{C}_2$ ,

$$\text{Cor}_\mu(\phi, \psi, n) \leq \rho(n).$$

The uniformly expanding systems in Section 2.2.1 have decay of correlations against  $L^1(\mu)$ , i.e., where  $\mathcal{C}_2 = L^1(\mu)$ . In that setting we also require a suitable space  $\mathcal{C}_1$ , a key feature being that the characteristic functions on our sets of interest, like the annuli  $A^{(j)}$  in (3.5) do not have large norm. In fact in the interval setting  $\mathcal{C}_2$  will be the  $BV$  norm which we recall here. If  $\psi : I \rightarrow \mathbb{R}$  is a measurable function on an interval  $I$  then its *variation* is defined as

$$\text{Var}(\psi) := \sup \left\{ \sum_{i=0}^{n-1} |\psi(x_{i+1}) - \psi(x_i)| \right\},$$

where the supremum is taken over all finite ordered sequences  $(x_i)_{i=0}^{n-1}$  in  $I$ . The  $BV$  norm is  $\|\psi\|_{BV} := \sup |\psi| + \text{Var}(\psi)$  and  $BV := \{\psi : I \rightarrow \mathbb{R} : \|\psi\|_{BV} < \infty\}$ . In Saussol's class of higher-dimensional expanding maps  $\mathcal{C}_1$  is a quasi-Hölder norm.

*Remark 4.3.* While our conditions on the *rate* of decay of correlations here may appear very weak, in fact summable decay of correlations against  $L^1(\mu)$  implies exponential decay of correlations for Hölder observables against  $L^\infty$ , as in [3, Theorem B].

We make a brief list of references where one can find the arguments to prove  $\mathbb{A}_{q_n}$  and  $\mathbb{A}'_{q_n}$  for the systems mentioned in Section 2.2. For non-invertible systems admitting decay of correlations against  $L^1$  and observables with maximal sets  $\mathcal{M}$  consisting on periodic points or a countable number of points in the same orbit, we note that our conditions on the system and the observable can be expressed as requiring, for  $A_{n,\ell}^{(q_n)}$ , as in (3.9) and (3.5),

- (1)  $\lim_{n \rightarrow \infty} \|\mathbb{1}_{A_{n,\ell}^{(q_n)}}\| n \rho(t_n) = 0$  for some sequence  $(t_n)_n$  with  $t_n = o(n)$ ,
- (2)  $\lim_{n \rightarrow \infty} \|\mathbb{1}_{A_{n,\ell}^{(q_n)}}\| \sum_{j=q_n}^n \rho(j) = 0$ .

These conditions can be found in, for example [5, 31]. Along with decay of correlations against  $L^1(\mu)$ , they imply  $\mathbb{A}_{q_n}, \mathbb{A}'_{q_n}$ : for proofs see those references. See also [30, Section 4.3].

Regarding the Benedicks-Carleson maps equipped with observables maximised at periodic points, the required estimates to satisfy  $\mathbb{A}_{q_n}$  can be found in [28] and [34, Sections 5,6].

For Anosov linear diffeomorphisms on the torus and observables maximised at periodic points, we refer to [15].

For the slowly mixing systems of the Manneville-Pomeau type and observables maximised at periodic points distinct from the indifferent fixed point, we could use a direct approach to prove conditions  $\mathbb{A}_{q_n}$  and  $\mathbb{A}'_{q_n}$  using the ideas in [36, Section 4], but since decay of correlations is stated for Hölder continuous functions, an approximation to indicator functions is needed which restricts the domain of application to  $\gamma \in (0, 0.289)$ , whereas one would expect the results to hold at least for  $\gamma < 1/2$  (when one has summable decay of correlations). In order to show that the results hold for observables with a spike at periodic points (or even with a finite number of spikes belonging to the same orbit) for all  $\gamma \in (0, 1/2)$  and even for  $\gamma \in [1/2, 1)$ , if we take the observables vanishing at the origin, in Section 4.4 we introduce a new point process which incorporates the idea of inducing.

**4.1. Proof of Theorem 2.5.** As a consequence of Theorem 4.1, since in this case we are assuming that  $g$  appearing in (2.2) is of type  $g_2$  and condition (2.19) holds, we obtain, by direct application of the CMT for the map  $\Xi^\#$  given in Section 3.5, that the convergence of point processes stated in Corollary 3.23 holds. Therefore, we are left to show that such point process convergence implies the convergence in  $F'$  of the continuous time process  $S_n$  to  $\underline{V}$ .

**Proposition 4.4.** *Let  $\mathbf{X}_0, \mathbf{X}_1, \dots$  be a stationary  $\mathbb{R}^d$ -valued process for which condition (2.19) holds and moreover  $N'_n$  defined in (3.39) converges weakly $^\#$  in  $\mathcal{N}_{\mathbb{R}_0^+ \times \tilde{l}_0 \setminus \{\mathbf{0}\}}^\#$  to  $N'$  given in (3.40) then, under the same assumptions of Theorem 2.5, the conclusion regarding the convergence of  $S_n$  to  $\underline{V}$  holds, in  $F'$ .*

*Proof.* Recall that the convergence in  $F'([0, 1], \mathbb{R}^d)$  consists of showing that the respective projections into  $E([0, 1], \mathbb{R}^d)$  and  $\tilde{D}([0, 1], \mathbb{R}^d)$  converge. The choice of metric in  $E([0, 1], \mathbb{R}^d)$  (see (2.10)) implies that the convergence in this space will follow from the convergence of the coordinate projections, in  $E([0, 1], \mathbb{R})$ , which follows immediately from [9, Theorem 4.5]. Hence, we are left to check the convergence of the  $\tilde{D}([0, 1], \mathbb{R}^d)$  counterparts, which we prove by splitting the argument into the same steps considered in [9, Theorem 4.5] so that we can keep track of the required adjustments.

We start by defining a projection  $\Upsilon: \mathcal{N}_{\mathbb{R}_0^+ \times \tilde{l}_0 \setminus \{\mathbf{0}\}}^\# \rightarrow F'([0, 1], \mathbb{R}^d)$ . Suppose we are given  $\gamma = \sum_{i=1}^\infty \delta_{(t_i, \tilde{x}^i)} \in \mathcal{N}_{\mathbb{R}_0^+ \times \tilde{l}_0 \setminus \{\mathbf{0}\}}^\#$ . At time  $t_i$  we have  $\tilde{x}^i \in \tilde{l}_0$ : let  $(\dots, x_{-1}, x_0, x_1, \dots)$  be a representative of this in  $l_0$  and define

$$\tilde{e}_x^{t_i}(t) = \sum_{i=-\infty}^{\lfloor v(t) \rfloor} x_i, \quad x(t) = \sum_{T_i \leq t} \tilde{e}_x^{T_i}(1), \quad S^x = \{t_i\}_i, \quad e_x^{t_i} = x(t_i^-) + \tilde{e}_x^{t_i}.$$

Note that since  $e_x^{t_i}$  is really an element of an equivalence class, the particular representative of  $\tilde{x}$  chosen does not matter. Finally, let  $\Upsilon(\gamma) = (x, S^x, \{e_x^{t_i}\}_i)$ .

We start assuming  $\alpha \in (0, 1)$ .

**Step 1.** For  $\varepsilon > 0$ , we define an  $\varepsilon$ -truncated projection  $\Upsilon^\varepsilon$ . We do this as follows. For  $\tilde{x} \in \tilde{l}_0$  set

$$u^\varepsilon(\tilde{x})(t) = \sum_{j=-\infty}^{\lfloor v(t) \rfloor} x_j \mathbb{1}_{\{|x_j| > \varepsilon\}}.$$

Then for  $\gamma = \sum_{i=1}^{\infty} \delta_{T_i, \tilde{x}^i} \in \mathcal{N}_{\mathbf{X}}^\#$ , define  $S^\varepsilon = \{T_i : \|\tilde{x}_i\|_\infty > \varepsilon\}$  and

$$\Upsilon^\varepsilon(\gamma) = \left( \left\{ \sum_{T_i \leq t} u^\varepsilon(\tilde{x}^i)(1) \right\}_{t \in [0,1]}, S^\varepsilon, \{u^\varepsilon(\tilde{x}^s)\}_{s \in S^\varepsilon} \right)$$

(here we understand  $\tilde{x}^s$  for  $s = T_i \in S^\varepsilon$  as being  $\tilde{x}^i$ ).

To show that given  $\gamma_n \rightarrow_{w^\#} \gamma$ , we have  $\tilde{\pi}(\Upsilon^\varepsilon(\gamma_n)) \rightarrow \tilde{\pi}(\Upsilon^\varepsilon(\gamma))$ , we first define a simpler space which is not essential here, but is intended to help the reader's visualisation of the situation.

Given an element  $z \in \tilde{D}$ , let  $(t_i)_i$  be the set  $\text{disc}(z)$  of discontinuities of a representative of  $z$ , which we abuse notation and also call  $z$ . Let  $v_i(z) = \sup_{t_i \leq s_1, s_2 < t_{i+1}} \|z(s_1) - z(s_2)\|$  and  $j(t_i) = \|z(t_i^-) - z(t_i)\|$ . Then for  $\varepsilon > 0$  set

$$\tilde{D}_\varepsilon := \{z \in \tilde{D} : \#\text{disc}(z) < \infty, v_i(z) = 0 \text{ and } j_i(z) > \varepsilon \text{ for all } i\}.$$

Now notice that since we are based in  $\mathcal{N}_{\mathbb{R}_0^+ \times \tilde{l}_0 \setminus \{0\}}^\#$ , which only allows a finite number of jumps of norm higher than  $\varepsilon$ , each element of  $\{u^\varepsilon(\tilde{x}^s)\}_{s \in S^\varepsilon}$  lies in  $\tilde{D}_\varepsilon$ . Moreover for  $\underline{x} = \Upsilon^\varepsilon(\gamma)$ ,  $\tilde{\pi}(\underline{x}) = \underline{x}^{\tilde{D}} \in \tilde{D}_\varepsilon$ . We can easily see in this simpler space that if  $\gamma_n \rightarrow_{w^\#} \gamma$  then for  $\underline{x}_n = \Upsilon^\varepsilon(\gamma_n)$  and  $\underline{x} = \Upsilon^\varepsilon(\gamma)$ , we have  $\underline{x}_n^{\tilde{D}} \rightarrow \underline{x}^{\tilde{D}}$ .

These arguments together complete the proof of Step 1.

**Step 2.** The aim here is to show that  $\tilde{\pi}(\Upsilon^\varepsilon(N')) = \underline{x}_\varepsilon^{\tilde{D}} \rightarrow \tilde{\pi}(\Upsilon(N')) = \underline{x}^{\tilde{D}}$  as  $\varepsilon \rightarrow 0$ . It is easy to see that

$$d_{\tilde{D}}(\underline{x}_\varepsilon^{\tilde{D}}, \underline{x}^{\tilde{D}}) \leq \sum_{i=1}^{\infty} \sum_{j \in \mathbb{Z}} U_i^{-\frac{1}{\alpha}} \|\xi(Q_{ij})\| \cdot \mathbb{1}_{\{U_i^{-\frac{1}{\alpha}} \|\xi(Q_{ij})\| \leq \varepsilon\}}$$

And this is shown to converge to zero almost surely because condition (2.22) implies that  $W_i = \sum_{j \in \mathbb{Z}} \|\xi(Q_{ij})\|$  is a.s. finite. Moreover, as noted in [9, Remark 4.6], we have

$$\sup_{i \in \mathbb{N}} \sum_{j \in \mathbb{Z}} U_i^{-\frac{1}{\alpha}} \|\xi(Q_{i,j})\| \mathbb{1}_{U_i^{-\frac{1}{\alpha}} \|\xi(Q_{i,j})\| \leq \varepsilon} \rightarrow 0, \quad \text{a.s. as } \varepsilon \rightarrow 0.$$

Note that condition (2.22), which is assumed by hypothesis for  $1 < \alpha < 2$ , holds for  $0 < \alpha < 1$ , as a byproduct of the convergence of the point processes stated in Theorem 3.18 as observed in [20, Theorem 2.6]. For  $\alpha = 1$ , a similar argument holds by making use of assumption (2.23).

**Step 3.** This step looks to compare the projection into  $\tilde{D}$  of empirical process  $\Upsilon(N'_n)$  with its  $\varepsilon$ -truncated version  $\Upsilon(N'_n)$ . Namely, we set  $\underline{x}_{n,\varepsilon}^{\tilde{D}} := \tilde{\pi}(\Upsilon^\varepsilon(N'_n))$ ,  $\underline{x}_n^{\tilde{D}} := \tilde{\pi}(\Upsilon(N'_n))$  and observe that

$$d_{\tilde{D}}(\underline{x}_{n,\varepsilon}^{\tilde{D}}, \underline{x}_n^{\tilde{D}}) \leq \sum_{j=1}^{k_n r_n} \frac{\|\mathbf{X}_j\|}{a_n} \mathbb{1}_{\{\|\mathbf{X}_j\| \leq a_n \varepsilon\}}.$$

Then, as in [9, Proof of Theorem 4.5], it follows by Markov's inequality and Karamata's Theorem that

$$\begin{aligned} \limsup_{n \rightarrow \infty} \mathbb{P}(d_{\tilde{D}}(\underline{x}_{n,\varepsilon}^{\tilde{D}}, \underline{x}_n^{\tilde{D}}) > \delta) &\leq \limsup_{n \rightarrow \infty} \frac{k_n r_n}{\delta a_n} \mathbb{E}(\|\mathbf{X}_1\| \mathbb{1}_{\{\|\mathbf{X}_1\| \leq a_n \varepsilon\}}) \\ &\leq \lim_{n \rightarrow \infty} \frac{n}{\delta a_n} \frac{\alpha a_n \varepsilon \mathbb{P}(\|\mathbf{X}_1\| > a_n \varepsilon)}{1 - \alpha} = \frac{\alpha}{\delta(1 - \alpha)} \varepsilon^{1-\alpha} \xrightarrow{\varepsilon \rightarrow 0} 0, \end{aligned}$$

since we are assuming  $\alpha < 1$ .

**Step 4.** Here we wish to compare the  $\tilde{D}$  component of the embedding of  $S_n$  into  $F'$  with the respective  $\tilde{D}$  projection of  $\Upsilon(N'_n)$ . But we need only notice that the representations of each of these elements is in fact identical so there is nothing to do: in the  $\Upsilon(N'_n)$  case all the jumps in a block are compressed into a single sequence, but the  $\tilde{D}$  component undoes this compression, capturing all the jumps as an element of  $D([0, 1])$ , which is equivalent to that obtained from  $S_n$ .

The proof of the  $1 \leq \alpha < 2$  case does not require further arguments, though we do need to employ condition (2.21) in order to bound the error term in the equivalent to step 3 above.  $\square$

**4.2. Proof of Theorem 2.10.** We split the proof of Theorem 2.10 into several steps given in a series of lemmas.

We start by defining a map  $\Upsilon: \mathcal{N}_{\mathbb{R}_0^+ \times \tilde{l}_\infty \setminus \{\infty\}}^\# \rightarrow F'((0, \infty), \mathbb{R})$ , which to each  $\gamma \in \mathcal{N}_{\mathbb{R}_0^+ \times \tilde{l}_\infty \setminus \{\infty\}}^\#$ , such that  $\gamma = \sum_{i=1}^\infty \delta_{(t_i, \tilde{\mathbf{x}}^i)}$ , assigns

$$\Upsilon(\gamma) = (h(\gamma), \{t_i: i \in \mathbb{N}\}, e_\gamma^{t_i}),$$

where  $h(\gamma)$  is defined for all  $t > 0$  by

$$h(\gamma)(t) := \begin{cases} \inf\{\|\tilde{P}(\tilde{\mathbf{x}}^i)\|_\infty^{-1}: t_i \leq t\} & \text{if } t > \underline{t} \\ \bar{y} & \text{if } t \leq \underline{t} \end{cases} \quad (4.1)$$

with  $\underline{t} = \inf\{(t_i)_{i=1}^\infty\}$ ,  $\bar{y} = \sup\{\inf\{\|\tilde{P}(\tilde{\mathbf{x}}^i)\|_\infty^{-1}: t_i \leq t\}: t > \underline{t}\}$  and, for  $t \in [0, 1]$

$$e_\gamma^{t_i}(t) = \min\{h(\gamma)(t_i^-), \inf_{j \leq \lfloor v(t) \rfloor} \|x_j^i\|\}.$$

Note that since we are using  $P$  here, we are looking for our observations to go *down* in value, so the objects here are all non-increasing functions.

Let  $\Lambda = \left\{ \gamma = \sum_{i=1}^\infty \delta_{(t_i, \tilde{\mathbf{x}}^i)}: t_i \neq t_j, \forall i \neq j; \gamma((0, a) \times \tilde{l}_\infty \setminus \{\infty\}) > 0, \forall a > 0 \right\}$ .

**Lemma 4.5.** *The map  $\Upsilon: \mathcal{N}_{\mathbb{R}_0^+ \times \tilde{l}_\infty \setminus \{\infty\}}^\# \rightarrow F'(\mathbb{R}^+, \mathbb{R})$  is continuous on  $\Lambda$ .*

*Proof.* Assume that  $\gamma_n \rightarrow_{w^\#} \gamma = \sum_{i=1}^\infty \delta_{(t_i, \tilde{\mathbf{x}}^i)} \in \Lambda$ . Let  $0 < a < b \in \{t_i: i \in \mathbb{N}\}^c$ , we need to verify that

$$m(\Gamma_{a,b,\underline{x}_n^E}, \Gamma_{a,b,\underline{x}^E}) + d_{\tilde{D}}(\underline{x}_{na,b}^{\tilde{D}}, \underline{x}_{a,b}^{\tilde{D}}) \rightarrow 0,$$

where  $\Gamma_{a,b,\underline{x}^E}$  and  $\underline{x}_{a,b}^{\tilde{D}}$  denote the restrictions to the time interval  $[a, b]$  of the respective objects. Let  $\alpha = h(\gamma)(a) \in (0, \infty)$  and consider the bounded set  $\tilde{B}_\alpha = \{\tilde{\mathbf{x}} \in \tilde{l}_\infty: \|\tilde{P}(\tilde{\mathbf{x}})\|_\infty^{-1} < \alpha\}$  of  $\tilde{l}_\infty \setminus \{\infty\}$ . There are finitely many  $t_i$  such that  $(t_i, \tilde{\mathbf{x}}^i) \in [a, b] \times \tilde{B}$ . Let  $(t_{i_1}, \tilde{\mathbf{x}}^{i_1}), \dots, (t_{i_k}, \tilde{\mathbf{x}}^{i_k})$  be an enumeration of those points. Then, as in [72, Proposition 3.13], one can show that for  $n$  sufficiently large there exists a shadow enumeration  $(t_{i_1}^{(n)}, \tilde{\mathbf{x}}^{(n), i_1}), \dots, (t_{i_k}^{(n)}, \tilde{\mathbf{x}}^{(n), i_k})$  of the mass points of  $\gamma_n$  in  $[a, b] \times \tilde{B}_\alpha$  so that

$$t_{i_j}^{(n)} \rightarrow t_{i_j} \quad \text{and} \quad \tilde{\mathbf{x}}^{(n), i_j} \rightarrow \tilde{\mathbf{x}}^{i_j}, \quad \text{as } n \rightarrow \infty, \text{ for all } j = 1, \dots, k \quad (4.2)$$

Let  $t_i$  be such that  $h(\gamma)(a) = \|\tilde{P}(\tilde{\mathbf{x}}^i)\|_\infty^{-1}$ . Then applying the same argument to the bounded set  $[t_i - \epsilon, a] \times B_{\alpha+\epsilon}$  where  $\epsilon$  is chosen so that no mass point of  $\gamma$  lies on the border of this set, one obtains that  $h(\gamma_n)(a) \rightarrow h(\gamma)(a)$ , as  $n \rightarrow \infty$ . It follows that

$$h(\gamma_n)(t) \xrightarrow{n \rightarrow \infty} h(\gamma)(t), \quad \forall t \in [a, b] \setminus \left( \bigcup_{n \in \mathbb{N}} \{t_{i_1}^{(n)}, \dots, t_{i_k}^{(n)}\} \cup \{t_{i_1}, \dots, t_{i_k}\} \right). \quad (4.3)$$

The convergence stated in (4.2) regarding the quotient space  $\tilde{l}_\infty \setminus \{\tilde{\infty}\}$  means that we can take representatives for which the convergence is obtained componentwise, which allows us to choose corresponding representatives on the space of excursions  $\tilde{D}([0, 1], \mathbb{R})$  for which the following holds

$$e_{\gamma_n}^{t_i^{(n)}}(t) \rightarrow e_{\gamma}^{t_i}(t) \quad \text{as } n \rightarrow \infty, \text{ for all } t \in [0, 1], \text{ which, in turn, implies} \quad (4.4)$$

$$\inf \left\{ e_{\gamma_n}^{t_i^{(n)}}(t) : t \in [0, 1] \right\} \xrightarrow{n \rightarrow \infty} \inf \{ e_{\gamma}^{t_i}(t) : t \in [0, 1] \}, \quad (4.5)$$

From (4.2), (4.5) and (4.3), it follows that  $m(\Gamma_{a,b,\underline{x}_n^E}, \Gamma_{a,b,\underline{x}^E}) \xrightarrow{n \rightarrow \infty} 0$ . In a similar way, (4.2), (4.4) and (4.3) imply that  $d_{\tilde{D}}(\underline{x}_{n,a,b}^{\tilde{D}}, \underline{x}_{a,b}^{\tilde{D}}) \xrightarrow{n \rightarrow \infty} 0$ .  $\square$

**Lemma 4.6.** *We have that  $d_{F',\infty}(Z_n, \Upsilon(N_n)) \rightarrow 0$  as  $n \rightarrow \infty$ , in probability, where  $N_n$  is defined in (3.33)*

*Proof.* The processes  $Z_n$  and  $\Upsilon(N_n)$  have different internal clocks, but  $h(\Upsilon(N_n))(i/k_n) = Z_n(ir_n/n)$ , for all  $i \in \mathbb{N}_0$ . Observe that  $h(\Upsilon(N_n))$  is constant between  $i/k_n$  and  $(i+1)/k_n$ , having a possible jump at  $(i+1)/k_n$  if  $\inf\{u_n^{-1}(\mathbf{X}_j) : j = ir_n/n, \dots, (i+1)r_n/n - 1\} < h(\Upsilon(N_n))(i/k_n)$ . Note that when such a jump occurs,  $Z_n$  is not necessarily constant between  $ir_n/n$  and  $(i+1)r_n/n$ , namely, it may jump several times corresponding to the several moments  $j = ir_n/n, \dots, (i+1)r_n/n$  at which  $u_n^{-1}(\mathbf{X}_j)$  beats the running minimum until the threshold  $\inf\{u_n^{-1}(\mathbf{X}_j) : j = ir_n/n, \dots, (i+1)r_n/n - 1\}$  is reached. Nevertheless, these oscillations are recorded by the excursion decorating  $\Upsilon(N_n)$  at the discontinuity point  $(i+1)/k_n$ .

Note that by definition of  $r_n$ ,

$$\frac{i}{k_n} \geq \frac{ir_n}{n} \geq \frac{i}{k_n} - \frac{i}{n} \quad \Rightarrow \quad \frac{i}{n} \geq \frac{i}{k_n} - \frac{ir_n}{n} \geq 0 \quad (4.6)$$

Hence, the maximum distance between one instant point of the time interval  $[i/k_n, (i+1)/k_n]$  and another from  $[ir_n/n, (i+1)r_n/n]$  is at most  $\frac{i+1}{n}$ .

Let  $0 < a < b \in \mathbb{R}$ ,  $i^- = \min\{j \in \mathbb{N}_0 : j/k_n \geq a\}$  and  $i^+ = \max\{j \in \mathbb{N}_0 : jr_n/n < b\}$ . Note that the range of the graphs  $\Gamma_{i^-jr_n/n, i^+jr_n/n, Z_n^E}$  and  $\Gamma_{i^-jr_n/n, i^+jr_n/n, \Upsilon(N_n)^E}$  is exactly the same, which means that the Hausdorff distance between the graphs is given by the time component, i.e.,

$$m\left(\Gamma_{i^-jr_n/n, i^+jr_n/n, Z_n^E}, \Gamma_{i^-jr_n/n, i^+jr_n/n, \Upsilon(N_n)^E}\right) \leq \frac{i+1}{n} \xrightarrow{n \rightarrow \infty} 0.$$

We only need to worry with the observations corresponding to  $j \in \mathbb{N}_0$  such that  $a \leq j/n < (i+1)/k_n$  where the range of the process  $Z_n$  may differ from the value of  $h(\Upsilon(N_n))(a) =: \tau$ , by either being above or below. Note that there are at most  $k_n$  such observations. This means that for  $|\epsilon| < \tau$ , we have that  $m\left(\Gamma_{a, i^-/k_n, Z_n^E}, \Gamma_{a, i^-/k_n, \Upsilon(N_n)^E}\right) > |\epsilon|$  implies that there must be at least one of these possible  $k_n$  observations such that  $\|\mathbf{X}_j\| > u_n(\tau + \epsilon)$ . Since  $k_n \mathbb{P}(\|\mathbf{X}_j\| > u_n(\tau + \epsilon)) \rightarrow 0$ , we obtain that  $m\left(\Gamma_{a, i^-/k_n, Z_n^E}, \Gamma_{a, i^-/k_n, \Upsilon(N_n)^E}\right)$  goes to 0 in probability. A similar argument applies for the time frame between  $i^+jr_n/n$  and  $b$ .

Hence,  $m\left(\Gamma_{a,b,Z_n^E}, \Gamma_{a,b,\Upsilon(N_n)^E}\right) \xrightarrow{n \rightarrow \infty} 0$ , in probability. This takes care of the  $E$  component of the distance in  $F'$  and we need now to conclude that the same applies to the projections into  $\tilde{D}$ .

Let  $i^- < i \leq i^+$  and assume that there is a jump of  $h(\Upsilon(N_n))$  at  $i/k_n$ . Recall that while  $h(\Upsilon(N_n))$  is constant between  $(i-1)/k_n$  and  $i/k_n$ , the process  $Z_n$  may have oscillations in the corresponding interval  $[(i-1)r_n/n, ir_n/n]$ . But, by construction, the excursion  $e_{\Upsilon(N_n)}^{i/k_n}$  exactly mimics these oscillations of  $Z_n$ . Observe now that the projection  $\Upsilon(N_n)^{\tilde{D}}$  reconstructs the behaviour of  $Z_n$  by incorporating in the time frame the excursions. Then, since the metric  $\tilde{D}$  allows for time deformations, it is clear that the distance between the projections  $\Upsilon(N_n)^{\tilde{D}}$  and  $Z_n^{\tilde{D}}$ , when restricted to  $[i^-r_n/n, ir_n/n]$  is actually equal to 0. Again, we are left to analyse the time periods  $[a, i^-r_n/n]$  and  $[i^+r_n/n, b]$ , where some pieces of information may be missing. However, as observed with the projections into  $E$ , the missing information corresponds to at most  $\max\{k_n, r_n\}$  random variables and the probability of these producing oscillations that will not be mimicked is bounded by  $\max\{k_n, r_n\}\mathbb{P}(\|\mathbf{X}_j\| > u_n(\tau + \epsilon)) \rightarrow 0$ . Again, we conclude that  $d_{\tilde{D}}\left(\Upsilon(N_n)_{a,b}^{\tilde{D}}, Z_{n,a,b}^{\tilde{D}}\right) \xrightarrow{n \rightarrow \infty} 0$ , in probability.  $\square$

*Proof of Theorem 2.10.* By Lemma 4.5, we may apply the CMT and conclude from Theorem 4.1 that  $\Upsilon(N_n)$  converges weakly to  $\Upsilon(N)$ , in  $F'((0, +\infty))$ . By Lemma 4.6 and a Slutsky argument we conclude that  $Z_n$  converges weakly to  $\Upsilon(N)$ , in  $F'((0, +\infty))$ . Therefore, we are only left to show that  $h(\Upsilon(N))$  has the prescribed finite-dimensional distributions of  $Z_H$ .

For each  $\alpha > 0$ , let  $\tilde{B}_\alpha = \{\tilde{\mathbf{x}} \in: \|\tilde{P}(\tilde{\mathbf{x}})\|_\infty^{-1} < \alpha\} \subset \tilde{l}_\infty \setminus \{\tilde{\infty}\}$ . For the unidimensional distribution, with  $t, y \geq 0$  and  $\bar{H}(\tau) = e^{-\theta\tau}$ , we have

$$\mathbb{P}(h(\Upsilon(N))(t) \geq y) = \mathbb{P}(N([0, t] \times \tilde{B}_y) = 0) = e^{-\theta ty} = \bar{H}^t(y).$$

For the bidimensional distribution, with  $0 \leq t_1 < t_2$  and  $y_1 \geq y_2 \geq 0$ ,

$$\begin{aligned} \mathbb{P}(h(\Upsilon(N))(t_1) \geq y_1, h(\Upsilon(N))(t_2) \geq y_2) &= \\ &= \mathbb{P}(N([0, t_1] \times B_{y_1}) = 0, N((t_1, t_2] \times B_{y_2}) = 0) = \bar{H}^{t_1}(y_1) \bar{H}^{t_2-t_1}(y_2). \end{aligned}$$

In case  $0 \leq y_1 < y_2$ ,

$$\mathbb{P}(h(\Upsilon(N))(t_1) \geq y_1, h(\Upsilon(N))(t_2) \geq y_2) = \mathbb{P}(h(\Upsilon(N))(t_2) \geq y_2) = \bar{H}^{t_2}(y_2),$$

so in general

$$\mathbb{P}(h(\Upsilon(N))(t_1) \geq y_1, h(\Upsilon(N))(t_2) \geq y_2) = \bar{H}^{t_1}(y_1 \vee y_2) \bar{H}^{t_2-t_1}(y_2).$$

By induction we get for the  $k$ -dimensional distribution

$$\begin{aligned} \mathbb{P}(h(\Upsilon(N))(t_1) \geq y_1, h(\Upsilon(N))(t_2) \geq y_2, \dots, h(\Upsilon(N))(t_k) \geq y_k) &= \\ &= \bar{H}^{t_1} \left( \bigvee_{i=1}^k \{y_i\} \right) \bar{H}^{t_2-t_1} \left( \bigvee_{i=2}^k \{y_i\} \right) \dots \bar{H}^{t_k-t_{k-1}}(y_k) \\ &= \mathbb{P}(Z_H(t_1) \geq y_1, \dots, Z_H(t_k) \geq y_k). \end{aligned}$$

Regarding the statements about the Markov structure of the process  $Z_H$ , we give brief indications how to prove them and refer to [72, Chapter 4.1] for some useful properties of the theory of Markov processes. The fact that  $Z_H$  is Markov with the given transition probability follows from the form of the finite-dimensional distributions. Namely, for  $0 < y < z$ ,

$$\begin{aligned} \mathbb{P}(Z_H(t+s) \geq y \mid Z_H(s) = z) &= \mathbb{P}(h(\Upsilon(N))(t+s) \geq y \mid h(\Upsilon(N))(s) = z) \\ &= \mathbb{P}(N((s, s+t] \times B_y) = 0) = e^{-\theta ty}. \end{aligned} \tag{4.7}$$



The exponential holding time parameter at state  $z$ , which is denoted by  $\lambda(z)$ , can be derived from

$$\mathbb{P}(Z_H(t+s) = Z_H(s) \mid Z_H(s) = z) = \mathbb{P}(N((s, s+t] \times B_z) = 0) = e^{-\theta tz}.$$

Thus,  $\lambda(z) = \theta z$ . The jump distribution  $\Pi(z, [0, y))$  can be computed from

$$\lim_{t \rightarrow 0} t^{-1} \mathbb{P}(Z_H(t+s) < y \mid Z_H(s) = z) = \lambda(z) \Pi(z, [0, y)).$$

But from (4.7), for  $y < z$ , we have  $t^{-1} \mathbb{P}(Z_H(t+s) < y \mid Z_H(s) = z) = t^{-1} (1 - e^{-\theta ty}) \xrightarrow[t \rightarrow 0]{} \theta y$  and therefore  $\Pi(z, [0, y)) = y/z$ .  $\square$

**4.3. Proof of Theorem 2.12.** In order to count the number of records during the clusters we define, for  $y \geq 0$  and  $\mathbf{x} = (x_j)_j \in l_\infty$ ,

$$R^{\mathbf{x}}(y) = \sum_{j \in \mathbb{Z}} \mathbb{1}_{\{\|x_j\| < (y \wedge \inf_{i < j} \|x_i\|)\}}, \quad (4.8)$$

which gives the number of record asymptotic frequencies corresponding to the smallest observations in  $\mathbf{x}$  that have beaten the frequency  $y$ , which must be finite because  $\lim_{|j| \rightarrow \infty} \|x_j\| = \infty$ . For  $\gamma = \sum_{i=1}^{\infty} \delta_{(t_i, \tilde{\mathbf{x}}^i)} \in \mathcal{N}_{\mathbb{R}_0^+ \times \tilde{l}_\infty \setminus \{\infty\}}^\#$ , let  $h(\gamma) \in D((0, \infty), \mathbb{R})$  be defined as in (4.1) and define the record point process in  $\mathcal{N}_{(0, +\infty)}^\#$  by

$$R_\gamma = \sum_{i \in \mathbb{N}} \delta_{t_i} R^{\mathbf{x}^i}(h(\gamma)(t_i^-)), \quad \text{where } \mathbf{x}^i \in l_\infty \text{ is any chosen representative of } \tilde{\mathbf{x}}^i.$$

In order to be able to relate to the actual count of the number of records at finite time we consider the empirical process:

$$N_n'' := \sum_{i=0}^{\infty} \delta_{\left(\frac{i}{n}, u_n^{-1}(\|\mathbf{X}_i\|) \frac{\mathbf{x}_i}{\|\mathbf{x}_i\|}\right)},$$

which we consider as defined in  $\mathcal{N}_{\mathbb{R}_0^+ \times \tilde{l}_\infty \setminus \{\infty\}}^\#$  by embedding the second coordinate in  $\mathbb{R}^d$  into  $l_\infty$  by adding a sequence of  $\infty$  before and after that entrance as in (3.19). Now, observe that  $R_{N_n''}$  coincides with  $\mathfrak{R}_n$  given in (2.32) and indeed counts the number of records of the process  $\mathbf{X}_0, \mathbf{X}_1, \dots$

Consider the subset of  $\mathcal{N}_{\mathbb{R}_0^+ \times \tilde{l}_\infty \setminus \{\infty\}}^\#$  defined by

$$\Lambda = \left\{ \gamma = \sum_{i=1}^{\infty} \delta_{(t_i, \tilde{\mathbf{x}}^i)} : t_j \neq t_\ell \text{ and } \|x_j^i\| = \|x_\ell^i\| \Rightarrow j = \ell \text{ or } \|x_j^i\| = \infty \right\}.$$

Since  $(0, \infty)$  is locally compact and separable, then the weak<sup>#</sup> topology in  $\mathcal{N}_{(0, +\infty)}^\#$  coincides with the vague topology and by a trivial adjustment of [9, Lemma 5.1] one obtains that the map  $\gamma \rightarrow R_\gamma$  from  $\mathcal{N}_{\mathbb{R}_0^+ \times \tilde{l}_\infty \setminus \{\infty\}}^\#$  to  $\mathcal{N}_{(0, +\infty)}^\#$  is continuous at every  $\gamma \in \Lambda$ .

**Proposition 4.7.** *Let  $\mathbf{X}_0, \mathbf{X}_1, \dots$  be a stationary  $\mathbb{R}^d$ -valued sequence, with tail index  $\alpha > 0$ , for which the point process  $N_n$  defined in (3.33) converges weakly in  $\mathcal{N}_{\mathbb{R}_0^+ \times \tilde{l}_\infty \setminus \{\infty\}}^\#$  to the Poisson point process  $N$  given by (3.34). Then, under the same assumptions of Theorem 2.12, then the conclusion regarding the convergence of  $R_{N_n''}$  to  $R_N$  holds, in  $\mathcal{N}_{(0, +\infty)}^\#$ , and the limit process has the representation given there.*

Together with Theorem 3.18, this concludes the proof of Theorem 2.12.

We elaborate a bit further on the representation of the limit record point process, which in this case is given in a more natural way through the use of the variable  $U$  with a uniform

distribution. Namely, the limiting process  $R_N$  can be written as  $\mathfrak{R}$ , in (2.33), where the  $\kappa_i$  have the same distribution as the integer-valued random variable  $R^{\mathbf{Q}}(U^{-1})$ , defined as in (4.8), where  $U$  is a uniformly distributed random variable independent of  $\mathbf{Q} = (Q_j)_j$ . The rest of the proof of this proposition follows from trivial adjustments to the proof of [9, Theorem 5.2].

Let  $M = \sum_{i=1}^{\infty} \delta_{(T_i, U_i)}$  be the Poisson point process on  $\mathbb{R}_0^+ \times \mathbb{R}_0^+$  defining the point process  $N$  in (3.34). Recall that  $\|Q_j\| \geq 1$  for all  $j$ , but there exists one index  $j$  for which the equality holds. Let  $(\tau_n)_{n \in \mathbb{Z}}$  denote the ordered sequence of record times of  $M$  counted by  $R_M$ . For definiteness, fix a certain  $s > 0$  and assume that  $\tau_1$  is the first record time larger than  $s$ , *i.e.*,  $\tau_1 = \inf\{\tau_j : \tau_j > s\}$ , and now denote the respective record frequencies by  $Y_n = \inf_{T_i \leq \tau_n} U_i$ , with  $Y_0 = \inf_{T_i \leq s} U_i$ . The fact that  $\sum_{n \in \mathbb{Z}} \delta_{\tau_n}$  is a Poisson process with intensity  $x^{-1} dx$  follows for example from [30, Theorem 5.7]. Note that for all  $n$  we have  $Y_n/Y_{n-1} < 1$  and, in fact, from the jump distribution  $\Pi$  computed in Theorem 2.10 we obtain that  $(Y_n/Y_{n-1})_{n \in \mathbb{N}}$  is a sequence of independent and uniformly distributed random variables.

The number of records of  $R_N$  observed at  $\tau_n = T_i$ , say, corresponds to the number of  $j$ 's for which  $Y_n \|Q_{i,j}\| < (Y_{n-1} \wedge \inf_{\ell < j} \{Y_n \|Q_{i,\ell}\| \}) \Leftrightarrow \|Q_{i,j}\| < (Y_{n-1}/Y_n \wedge \inf_{\ell < j} \{\|Q_{i,\ell}\|\})$ , *i.e.*, is equal to  $R^{\mathbf{Q}_i}(Y_{n-1}/Y_n)$ , where  $\mathbf{Q}_i = (Q_{i,j})_j$  is independent of the sequence  $(Y_n/Y_{n-1})_{n \in \mathbb{N}}$  for all  $i$ . Since  $s$  was arbitrary, the given representation of  $R_N$  holds.

**4.4. Point processes encompassing inducing.** Inducing is a very powerful technique used to study statistical properties of non-uniformly hyperbolic systems (see [2], for example). Let  $T: \mathcal{X} \rightarrow \mathcal{X}$  be a dynamical system defined on a Lebesgue space  $(\mathcal{X}, \mathcal{B}_{\mathcal{X}}, m)$ . Let  $\Delta_0 \in \mathcal{B}_{\mathcal{X}}$  be a base set such that  $m(\Delta_0) > 0$ . Let  $\mathcal{P}$  be a countable mod 0 partition of  $\Delta_0$ , such that, for all  $\omega \in \mathcal{P}$ ,  $T|_{\omega}$  is invertible and there exists a return time function  $R: \Delta_0 \rightarrow \mathbb{N}$ , constant in each element of  $\mathcal{P}$  and such that  $T^{R(\omega)}(\omega)$  is a union elements of  $\mathcal{P}$ . We define the induced system  $\bar{T} = T^R: \Delta_0 \rightarrow \Delta_0$  by  $\bar{T}(x) = T^{R(x)}(x)$ . It is well known that if  $\mu_0$  is a  $\bar{T}$ -invariant probability measure, then, assuming that  $R \in L^1(\mu_0)$  (which we do throughout), its saturation defined by  $\mu(A) = \sum_{j=0}^{\infty} \mu_0(\{R > j\} \cap T^{-j}(A))$  is  $T$ -invariant. Note that  $R$  is not necessarily the first return to  $\Delta_0$ . We consider a measurable observable function  $\varphi: \mathcal{X} \rightarrow \mathbb{R}$  and consider the potentials  $\bar{\varphi}, \bar{\Phi}: \Delta_0 \rightarrow \mathbb{R}$  defined by

$$\bar{\varphi}(x) = \max_{j=0, \dots, R(x)-1} |\varphi(T^j(x))| \quad \bar{\Phi}(x) = \sum_{j=0, \dots, R(x)-1} |\varphi(T^j(x))|.$$

We start by considering the induced stochastic process  $X_0, X_1, \dots$  such that  $X_j = \bar{\varphi} \circ \bar{T}^j$ . Assume the existence of normalising levels  $(u_n(\tau))_{n \in \mathbb{N}}$  as in (2.15), *i.e.*, such that  $\lim_{n \rightarrow \infty} n\mu_0(\bar{\varphi} > u_n(\tau)) = \tau$ , for all  $\tau > 0$ . Then, for each  $n \in \mathbb{N}$ , define  $F_n: \mathcal{X} \rightarrow \mathbb{R}$  as

$$F_n(x) = u_n^{-1}(|\varphi(x)|) \frac{\varphi(x)}{|\varphi(x)|}.$$

Let  $d = 1$  and define  $l_{\infty}, l_0, \tilde{l}_{\infty}, \tilde{l}_0$  as above. For  $\mathbf{x} \in \tilde{l}_0$ , define  $|\tilde{\mathbf{x}}|'_{\infty} := \|\mathbf{x}\|_{\infty}$ , where  $\mathbf{x} \in l_0$  is such that  $\tilde{\mathbf{x}} = \tilde{\pi}(\mathbf{x})$ . For  $\tilde{\mathbf{x}} \in \tilde{l}_{\infty}$ , set  $|\tilde{\mathbf{x}}|_{\infty} := |\tilde{P}(\tilde{\mathbf{x}})|'_{\infty}$ . Define the product space

$$l_{\infty}^* = \left\{ \underline{\mathbf{x}} = (\tilde{\mathbf{x}}_j)_j \in \tilde{l}_{\infty}^{\mathbb{Z}} : \lim_{|j| \rightarrow \infty} |\tilde{\mathbf{x}}_j|_{\infty} = 0 \right\} \quad (4.9)$$

For  $\underline{\mathbf{x}} \in l_{\infty}^*$ , we define

$$|\underline{\mathbf{x}}|^* := \sup_{j \in \mathbb{Z}} |\tilde{\mathbf{x}}_j|_{\infty}.$$

and the complete metric defined on  $l_\infty^* \setminus \{\underline{\infty}\}$  ( $\underline{\infty}$  denotes the only element of  $l_\infty^*$  such that  $|\underline{\infty}|^* = 0$ ) given by:

$$d^*(\underline{\mathbf{x}}, \underline{\mathbf{y}}) = \left( \sup_{j \in \mathbb{Z}} \tilde{d}(\tilde{\mathbf{x}}_j, \tilde{\mathbf{y}}_j) \wedge 1 \right) \vee \left| \frac{1}{|\underline{\mathbf{x}}|^*} - \frac{1}{|\underline{\mathbf{y}}|^*} \right|.$$

We define the quotient spaces  $\tilde{l}_\infty^*$  and the respective metric  $\tilde{d}^*$ , accordingly. We embed the finite product space  $\tilde{l}_\infty^{*n}$  into the infinite product space  $l_\infty^*$  simply by adding a sequence of  $\tilde{\infty}$  before and after the  $n$  entrances of the elements of  $\tilde{l}_\infty^{*n}$  and write  $\tilde{\pi}^*((\tilde{\mathbf{x}}_1, \dots, \tilde{\mathbf{x}}_n)) \in \tilde{l}_\infty^*$  for the projection of the natural embedding of  $(\tilde{\mathbf{x}}_1, \dots, \tilde{\mathbf{x}}_n)$  into  $l_\infty^*$  to the quotient space  $\tilde{l}_\infty^*$ .

We define now the new point process  $N_n^*$  as a random element of  $\mathcal{N}_{\mathbb{R}_0^+ \times \tilde{l}_\infty^* \setminus \{\underline{\infty}\}}^\#$  of boundedly finite point measures on  $\mathbb{R}_0^+ \times \tilde{l}_\infty^* \setminus \{\underline{\infty}\}$ . For  $j \in \mathbb{N}_0$ , we consider

$$\tilde{\mathbf{X}}_j = \tilde{\pi} \left( F_n(\bar{T}^j(x)), F_n(T(\bar{T}^j(x))), \dots, F_n(T^{R(\bar{T}^j(x))-1}(\bar{T}^j(x))) \right) \in \tilde{l}_\infty^*$$

and define

$$N_n^* = \sum_{i=1}^{\infty} \delta_{(i/k_n, \tilde{\pi}^*(\tilde{\mathbf{X}}_{(i-1)r_n}, \dots, \tilde{\mathbf{X}}_{ir_n-1}))}. \quad (4.10)$$

Note that these point processes have in the time direction the information gathered by the induced dynamics  $\bar{T}$  (as with  $N_n$  defined in (3.33)) and, at each recorded time event,  $\tilde{\mathbf{X}}_j$  includes the information regarding the excursion performed during the induced time by the original dynamics  $T$ . In order to obtain the convergence of these point processes, we need to adapt the previous conditions and definitions, which assumed that  $\mathcal{V} = \mathbb{R}^d$  to the present situation where  $\mathcal{V} = \tilde{l}_\infty^*$ . In particular, for the definition of the transformed anchored tail process, we assume the existence of a process  $(\tilde{\mathbf{Y}}_j)_{j \in \mathbb{Z}} \in l_\infty^*$  satisfying the following assumptions:

$$\mathcal{L} \left( \frac{1}{\tau} (\tilde{\mathbf{X}}_{r_n+s}, \dots, \tilde{\mathbf{X}}_{r_n+t}) \mid |\tilde{\mathbf{X}}_{r_n}|_\infty > \tau^{-1} \right) \xrightarrow{n \rightarrow \infty} \mathcal{L} \left( (\tilde{\mathbf{Y}}_j)_{j=s, \dots, t} \right),$$

for all  $s < t \in \mathbb{Z}$  and all  $\tau > 0$ . Note that

$$\begin{aligned} |\tilde{\mathbf{X}}_{r_n}|_\infty > \tau^{-1} &\Leftrightarrow \max_{j=0, \dots, R(\bar{T}^{r_n}(x))-1} 1/u_n^{-1}(|\varphi(T^j(\bar{T}^{r_n}(x)))|) > \tau^{-1} \\ &\Leftrightarrow \min_{j=0, \dots, R(\bar{T}^{r_n}(x))-1} u_n^{-1}(|\varphi(T^j(\bar{T}^{r_n}(x)))|) < \tau \Leftrightarrow \bar{\varphi}(\bar{T}^{r_n}(x)) > u_n(\tau). \end{aligned}$$

The transformed anchored tail process is then defined by

$$\mathcal{L} \left( (\tilde{\mathbf{Z}}_j)_{j \in \mathbb{Z}} \right) = \mathcal{L} \left( (\tilde{\mathbf{Y}}_j)_{j \in \mathbb{Z}} \mid \sup_{j \leq -1} |\tilde{\mathbf{Y}}_j|_\infty \leq 1 \right), \quad (4.11)$$

whose spectral decomposition is given by

$$L_Z = \left( \sup_{j \in \mathbb{Z}} |\tilde{\mathbf{Z}}_j|_\infty \right)^{-1}, \quad \tilde{\mathbf{Q}}_j = \frac{\tilde{\mathbf{Z}}_j}{L_Z}, \quad \text{for all } j \in \mathbb{Z}. \quad (4.12)$$

Let  $\underline{\mathbf{Q}} = (\tilde{\mathbf{Q}}_j)_{j \in \mathbb{Z}}$ , which can be seen as a random element in  $\mathbb{S}^* = \{\underline{\mathbf{x}} \in \tilde{l}_\infty^* : |\underline{\mathbf{x}}|^* = 1\}$  (and thus of the form  $(\mathbf{Q}_{j,\kappa})_{j \in \mathbb{Z}, \kappa \in \mathbb{Z}}$ ). Observe also that conditions  $\mathcal{A}_{q_n}$ ,  $\mathcal{A}'_{q_n}$  also need to be adapted. Namely, the sets  $H_j \in \mathcal{F}_{\mathcal{V}}$  in the definition of  $\mathcal{F}$  in (3.6) have to be taken as sets of the form  $\tilde{H}_j$  as defined in (3.22).

**Proposition 4.8.** *Let  $T : \mathcal{X} \rightarrow \mathcal{X}$  be a dynamical system and consider an observable function  $\varphi : \mathcal{X} \rightarrow \mathbb{R}$ . We assume that  $T$  admits an induced system  $\bar{T} : \Delta_0 \rightarrow \Delta_0$  as described above and such that  $\bar{T}$  is uniformly expanding. We assume that the adapted conditions  $\bar{\Delta}_{q_n}$ ,  $\bar{\Delta}'_{q_n}$  are satisfied and, moreover, the transformed anchored tail process given in (4.11) is well defined. Consider the  $N_n^*$  defined as in (4.10), then  $N_n^*$  converges weakly in  $\mathcal{N}_{\mathbb{R}_0^+ \times \tilde{l}_\infty^* \setminus \{\infty\}}^\#$  to the Poisson point process  $N^*$ , which can be written as:*

$$N^* = \sum_{i=1}^{\infty} \delta_{(T_i, U_i, \underline{Q}_i)}, \quad (4.13)$$

where  $(T_i)_{i \in \mathbb{N}}$  and  $(U_i)_{i \in \mathbb{N}}$  are such that  $\sum_{i=1}^{\infty} \delta_{(T_i, U_i)}$  is a bidimensional Poisson point process on  $\mathbb{R}_0^+ \times \mathbb{R}_0^+$  with intensity measure  $\text{Leb} \times \theta \text{Leb}$  and  $(\underline{Q}_i)_{i \in \mathbb{N}}$  is an i.i.d. sequence of random elements in  $\mathbb{S}^*$  such that each  $\underline{Q}_i$  has a distribution given by (4.12). All the sequences  $(T_i)_{i \in \mathbb{N}}$ ,  $(U_i)_{i \in \mathbb{N}}$  and  $(\underline{Q}_i)_{i \in \mathbb{N}}$  are mutually independent.

The proof of this result follows as in the proof of Theorem 4.1 with the necessary straightforward adjustments.

**4.4.1. Functional limit theorem with clustering through inducing.** The point process convergence given in Proposition 4.8 allows to easily obtain the conclusions of Theorems 2.10 and 2.12 as they were obtained in Sections 4.2 and 4.3 from the convergence of the point processes stated in Theorem 4.1.

However, in order to obtain the conclusions of Theorem 2.5 from the convergence of the point processes, as we did in Proposition 4.4, one needs to add an extra condition to ensure that projection of the information registered during the inducing period does not pile up to create discontinuities in the sum. Essentially, we need to forbid the accumulation of many very small contributions which add up to have a significant impact on the sums: the sums should be mostly influenced by heavy tailed observations corresponding to entrances of the orbit near the set  $\mathcal{M}$ , where  $\varphi$  is maximised. Namely, we assume that, for every  $j \in \mathbb{Z}$ ,

$$\mathcal{L} \left( \frac{\bar{\Phi} \circ \bar{T}^{r_n+j}}{\tau^{-1/\alpha} a_n} \left| \left\{ \bar{\varphi} \circ \bar{T}^{r_n} > u_n(\tau) \right\}, \bigcap_{i=1}^{q_n} \left\{ \bar{\varphi} \circ \bar{T}^{r_n-i} \leq u_n(\tau) \right\} \right. \right) \rightarrow \mathcal{L} \left( \Sigma \left( \tilde{\Xi}(\tilde{\mathbf{Z}}_j) \right) \right) \quad (4.14)$$

where  $a_n$  is as in (2.19) and the function  $\Sigma : \tilde{l}_0 \rightarrow \mathbb{R}$ , to each  $\mathbf{x} \in \tilde{l}_0$ , assigns  $\Sigma(\mathbf{x}) := \sum_{j \in \mathbb{Z}} x_j$ , where  $\mathbf{x} = (\dots, x_{-1}, x_0, x_1, \dots) \in l_0$  is such that  $\tilde{\mathbf{x}} = \tilde{\pi}(\mathbf{x})$ . This condition guarantees some sort of tightness so that the aggregate effect of the excursions performed during the induced periods is completely captured by the sum of the entrances of each component,  $\tilde{\mathbf{Z}}_j$ , of the transformed anchored tail process.

In order to illustrate the adjustments needed to be performed to obtain the conclusions of Theorems 2.5, 2.10 and 2.12 from the point process convergence stated in Proposition 4.8, we will consider the most complicated case regarding the functional limit for sums. The excursions here have two levels: one corresponding to the clustering observed for the induced system (see the middle term of (4.15) below) and the digression performed during the induced period (see the last term of (4.15)): our time parametrisation  $v(t)$  thus runs through the contributions from the induced observable, which are added when  $v(t)$  is an integer, and the excursions during the inducing time, which are added when  $v(t)$  is between integers.

**Theorem 4.9.** *Let  $T : \mathcal{X} \rightarrow \mathcal{X}$  be a dynamical system as in Proposition 4.8 such that  $N_n^*$ , defined as in (4.10), converges weakly in  $\mathcal{N}_{\mathbb{R}_0^+ \times \tilde{l}_\infty^* \setminus \{\infty\}}^\#$  to the Poisson point process  $N^*$ , described in (4.13). Let  $\mathcal{M} \subset \mathcal{X}$  be such that  $\mu(\mathcal{M}) = 0$  and  $\varphi : \mathcal{X} \rightarrow \mathbb{R}$  be such that  $\varphi(x) = g(\text{dist}(x, \mathcal{M}))$ , where  $g$  is of type  $g_2$  and condition (2.19) holds, with  $0 < \alpha < 1$ . Assume also that condition (4.14) holds. Then  $S_n$  defined in (2.20) converges, in  $F'$ , to*

$\underline{V} := (V, \text{disc}(V), \{e_V^s\}_{s \in \text{disc}(V)})$ , where  $V$  is an  $\alpha$ -stable Lévy process on  $[0, 1]$  which can be written as

$$V(t) = \sum_{T_i \leq \mu(\Delta_0)t} \sum_{j \in \mathbb{Z}} U_i^{-\frac{1}{\alpha}} \Sigma(\tilde{Q}_{i,j})$$

and the excursions can be represented for  $t \in [0, 1]$  by

$$e_V^{T_i}(t) = V(T_i^-) + U_i^{-\frac{1}{\alpha}} \sum_{j < \lfloor v(t) \rfloor} \Sigma(\tilde{Q}_{i,j}) + U_i^{-\frac{1}{\alpha}} \sum_{\kappa \leq \lfloor v(v(t) - \lfloor v(t) \rfloor) \rfloor} \mathcal{Q}_{i, \lfloor v(t) \rfloor, \kappa}, \quad (4.15)$$

where  $(\mathcal{Q}_{i,j,\kappa})_{i \in \mathbb{N}, j \in \mathbb{Z}, \kappa \in \mathbb{Z}}$  is a representative of  $\tilde{Q}_{i,j} = \tilde{\Xi}(\tilde{Q}_{i,j})$  and  $(\tilde{Q}_{i,j})_{i \in \mathbb{N}, j \in \mathbb{Z}}$ ,  $(T_i)_{i \in \mathbb{N}}$ ,  $(U_i)_{i \in \mathbb{N}}$  are as in Proposition 4.8.

*Remark 4.10.* This result can be generalised to include the cases  $1 \leq \alpha \leq 2$  and  $d > 1$  as we did earlier.

*Proof.* We start by defining a projection  $\Upsilon: \mathcal{N}_{\mathbb{R}_0^+ \times \tilde{l}_\infty^* \setminus \{\infty\}}^\# \rightarrow F'([0, 1], \mathbb{R}^d)$ . Suppose we are given  $\gamma = \sum_{i=1}^\infty \delta_{(t_i, \mathbf{x}_i)} \in \mathcal{N}_{\mathbb{R}_0^+ \times \tilde{l}_\infty^* \setminus \{\infty\}}^\#$ . At time  $t_i$  we have  $\mathbf{x}_i \in \tilde{l}_\infty^*$ : let  $(\dots, \tilde{\mathbf{x}}_{i,-1}, \tilde{\mathbf{x}}_{i,0}, \tilde{\mathbf{x}}_{i,1}, \dots)$  be a representative of this in  $\tilde{l}_\infty^*$  and for each  $i, j \in \mathbb{Z}$  let  $(\dots, x_{i,j,-1}, x_{i,j,0}, x_{i,j,1}, \dots)$  be a representative of  $\tilde{\mathbf{x}}_{i,j}$  in  $l_\infty$ . Define

$$\hat{e}_x^{t_i}(t) = \sum_{j=-\infty}^{\lfloor v(t) \rfloor - 1} \Sigma(\tilde{\Xi}(\tilde{\mathbf{x}}_{i,j})) + \sum_{\kappa \leq \lfloor v((v(t) - \lfloor v(t) \rfloor) - \frac{1}{2}) \rfloor} \xi(x_{i, \lfloor h(t) \rfloor, \kappa}), \quad x(t) = \sum_{t_i \leq t} \hat{e}_x^{t_i}(1)$$

and  $S^x = \{t_i\}_i$ . Finally, set  $e_x^{t_i} = x(t_i^-) + \hat{e}_x^{t_i}$  and let  $\Upsilon(\gamma) = (x, S^x, \{e_x^{t_i}\}_i)$ . Following the same steps as in the proof of Proposition 4.4 and using condition (4.14) to guarantee non-degeneracy of the limits of the sums calculated during the induced excursions we obtain that  $\Upsilon(N_n^*) = \underline{V}_n = (\bar{V}_n, \text{disc}(\bar{V}_n), \{e_{\bar{V}_n}^s\})$  converges in  $F'$  to  $\Upsilon(N^*) = \underline{V} = (\bar{V}, \text{disc}(\bar{V}), \{e_{\bar{V}}^s\})$ , where

$$\bar{V}(t) = \sum_{T_i \leq t} \sum_{j \in \mathbb{Z}} U_i^{-\frac{1}{\alpha}} \Sigma(\tilde{Q}_{i,j}),$$

and the excursions (defined in  $\tilde{D}$ , which allows time deformation) are as in (4.15).

We define the  $n$ -th return time  $R_n: \Delta_0 \rightarrow \mathbb{N}$  by  $R_n(x) = \sum_{j=0}^{n-1} R \circ \bar{T}^j(x)$  and the occupation times  $N_n: \Delta_0 \rightarrow \mathbb{N}$  by  $N_n(x) = \max\{j \in \mathbb{N}_0: R_j \leq n\}$ . By the ergodic theorem, we have that  $R_n/n \rightarrow \int R d\mu_0 = \mu(\Delta_0)^{-1}$ , a.e. As a consequence, we obtain a strong law for the renewal process, which implies that  $N_n/n \rightarrow \mu(\Delta_0)$ , a.e. and, in fact, we have  $\sup_{t \in [0, 1]} |N_{\lfloor tn \rfloor}/n - t\mu(\Delta_0)| \xrightarrow{n \rightarrow \infty} 0$ , a.e.

We produce a time change to the process  $\bar{V}_n$  in order to approximate  $S_n$ . Namely, we consider the sequence random elements of  $F'$  denoted by  $\underline{V}_n = (V_n, \text{disc}(V_n), \{e_{V_n}^s\})$ , where  $V_n(t) = \bar{V}_n(\mu(\Delta_0)t)$ ,  $s \in \text{disc}(V_n)$  iff  $\mu(\Delta_0)s \in \text{disc}(\bar{V}_n)$  and  $e_{V_n}^s = e_{\bar{V}_n}^{\mu(\Delta_0)s}$ . Note that clearly,  $\underline{V}_n$  converges in  $F'$  to  $\underline{V}$  given in the statement of the proposition.

By a Slutsky argument, the conclusion follows once we prove that

$$d_{F'}(S_n, \underline{V}_n) \xrightarrow{n \rightarrow \infty} 0 \quad \text{in probability.}$$

Note that the excursions of  $\underline{V}_n$  keep track of all the oscillations of  $S_n$ , except for a possible discrepancy near  $t = 1$ , where there may be a lack or excess of data. Away from  $t = 1$ , for  $n$  large enough, since  $\tilde{D}$  allows for time deformation, the distance in  $F'$  comes from the projection into the  $E$  component. But since the excursions keep all the information, then the range of values in vertical direction will coincide and the distance between the graphs will result from the time component. The time correction needed to align the graphs

comes from two sources. One is deterministic and results from the fact that the clock of  $S_n$  moves at steps  $1/n$  while the clock of  $\bar{V}_n$  has a step size of  $1/k_n$ . This means we need to match intervals  $[ir_n/n, (i+1)r_n/n]$  and  $[i/k_n, (i+1)/k_n]$ , whose time gap is bounded by  $2k_n/n$ . In fact,  $ir_n/n \leq i/k_n$  and  $i/k_n - ir_n/n \leq i/k_n(1 - \lfloor n/k_n \rfloor / (n/k_n)) \leq k_n/n$ . The second is random and depends on the distance between  $N_{\lfloor nt \rfloor}/n$  and  $\mu(\Delta_0)t$ , which converges uniformly in  $t$  to 0,  $\mu$ -a.e. Near  $t = 1$ , we have to consider also possible mismatches resulting from the fact that  $n - k_n r_n = O(1/k_n)$ , which means that, besides the time contributions for  $d_{F'}$  already considered earlier, we need to consider differences in the vertical range that are bounded by  $\sum_{j=k_n r_n+1}^n \frac{\bar{\Phi}(\bar{T}^j(x))}{a_n}$ , which converges to 0 in probability.  $\square$

**4.4.2. Application to the Manneville-Pomeau case.** We apply these tools to the LSV maps  $T_\gamma$  defined in (2.7). Recall that, for  $\gamma \in (0, 1)$ , the map  $T_\gamma$  has an invariant measure  $\mu_\gamma$  absolutely continuous with respect to Lebesgue such that its density  $h_\gamma = \frac{d\mu_\gamma}{dx}$  is Lipschitz on any interval of the form  $(\varepsilon, 1]$  and  $\lim_{x \rightarrow 0} \frac{h(x)}{x^{-\gamma}} = C_0 > 0$  (see [46], for example). Consider the canonical inducing domain  $\Delta_0 = (1/2, 1]$  and define  $R : \Delta_0 \rightarrow \mathbb{N}$  as the first return time, i.e.,  $R(x) = \inf\{j \in \mathbb{N} : T_\gamma^j(x) \in \Delta_0\}$ . As before we denote by  $\mu_0$  the restriction of  $\mu_\gamma$  to  $\Delta_0$ . The induced map  $\bar{T}_\gamma : \Delta_0 \rightarrow \Delta_0$  given by  $\bar{T}_\gamma = T_\gamma^{R(x)}(x)$  qualifies as a Rychlik map and therefore it has exponential decay of correlations of BV functions against  $L^1$ , so conditions  $\bar{\mathcal{A}}_{q_n}, \bar{\mathcal{A}}'_{q_n}$  can be shown to hold by adapting the argument used in the proof of [30, Theroem 4.3]. With the application to heavy tailed sums in mind, for definiteness we assume that  $\varphi : [0, 1] \rightarrow \mathbb{R}$  is such that  $\varphi(x) = |x - \zeta|^{-1/\alpha}$ , with  $0 < \alpha < 1$  and  $\zeta$  the period two point belonging to  $\Delta_0$ . The fact that the inducing map has such nice mixing properties allows us to apply Proposition 4.8, where  $\underline{\mathbf{Q}}_i = (\dots, \infty, \infty, \tilde{\mathbf{K}}_0, \tilde{\mathbf{K}}_1, \tilde{\mathbf{K}}_2, \dots)$ ,  $\tilde{\mathbf{K}}_j = (\dots, \infty, \infty, \chi^j, \infty, \infty, \dots)$  and  $\chi = DT_\gamma^2(\zeta)$ . The computation of this transformed anchored tail process follows the exact same argument used in Appendix D.1. The fact that all entries but one are  $\infty$  in  $\tilde{\mathbf{K}}_j$  results from the canonical embedding and from the fact that during the induced time the orbits are always outside  $\Delta_0$ , which means that they are at fixed distance from  $\zeta$  and therefore the normalisation dictates that  $\infty$  should appear for the corresponding limits (see also Appendix D.1 for an argument on how the normalisation leads to  $\infty$  for points at a fixed distance from  $\mathcal{M}$ ).

We focus now on condition (4.14). Notice that if  $|\varphi(0)| > 0$ , then, for an high value  $u$ , the set

$$\left\{ x \in \Delta_0 : \sum_{i=0}^{R(x)-1} |\varphi(T^i(x))| > u \right\} = A_u \cup B_u$$

is the union of two intervals. One corresponding to a small neighbourhood of  $\zeta$ ,  $A_u = (\zeta - \delta_u, \zeta + \delta_u)$ , where the value of  $\varphi$  is already sufficiently high and the other corresponds to an interval of type  $B_u = (1/2, 1/2 + \varepsilon_u)$  because  $T_\gamma(1/2) = 0$  and since  $DT_\gamma(0) = 1$  this means that the orbit will linger for a very long time near 0 so that the sum  $\sum_{i=0}^{R(x)-1} |\varphi(T^i(x))|$  will add up  $|\varphi(0)|$  so many times that it will ultimately exceed the level  $u$ . The shape of the observable near  $\zeta$  dictates that  $\mu_0(A_u) = O(u^{-\alpha})$ . We note that  $\mu_0$  is equivalent to Lebesgue measure on  $\Delta_0$ , in the sense that its density is bounded above and away from 0 (see [55, Lemma 2.3]). Using the same computation as in the proof of Theorem 1.3 of [37], one obtains that  $\mu_0(B_u) = O(u^{-1/\gamma})$ . Now, since  $\alpha < 1 < 1/\gamma$ , then clearly  $\mu_0(B_u) = o(\mu_0(A_u))$  and therefore condition (4.14) is satisfied.

Note that we could consider  $1 < \alpha < 2$ , but in order to guarantee that (4.14) still holds when  $\gamma \in (1/2, 1)$ , we would need the extra assumptions:  $\varphi(0) = 0$  and  $\varphi(x) < Cx^\beta$  for some  $\beta > \gamma - 1/2$ . These would still guarantee that  $\mu_0(B_u) = o(\mu_0(A_u))$ . (See proof of Theorem 1.3 of [37]).

As in Example 2.7, we can also consider an oscillatory behaviour creating overshooting by considering  $\mathcal{M} = \{\zeta, T_\gamma(\zeta)\}$  and the corresponding observable  $\psi(x) = |x - \zeta|^{-1/\alpha} - |x - T_\gamma(\zeta)|^{-1/\alpha}$ , where  $\zeta$  is as above. For simplicity, we assume that  $0 < \alpha < 1$ , so that in a similar way we have that condition (4.14) is easily satisfied. In this case, to describe  $\underline{\mathbf{Q}}_i$ , we consider the random variable  $E_i$  taking values in  $\{-1, 1\}$  and such that

$$\mathbb{P}(E_i = 1) = \lim_{u \rightarrow \infty} \frac{\mu_\gamma(\{|\psi| > u\} \cap \Delta_0)}{\mu_\gamma(\{|\psi| > u\})}.$$

In a similar way to the computation of the transformed anchored tail process in Appendix D.2 we obtain that  $\underline{\mathbf{Q}}_i = (\dots, \tilde{\infty}, \tilde{\infty}, \tilde{\mathbf{K}}_0(E_i), \tilde{\mathbf{K}}_1(E_i), \tilde{\mathbf{K}}_2(E_i), \dots)$ , where

$$\begin{aligned} \tilde{\mathbf{K}}_0(-1) &= (\dots, \infty, \infty, -1, \infty, \infty, \dots) \\ \tilde{\mathbf{K}}_j(-1) &= (\dots, \infty, \infty, (-1)^{2j} \chi_{-1}^j \chi_1^{j-1}, (-1)^{2j+1} \chi_{-1}^j \chi_1^j, \infty, \infty, \dots) \\ \tilde{\mathbf{K}}_0(1) &= (\dots, \infty, \infty, 1, \chi_1, \infty, \infty, \dots) \\ \tilde{\mathbf{K}}_j(1) &= (\dots, \infty, \infty, (-1)^{2j-1} \chi_{-1}^j \chi_1^j, (-1)^{2j} \chi_{-1}^j \chi_1^{j+1}, \infty, \infty, \dots) \end{aligned}$$

with  $\chi_1 = DT_\gamma(\zeta)$  and  $\chi_{-1} = DT_\gamma(T_\gamma(\zeta))$ . The resulting excursion is:

$$e_V^{T_i}(t) = V(T_i^-) + U_i^{-1/\alpha} \sum_{0 \leq j \leq \lfloor v(t) \rfloor} E_i(-1)^j \left( \chi_{-1}^{\lfloor j/2 \rfloor - (E_i+1)/2|_+} \chi_1^{\lfloor j/2 \rfloor + (E_i-1)/2|_+} \right)^{-1/\alpha},$$

where  $t \in [0, 1]$  and  $|\cdot|_+ = \max\{0, \cdot\}$ .

#### APPENDIX A. COMPLETENESS AND SEPARABILITY OF THE SPACE $F'$

As in Whitt's space  $E$ , there are two natural metrics to use for the  $E$  component of our space  $F'$ . These are denoted  $m_E$  and  $m_E^*$  (see (2.10) and (2.11)). We keep the same metric on the  $\tilde{D}$  part.

**Lemma A.1.**  *$D$  with the sup norm is complete.*

*Proof.* Let  $x_n \in D$  define a Cauchy sequence w.r.t.  $\|\cdot\|_\infty$ . Then set  $x(t) = \lim_{n \rightarrow \infty} x_n(t)$ . Note that  $\cup_n \text{disc}(x_n)$  is an at most countable set. Observe also that in this metric (unlike  $M_1$  say), each discontinuity  $t$  of  $x$  must correspond to a limit of discontinuities  $(t_n)_n$  of  $(x_n)_n$  (indeed we can take a sequence  $(t_n)_n$  which is eventually constant). So  $x$  must have at most countably many discontinuities  $\text{disc}(x)$ . The fact that  $x$  is continuous on the right of the discontinuities follow from the continuity of  $x_n$  on the right. Hence  $x \in D$ .  $\square$

**Lemma A.2.**  *$\tilde{D}$  is separable and complete.*

*Proof.* Completeness follows since  $D$  with the sup norm is complete: if  $[x_n] \in \tilde{D}$  gives a Cauchy sequence then there are  $y_n \in [x_n]$  such that  $(y_n)_n$  is Cauchy in  $D$  with the sup norm. Then the limit  $y$ , which exists as in Lemma A.1, defines a class in  $\tilde{D}$ .

Separability follows since the set of piecewise constant  $\mathbb{Q}$ -valued functions in  $D$  with rational discontinuity points (the discontinuity points can be any countable set here) is countable and defines a set of equivalence classes which is dense in  $\tilde{D}$ .  $\square$

As shown in [79, Theorem 15.4.3],  $(E, m_E)$  is separable, and in [79, Example 15.4.2] this space is not complete. On the other hand for  $(E, m_E^*)$  is complete, but not separable.

Assume we are dealing with real-valued functions rather than  $\mathbb{R}^d$ -valued for  $d \geq 2$ .

**Proposition A.3.** (a)  *$F'$  with the  $m_E$ -metric on the  $E$ -component is not complete, but is separable.*

(b)  *$F'$  with the  $m_E^*$ -metric on the  $E$ -component is complete, but is not separable.*

*Proof.* (a) [79, Example 15.4.2] would not give a convergent sequence in  $F'$  since we would not be able to get convergent excursions: there are large jumps in adjacent terms of the sequence which cannot be matched up. Hence we need an alternative example for non-completeness.

For each  $n$  we define

$$\phi_n(x) = \begin{cases} nx & \text{if } x \in [0, \frac{1}{n}) \\ 1 & \text{if } x \in [\frac{1}{n}, 1] \end{cases}$$

Also let, for each  $k \in \mathbb{N}$ ,  $x_{k,1}, \dots, x_{k,n+k}$  be the set of  $n+k$  points in  $[\frac{1}{2^k}, \frac{1}{2^{k-1}}]$  so that these points, along with the boundary points, are equidistributed through this interval. Define  $\varepsilon_{n,k} = \frac{1}{n2^{k+2}}$ . Then define

$$x_n = \sum_{k=1}^{\infty} \sum_{i=1}^{n+k} \phi_n \cdot \mathbb{1}_{[x_{k,i}-\varepsilon_{n,k}, x_{k,i}+\varepsilon_{n,k})}.$$

Then  $S_n = \{x_{k,i} \pm \varepsilon_{n,k}\}_{k,n}$  and the excursions are for example  $\phi_n(x_{k,i} - \varepsilon_{n,k}) \cdot \mathbb{1}_{[1/2,1]}$  at the points  $x_{k,i} - \varepsilon_{n,k}$  and  $\phi_n(x_{k,i} + \varepsilon_{n,k}) \cdot \mathbb{1}_{[0,1/2]}$  at the points  $x_{k,i} + \varepsilon_{n,k}$ , producing an element  $\underline{x}_n \in F'$  (note the damping effect of  $\phi_n$  ensuring that (2.9) holds). Then  $(\underline{x}_n)_n$  is Cauchy with the  $m$ -metric, but this sequence does not have a limit in  $F'$  (eg the graph of the limit would have to be  $[0, 1] \times [0, 1]$  as in [79, Example 15.4.2]). Hence the space is not complete.

To show separability, we use the usual approximation by objects with rational coordinates. As shown in [79, Theorem 15.4.3], the  $E$ -part of the space is separable. We then use Lemma A.2 to give separability of the excursion part.

(b) Completeness follows from the  $E$ -component being complete as in [79, Section 15.4], which is the same type of argument as completeness in  $D$  (see Lemma A.1) and  $\tilde{D}$  since this is uniform convergence. The latter facts give convergence to a limit in  $\tilde{D}$  for the excursions also.

Non-separability of  $E$  (with  $m^*$ ) implies the non-separability of  $F'$ : in fact this is the same type of argument as for  $D$ . Note that for non-separability it is sufficient, for a given positive distance, to find an uncountable collection of elements of the set all at least that distance apart. So for example for  $D$ ,  $\{\mathbb{1}_{[a,1]}\}_{a \in [0,1]}$  are all distance 1 apart in the  $m^*$  metric.  $\square$

*Remark A.4.* [79, Example 15.4.2] would not give a convergent sequence in  $F'$  since we would not be able to get convergent excursions. Hence we needed an alternative example above for non-completeness.

## APPENDIX B. WEAK AND WEAK<sup>#</sup> CONVERGENCE

The purpose of this section is to review notions of convergence of measures on the metric spaces  $\tilde{l}_\infty$  and  $\tilde{l}_0$ . We are particularly interested in weak convergence of probability measures and, since we want to consider point processes which are random elements corresponding to boundedly finite measures, we are also interested in weak<sup>#</sup> convergence. Recall that  $\tilde{l}_\infty$  and  $\tilde{l}_0$  are complete, separable, metric spaces, which are not locally compact and therefore vague convergence is not useful here. We remark that weak<sup>#</sup> convergence is equivalent to vague convergence when the ambient space is locally compact (see [17, Appendix A2.6]).

The Portmanteau Theorem (see for example [12, Theorem 2.1]) is the following.

**Theorem B.1.** *Given a metric space, the following conditions are equivalent to weak convergence of Borel probability measures.*

- (1)  $\lim_{n \rightarrow \infty} \int f d\mathbb{P}_n = \int f d\mathbb{P}$ , for all bounded, uniformly continuous real  $f$ ;
- (2)  $\limsup_{n \rightarrow \infty} \mathbb{P}_n(F) \leq \mathbb{P}(F)$  for all closed set  $F$ ;



- (3)  $\liminf_{n \rightarrow \infty} \mathbb{P}_n(E) \geq \mathbb{P}(E)$  for all open set  $E$ ;  
 (4)  $\lim_{n \rightarrow \infty} \mathbb{P}_n(A) = \mathbb{P}(A)$  for all set  $A$  such that  $\mathbb{P}(\partial A) = 0$ .

A set  $A$  such that  $\mathbb{P}(\partial A) = 0$  is called a  $\mathbb{P}$ -continuity set. A class  $\mathcal{U}$  of sets is called a *convergence determining class* if the convergence  $\lim_{n \rightarrow \infty} \mathbb{P}_n(A) = \mathbb{P}(A)$  for all  $\mathbb{P}$ -continuity sets  $A \in \mathcal{U}$  implies the weak convergence of  $\mathbb{P}_n$  to  $\mathbb{P}$ .

**Lemma B.2.** *The class of sets  $\tilde{\mathcal{J}}$  as in (3.22) is a convergence determining class for weak convergence on the metric space  $\tilde{l}_\infty \setminus \{\tilde{\infty}\}$ .*

*Proof.* We start by proving a claim.

**Claim.** *For all  $\tilde{\mathbf{x}} \in \tilde{l}_\infty \setminus \{\tilde{\infty}\}$  and all  $\varepsilon > 0$ , there exists a  $\mathbb{P}$ -continuity set  $\tilde{A} \in \tilde{\mathcal{J}}$  such that  $\tilde{\mathbf{x}} \in \tilde{A} \subset B(\tilde{\mathbf{x}}, \varepsilon)$ .*

*Proof of Claim.* Let  $\mathbf{z}, \mathbf{w} \in l_0 \setminus \{\mathbf{0}\}$ . Since  $|\|\mathbf{z}\|_\infty - \|\mathbf{w}\|_\infty| \leq \|\mathbf{z} - \mathbf{w}\|_\infty$  and  $h(x) = \frac{1}{x}$  is continuous on  $\mathbb{R} \setminus \{0\}$ , then there exists  $\delta > 0$  such that  $\|\mathbf{z} - \mathbf{w}\|_\infty < \delta$  implies that  $d'(\mathbf{z}, \mathbf{w}) < \varepsilon$ , where  $d'$  is the metric on  $l_0 \setminus \{\mathbf{0}\}$  defined in (3.17).

Let  $\mathbf{z} = P(\mathbf{x}) \in l_0 \setminus \{\mathbf{0}\}$  and let  $k \in \mathbb{N}$  be such that  $\|z_j\| < \delta$  for all  $|j| \geq k$ . Define

$$B_\rho = \bigcap_{|j| < k} \sigma^{-j}([z_{j_1} - \rho, z_{j_1} + \rho) \times \dots \times [z_{j_d} - \rho, z_{j_d} + \rho)),$$

for some  $\rho < \delta/\sqrt{d}$ . Let  $A_\rho = P^{-1}(B_\rho)$  and let  $\tilde{A}_\rho \in \tilde{\mathcal{J}}$  be associated to  $A_\rho$  as in (3.22). We choose  $\rho$  so that  $\tilde{A}_\rho$  is a  $\mathbb{P}$ -continuity set. We can always choose such  $\rho$  because  $\partial \tilde{A}_\rho \subset \tilde{\pi}(\cup_{i \in \mathbb{Z}} (\sigma^{-i}(\cup_{|j| < k} \cup_{m=1}^d \{\mathbf{w} : |w_{j_m} - z_{j_m}| = \rho\})))$ , which means that each  $\partial \tilde{A}_\rho$  can intersect at most countably many other such sets. Therefore, there is an uncountable number of disjoint sets  $\partial \tilde{A}_\rho$ , for  $0 < \rho < \delta/\sqrt{d}$ , and since there cannot be an uncountable number of them with positive probability, at least one of them must have 0 probability, which means one of the  $\tilde{A}_\rho$  is a  $\mathbb{P}$ -continuity set.

We claim that  $\tilde{A}_\rho \subset B(\tilde{\mathbf{x}}, \varepsilon)$ . To see this, let  $\mathbf{w} \in l_\infty$  be such that  $\tilde{\mathbf{w}} = \tilde{\pi}(\mathbf{w}) \in \tilde{A}_\rho$ . By definition of  $\tilde{A}_\rho$  we must have  $\sigma^\ell(\mathbf{w}) \in A_\rho$ , for some  $\ell \in \mathbb{Z}$ . Hence,  $P(\sigma^\ell(\mathbf{w})) = \sigma^\ell(P(\mathbf{w})) \in B_\rho$ , which means that  $\|P(\mathbf{x}) - \sigma^\ell(P(\mathbf{w}))\|_\infty < \delta$  and therefore  $d'(P(\mathbf{x}), \sigma^\ell(P(\mathbf{w}))) < \varepsilon$ . It follows that, by definition of the metric  $\tilde{d}'$  on  $\tilde{l}_0$ , given in (3.21), we have  $\tilde{d}'(\tilde{P}(\tilde{\mathbf{x}}), \tilde{P}(\tilde{\mathbf{w}})) < \varepsilon$  and hence  $\tilde{d}(\tilde{\mathbf{x}}, \tilde{\mathbf{w}}) < \varepsilon$ .  $\square$

Let  $E$  be an open set. By the claim, for each  $\tilde{\mathbf{x}} \in E$  there exists a  $\mathbb{P}$ -continuity set  $\tilde{A} \in \tilde{\mathcal{J}}$  such that  $\tilde{\mathbf{x}} \in \tilde{A} \subset B(\tilde{\mathbf{x}}, \varepsilon)$ . This means we have an open cover of  $E$  by sets  $\tilde{A}$ , where  $\tilde{A} \in \tilde{\mathcal{J}}$  is a  $\mathbb{P}$ -continuity set. Since on a separable metric space (such as  $\tilde{l}_\infty \setminus \{\tilde{\infty}\}$ ) each open cover admits a countable subcover (see [12, Appendix M3], for example) then there exists a sequence  $(\tilde{A}_i)_{i \in \mathbb{N}}$  such that each  $\tilde{A}_i \in \tilde{\mathcal{J}}$  is a  $\mathbb{P}$ -continuity set,  $\tilde{A}_i \subset E$  and, moreover,  $E \subset \cup_{i \in \mathbb{N}} \tilde{A}_i$ , which means that  $E = \cup_{i \in \mathbb{N}} \tilde{A}_i$ .

We are assuming that  $\lim_{n \rightarrow \infty} \mathbb{P}_n(\tilde{A}) = \mathbb{P}(\tilde{A})$  for all  $\mathbb{P}$ -continuity sets  $\tilde{A} \in \tilde{\mathcal{J}}$ . Since  $\tilde{\mathcal{J}}$  is closed for finite unions and  $\partial(\tilde{A} \cup \tilde{B}) \subset \partial \tilde{A} \cup \partial \tilde{B}$  (see [17, Proposition A1.2.I]), we have that  $\cup_{i=1}^m \tilde{A}_i \in \tilde{\mathcal{J}}$  is a  $\mathbb{P}$ -continuity set and therefore

$$\lim_{n \rightarrow \infty} \mathbb{P}_n(\cup_{i=1}^m \tilde{A}_i) = \mathbb{P}(\cup_{i=1}^m \tilde{A}_i).$$

Since  $E = \cup_{i \in \mathbb{N}} \tilde{A}_i$ , for  $\varepsilon > 0$ , let  $m \in \mathbb{N}$  be such that  $\mathbb{P}(\cup_{i=1}^m \tilde{A}_i) > \mathbb{P}(E) - \varepsilon$ . Then,

$$\mathbb{P}(E) - \varepsilon < \mathbb{P}(\cup_{i=1}^m \tilde{A}_i) = \lim_{n \rightarrow \infty} \mathbb{P}_n(\cup_{i=1}^m \tilde{A}_i) \leq \liminf_{n \rightarrow \infty} \mathbb{P}_n(E).$$

Since the last inequalities hold for all  $\varepsilon > 0$  then

$$\mathbb{P}(E) \leq \liminf_{n \rightarrow \infty} \mathbb{P}_n(E), \quad \text{for all } E \text{ open}$$

and therefore by Theorem B.1(3),  $\mathbb{P}_n$  converges weakly to  $\mathbb{P}$ .  $\square$

In order to study point processes, we need to consider  $\sigma$ -finite measures taking finite measures on bounded sets. Namely, we define:

**Definition B.3.** A Borel measure  $\mu$  on a complete, separable, metric space is boundedly finite if  $\mu(A) < \infty$  for every bounded Borel set  $A$ .

Let  $\mathcal{X}$  denote a complete, separable, metric space such as  $\tilde{l}_\infty \setminus \{\tilde{\infty}\}$ . We denote by  $\mathcal{M}_\mathcal{X}^\#$  the space of boundedly finite Borel measures on  $\mathcal{X}$ . Following [17, 18], we consider a notion of convergence in  $\mathcal{M}_\mathcal{X}^\#$  called the *weak hash* or *weak<sup>#</sup>* convergence, denoted by  $\mu_k \rightarrow_{w^\#} \mu$  which can be defined by any of the following equivalent conditions (see [17, Proposition A2.6.II]):

- (i)  $\lim_{k \rightarrow \infty} \int f d\mu_k = \int f d\mu$  for all bounded continuous functions  $f$  defined on  $\mathcal{X}$  and vanishing outside a bounded set;
- (ii) There exists an increasing sequence of bounded open sets  $B_n$  converging to  $\mathcal{X}$  such that if  $\mu_k^{(n)}$  and  $\mu^{(n)}$  denote the restrictions of the measures  $\mu_k$  and  $\mu$  to  $B_n$ , respectively, then  $\mu_k^{(n)}$  converges weakly to  $\mu^{(n)}$ , as  $k \rightarrow \infty$ , for all  $n \in \mathbb{N}$ . Note that  $\mu_k^{(n)}$  and  $\mu^{(n)}$  are not necessarily probability measures, so when we say that there is weak convergence we mean that either (1) and (4) from Theorem B.1 apply or (2) and (3) apply with the extra assumption that  $\lim_{k \rightarrow \infty} \mu_k^{(n)}(B_n) = \mu^{(n)}(B_n)$ , for all  $n \in \mathbb{N}$ .
- (iii)  $\lim_{k \rightarrow \infty} \mu_k(A) = \mu(A)$  for all bounded Borelean  $A$  such that  $\mu(\partial A) = 0$ .

Similarly to Lemma B.2, we show that in order to prove weak<sup>#</sup> convergence on  $\tilde{l}_\infty \setminus \{\tilde{\infty}\}$ , we only need to check  $\lim_{k \rightarrow \infty} \mu_k(\tilde{A}) = \mu(\tilde{A})$  for all bounded,  $\mu$ -continuity sets  $\tilde{A} \in \tilde{\mathcal{J}}$ , as in property (iii).

**Lemma B.4.** *The class of bounded sets in  $\tilde{\mathcal{J}}$  is a convergence determining class for weak<sup>#</sup> convergence on the metric space  $\tilde{l}_\infty \setminus \{\tilde{\infty}\}$ .*

*Proof.* Suppose  $\mu_k \rightarrow_{w^\#} \mu$ . For all  $a > 1$ , let  $\mathcal{F} \ni B_a := \{\mathbf{x} \in \dot{\mathcal{V}}^\mathbb{Z} : 1/a < \|x_0\| < a\}$  and  $\tilde{B}_a \in \tilde{\mathcal{J}}$  be associated to  $B_a$  as in (3.22). Observe that  $\tilde{d}(\tilde{\mathbf{x}}, \tilde{\mathbf{w}}) \leq a - 1/a$ , for all  $\tilde{x}, \tilde{w} \in \tilde{B}_a$ . Since there are uncountably many sets of the form  $B_a$  and  $\partial B_a \subset \{\mathbf{x} \in \dot{\mathcal{V}}^\mathbb{Z} : 1/a = \|x_0\| \text{ or } \|x_0\| = a\}$ , then we can find a strictly increasing, diverging sequence  $(a_n)_{n \in \mathbb{N}}$ , such that  $B_n := B_{a_n}$  is such that  $\mu(\partial \tilde{B}_n) = 0$ , for all  $n \in \mathbb{N}$ .

Hence,  $(B_n)_{n \in \mathbb{N}}$  is an increasing sequence of bounded, open,  $\mu$ -continuity sets converging to  $\tilde{l}_\infty \setminus \{\tilde{\infty}\}$ . Note that by hypothesis we are assuming that  $\lim_{k \rightarrow \infty} \mu_k(\tilde{A}) = \mu(\tilde{A})$  for all bounded,  $\mu$ -continuity sets  $\tilde{A} \in \tilde{\mathcal{J}}$  and therefore that applies to all the  $\tilde{B}_n$ , namely,  $\lim_{k \rightarrow \infty} \mu_k(\tilde{B}_n) = \mu(\tilde{B}_n)$ . Therefore, we can apply (3) to prove weak convergence of  $\mu_k^{(n)}$  (the restriction of  $\mu_k$  to  $B_n$ ) to  $\mu^{(n)}$  (the restriction of  $\mu$  to  $B_n$ ), which was the main tool used in the proof of Lemma B.2. It follows then, by assumption and Lemma B.2 that  $\mu_k^{(n)}$  converges weakly to  $\mu^{(n)}$ , for all  $n \in \mathbb{N}$  and, by (ii), we obtain the weak<sup>#</sup> convergence of  $\mu_k$  to  $\mu$ .  $\square$

## APPENDIX C. CONVERGENCE OF POINT PROCESSES ON NON LOCALLY COMPACT SPACES

We closely follow [17, Appendix A2.6] and [18, Sections 9 and 11]. As before, let  $\mathcal{X}$  be a complete separable metric space. The notion of weak<sup>#</sup> convergence is metrizable, in the sense that  $\mathcal{M}_\mathcal{X}^\#$  admits a metric generating what the so-called  $w^\#$ -topology so that the

weak<sup>#</sup> convergence corresponds to convergence in the  $w^\#$ -topology. Denote by  $\mathcal{B}(\mathcal{M}_\mathcal{X}^\#)$  the corresponding Borel  $\sigma$ -algebra. A *random measure* is a random element in  $(\mathcal{M}_\mathcal{X}^\#, \mathcal{B}(\mathcal{M}_\mathcal{X}^\#))$ . A point process  $N$  is an integer-valued random measure. Let  $\mathcal{N}_\mathcal{X}^\#$  denote the space of boundedly finite integer-valued measures. We have that  $\mathcal{N}_\mathcal{X}^\#$  is a closed subset of  $\mathcal{M}_\mathcal{X}^\#$  ([18, Proposition 9.1.V]) and let  $\mathcal{B}(\mathcal{N}_\mathcal{X}^\#)$  denote the corresponding  $\sigma$ -algebra for the  $w^\#$ -topology. We remark that  $\mu \in \mathcal{N}_\mathcal{X}^\#$  has the following form ([18, Proposition 9.1.III]):

$$\mu = \sum_{i \in \mathbb{N}} k_i \delta_{x_i}, \quad \text{where } \delta_{x_i} \text{ is the Dirac measure at distinct } x_i \in \mathcal{X}, \text{ and } k_i \in \mathbb{N}. \quad (\text{C.1})$$

For any  $\mu \in \mathcal{N}_\mathcal{X}^\#$  given by (C.1), we define its *support counting measure*  $\mu^*$  as:

$$\mu^* = \sum_{i \in \mathbb{N}} \delta_{x_i}. \quad (\text{C.2})$$

The boundedly finite measure  $\mu \in \mathcal{N}_\mathcal{X}^\#$  is simple if and only if  $\mu = \mu^*$ , which is equivalent to say that  $k_i = 1$  for all  $i \in \mathbb{N}$ .

Now, we give formally now the definition of point process.

**Definition C.1.** A point process  $N$  on state space  $\mathcal{X}$  is measurable mapping from a probability space  $(\Omega, \mathcal{B}, \mathbb{P})$  into  $(\mathcal{N}_\mathcal{X}^\#, \mathcal{B}(\mathcal{N}_\mathcal{X}^\#))$ . A point process  $N$  is said to be *simple* if  $\mathbb{P}(N(\{x\}) > 1) = 0$ , for all  $x \in \mathcal{X}$ . To each point process  $N$  on state space  $\mathcal{X}$  we denote by  $N^*$  the corresponding *support point process* obtained from  $N$  as in (C.2).

Let  $(N_n)_{n \in \mathbb{N}}$  be a sequence of point processes and  $N$  another point process, all with state space  $\mathcal{X}$ . We say that  $N_n$  converges weakly to  $N$ , when the respective distributions  $P_n$ , defined by  $P_n(A) := \mathbb{P}(N_n \in A)$ , for all  $A \in \mathcal{B}(\mathcal{N}_\mathcal{X}^\#)$ , converge weakly (in the sense of weak convergence of probability measures on the metric space  $\mathcal{N}_\mathcal{X}^\#$ ) to the distribution  $P$  associated to  $N$ .

Tightness is a very useful property which gives relative compactness and ultimately weak convergence. We state two conditions that are necessary and sufficient to show that a family of probability measures  $\{P_t, t \in \mathcal{T}\}$  is *uniformly tight*, see [18, Proposition 11.1.VI]. Given any closed sphere  $S \subset \mathcal{X}$  and any  $\epsilon, \delta > 0$ , there exists a real number  $M$  and a compact set  $C \subset S$  such that for  $t \in \mathcal{T}$ ,

$$P_t(N_t(S) > M) < \epsilon, \quad (\text{C.3})$$

$$P_t(N_t(S \setminus C) > \delta) < \epsilon. \quad (\text{C.4})$$

We remark that if  $\mathcal{X}$  was locally compact, (C.4) would be unnecessary.

Based on these criteria for tightness, in [18, Proposition 11.1.VII] it is shown that weak convergence of  $N_n$  to  $N$  follows from the convergence:

$$(N_n(A_1), \dots, N_n(A_k)) \rightarrow (N(A_1), \dots, N(A_k)),$$

of joint distributions as random vectors in  $\mathbb{R}^k$ , for all finite collections  $\{A_1, \dots, A_k\}$  of bounded continuity sets  $A_i \in \mathcal{B}_\mathcal{X}$ , for all  $i = 1, \dots, k$  and all  $k \in \mathbb{N}$ . Here, *continuity set* means that  $N(\partial A_i) = 0$  a.s.

When the limiting point process is simple, a simpler criterion for convergence can be used.

**Proposition C.2.** Let  $(N_n)_{n \in \mathbb{N}}$  be a sequence of point processes on the state space  $\mathcal{X}$  and let  $N$  be simple point process on the same state space. Let  $\mathcal{R}$  be a covering dissecting ring<sup>4</sup>

<sup>4</sup>A dissecting ring is a ring generated by finite unions and intersections of the elements of a dissecting system, which consists on a nested sequence of finite partitions of the whole space that eventually separate points. See [17, Appendices A1.6 and A2.1] for further details.

of continuity sets of  $\mathcal{X}$ . Then,  $N_n$  converges weakly to  $N$  if the two following conditions hold:

- (i)  $\lim_{n \rightarrow \infty} \mathbb{P}(N_n(A) = 0) = \mathbb{P}(N(A) = 0)$  for all bounded  $A \in \mathcal{R}$ ;
- (ii) for all bounded  $A \in \mathcal{R}$  and a nested sequence of partitions  $\mathcal{T}_r = \{A_{ri} : i = 1, \dots, k_r\}$  of  $A$  by sets of  $\mathcal{R}$  that ultimately separate the points of  $A$ ,

$$\inf_{\mathcal{T}_r} \limsup_{n \rightarrow \infty} \sum_{i=1}^{k_r} \mathbb{P}(N_n(A_{ri}) \geq 2) = 0.$$

Alternatively, we may replace condition (ii) by the following:

$$(II) \lim_{n \rightarrow \infty} \mathbb{E}(N_n(A)) = \mathbb{E}(N(A)), \text{ for all bounded } A \in \mathcal{R}.$$

*Remark C.3.* This proposition is very similar to [18, Proposition 11.1.IX]. However, note that the corresponding statement to our condition (ii), namely [18, equation (11.1.4)], is incorrectly stated there. Essentially, this condition is supposed to require that there is no accumulation of mass points as the point processes approach the limiting point process. This is related to a property called *ordinary* in [18, equation (11.1.4)], which is then used in the proof of [18, Proposition 11.1.IX]. When one observes the condition giving the notion of ordinary point process one realises the need to use the infimum over all partitions, as we did in (ii), which contrasts with the use of the supremum over all partitions used in [18, equation (11.1.4)].

For completeness we redo the proof of [18, Proposition 11.1.IX], correcting this typo. We also extend the result by showing that condition (II) can be used to replace condition (ii) when trying to prove weak convergence to a simple point process.

*Proof.* We closely follow the proof of [18, Proposition 11.1.IX]. At the core of the proof this proposition is a result attributed to Rényi and Mönch, see [18, Theorem 9.2.XII], which states that the distribution of a simple point process on a complete separable metric spaces  $\mathcal{X}$  is determined by the values of the avoidance function  $P_0$  on the bounded sets of a dissecting ring  $\mathcal{R}$  for  $\mathcal{X}$ , where  $P_0$  is defined by:

$$P_0(A) = \mathbb{P}(N(A) = 0), \quad \text{for } A \in \mathcal{R}.$$

We are then left to show that the family  $\{P_n : n \geq n_0\}$  is uniformly tight and that the limit of any convergent subsequence is a simple point process, where  $P_n$  is the distribution of the point process  $N_n$ .

Assuming that (ii) holds, let  $A$  be a closed ball in  $\mathcal{R}$ . Observing that  $\{N_n(A) > k_r\}$  implies  $\{N_n(A_{ri}) \geq 2\}$  for at least one  $i$ , then

$$\sum_{i=1}^{k_r} \mathbb{P}(N_n(A_{ri}) \geq 2) \geq \mathbb{P}(N_n(A) \geq k_r).$$

Note that the sum on the left is non-increasing with  $r$ . Hence, given  $\epsilon > 0$ , by (ii), there is an  $r_0$  such that  $\forall r \geq r_0$ , we have  $\limsup_{n \rightarrow \infty} \sum_{i=1}^{k_r} \mathbb{P}(N_n(A_{ri}) \geq 2) < \epsilon$ . In particular, there exists  $n_0 \in \mathbb{N}$  such that for all  $n > n_0$ , we have  $\mathbb{P}(N_n(A) \geq k_{r_0}) \leq \sum_{i=1}^{k_{r_0}} \mathbb{P}(N_n(A_{r_0i}) \geq 2) < \epsilon$ . Therefore, choosing  $M > k_{r_0}$  large enough, we have  $\mathbb{P}(N_n(A) \geq M) < \epsilon$ , for all  $n \in \mathbb{N}$ , which means that C.3 holds.

Now, we verify that the same happens when we assume (II) instead of (ii). Using Chebyshev's inequality, we obtain for all  $n \in \mathbb{N}$

$$\mathbb{P}(N_n(S) > M) \leq \frac{\mathbb{E}(N_n(S))}{M}.$$

By (II), there exists  $K \in \mathbb{N}$  such that  $\mathbb{E}(N_n(S)) < K$ , for all  $n \in \mathbb{N}$ . Therefore, taking  $M$  sufficiently large so that  $K/M < \epsilon$ , we obtain that  $\mathbb{P}(N_n(S) > M) < \epsilon$ , for all  $n \in \mathbb{N}$  and, therefore, C.3 holds.

We next show that (i) implies C.4, which can be restated here in the following form: given  $\epsilon > 0$ , there exists a compact set  $C$  such that  $\mathbb{P}(N_n(S \setminus C) = 0) > 1 - \epsilon$ , for all  $n \in \mathbb{N}$ . We choose  $C$  so that for the limit distribution (which corresponds to that of a simple point process), we have

$$\mathbb{P}(N(S \setminus C) = 0) > 1 - \epsilon/2.$$

From (i), we have that  $\lim_{n \rightarrow \infty} \mathbb{P}(N_n(S \setminus C) = 0) = \mathbb{P}(N(S \setminus C) = 0)$ , from which we obtain that  $\mathbb{P}(N_n(S \setminus C) = 0) > 1 - \epsilon$ , for all  $n$  sufficiently large and, by taking a larger  $C$  if necessary, for all  $n \in \mathbb{N}$ , which means C.4 holds.

Since, both conditions C.3 and C.4 hold then  $N_n$  admits a weakly convergent subsequence, say  $(N_{n_\ell})_{\ell \in \mathbb{N}}$ , which converges weakly to  $\bar{N}$ . By (i), we have

$$\mathbb{P}(N(A) = 0) = \mathbb{P}(\bar{N} = 0),$$

but we still do not know if  $\bar{N}$  is simple, hence this only gives us that  $N$  and  $\bar{N}^*$  have the same distribution, where  $\bar{N}^*$  is the support point process associated to  $\bar{N}$ .

Hence, we are left to prove that both (ii) and (II) imply that  $\bar{N}$  is simple, *i.e.*,  $\bar{N} = \bar{N}^*$ , a.s.

Note that for all  $r$ , we have

$$\sum_{i=1}^{k_r} \mathbb{P}(\bar{N}(A_{ri}) \geq 2) = \lim_{\ell \rightarrow \infty} \sum_{i=1}^{k_r} \mathbb{P}(N_{n_\ell}(A_{ri}) \geq 2).$$

Assuming (ii), we obtain

$$\inf_{\mathcal{T}_r} \sum_{i=1}^{k_r} \mathbb{P}(\bar{N}(A_{ri}) \geq 2) \leq \inf_{\mathcal{T}_r} \limsup_{n \rightarrow \infty} \sum_{i=1}^{k_r} \mathbb{P}(N_n(A_{ri}) \geq 2) = 0.$$

This means that the point process  $\bar{N}$  is ordinary, as defined in [18, Definition 9.3.XI], and, therefore, by [18, Proposition 9.3.XII],  $\bar{N}$  is simple.

Now, assuming (II) instead, we have for all bounded  $A \in \mathcal{R}$ ,

$$\mathbb{E}(N(A)) = \mathbb{E}(\bar{N}^*(A)) \leq \mathbb{E}(\bar{N}(A)) = \lim_{\ell \rightarrow \infty} \mathbb{E}(N_{n_\ell}(A)) = \mathbb{E}(N(A)),$$

which means that  $\mathbb{E}(\bar{N}^*(A)) = \mathbb{E}(\bar{N}(A))$  for all bounded  $A \in \mathcal{R}$ , which implies that  $\bar{N}$  must be simple.  $\square$

*Remark C.4.* In fact, since given any sequence of positive random variables  $(X_n)_{n \in \mathbb{N}}$  converging in distribution to  $X$ , we have  $\liminf_{n \rightarrow \infty} \mathbb{E}(X_n) \geq \mathbb{E}(X)$ , then we can strengthen Proposition C.2 by replacing (II) by the condition  $\limsup_{n \rightarrow \infty} \mathbb{E}(N_n(A)) \leq \mathbb{E}(N(A))$ , which actually only need verifying for all bounded sets in semi-ring generating  $\mathcal{R}$ .

We consider now the particular case where  $\mathcal{X} = \mathbb{R}_0^+ \times \tilde{l}_\infty \setminus \{\tilde{\infty}\}$ . In this space, we consider the ring  $\mathcal{R}$  and its subclass of sets  $\mathcal{I}$  defined by:

$$\mathcal{R} := \left\{ \bigcup_{\ell=1}^m J_\ell \times \tilde{A}_\ell : m \in \mathbb{N}, J_\ell = [a_\ell, b_\ell), \text{ and } \tilde{A}_\ell \in \tilde{\mathcal{R}} \right\}; \quad (\text{C.5})$$

$$\mathcal{I} := \left\{ \bigcup_{\ell=1}^m J_\ell \times \tilde{A}_\ell : m \in \mathbb{N}, J_\ell = [a_\ell, b_\ell), \text{ and } \tilde{A}_\ell \in \tilde{\mathcal{J}} \right\}. \quad (\text{C.6})$$

**Lemma C.5.** *The conclusion of Proposition C.2 holds if the conditions are verified only for all bounded sets of  $\mathcal{I}$ .*

See related results in [49] (for example Theorem 2.2).

*Proof.* We focus on the condition (i), which is the strongest. The others follow easily as one can already guess from Remark C.4. The statement follows from the fact that  $\tilde{\mathcal{J}}$  is closed for unions and then using the inclusion exclusion formula, one can write the probability involving any set  $\tilde{A}_\ell \in \tilde{\mathcal{R}}$  using the probability of events involving only elements of  $\tilde{\mathcal{J}}$ .

For example, assume that  $\tilde{A}_1 = \tilde{B}_1 \cap \tilde{D}_1$ , where  $\tilde{B}_1, \tilde{D}_1 \in \tilde{\mathcal{J}}$ . Observe that

$$\begin{aligned} \mathbb{P}(N(\cup_{\ell=1}^m J_\ell \times \tilde{A}_\ell) = 0) &= \mathbb{P}\left(N(J_1 \times (\tilde{B}_1 \cap \tilde{D}_1)) = 0, N(\cup_{\ell=1}^m J_\ell \times \tilde{A}_\ell) = 0\right) \\ &= \mathbb{P}\left(N(J_1 \times \tilde{B}_1) = 0, N(\cup_{\ell=1}^m J_\ell \times \tilde{A}_\ell) = 0\right) \\ &\quad + \mathbb{P}\left(N(J_1 \times \tilde{D}_1) = 0, N(\cup_{\ell=1}^m J_\ell \times \tilde{A}_\ell) = 0\right) \\ &\quad - \mathbb{P}\left(N(J_1 \times (\tilde{B}_1 \cup \tilde{D}_1)) = 0, N(\cup_{\ell=1}^m J_\ell \times \tilde{A}_\ell) = 0\right), \end{aligned}$$

which means that all events on the 3 last terms correspond to the value of the avoidance function over sets of  $\mathcal{I}$ , as we wanted. In the same way, if  $\tilde{A}_1 = \tilde{B}_1 \setminus \tilde{D}_1$ , for  $\tilde{B}_1, \tilde{D}_1 \in \tilde{\mathcal{J}}$ , for example, we could use the fact that  $\mathbb{P}(N(\tilde{A}_1) = 0) = \mathbb{P}(N(\tilde{B}_1 \cup \tilde{D}_1) = 0) - \mathbb{P}(N(\tilde{D}_1) = 0)$  and obtain a similar formula for the avoidance function using only sets in  $\mathcal{I}$  as we did above.

Hence, we have just shown that we can handle unions, intersections and exclusions of sets of  $\tilde{\mathcal{J}}$ . Noting that  $\tilde{A}_\ell \in \tilde{\mathcal{R}}$  can always be written by using a finite number of these set operations involving elements of  $\tilde{\mathcal{J}}$ , the conclusion follows.  $\square$

#### APPENDIX D. COMPUTATION OF THE TRANSFORMED ANCHORED TAIL PROCESSES FOR THE DYNAMICAL EXAMPLES

Section 2.2 gave concrete dynamical examples where the theory in this paper holds. As noted in the discussion after Theorem 4.1, previous work shows that many of the required conditions hold. In this section we fill in the missing proofs by showing that the transformed anchored tail process is well-defined for observables whose norm is maximised at periodic points. Note that the existence of the limit in (3.15) follows from a simpler version of the ideas here.

**D.1. Applications to observables maximised at hyperbolic periodic points for general 1D systems.** Recall that we are dealing with an acip  $\mu$  and that for our periodic point  $\zeta$  we are always assuming that the density exists and is bounded, say  $\frac{d\mu}{d\text{Leb}} = D \in (0, \infty)$ . For definiteness assume that we are dealing with a non-invertible map, the observable is as in (2.6),  $\mathcal{M} = \{\zeta\}$  and  $\zeta$  is a repelling fixed point (therefore  $p = 1$ ). In this case the process  $(Y_j)_{j \in \mathbb{Z}}$  is such that, for some  $s \in \mathbb{N}_0$ , we have  $Y_j = \infty$ , for all  $j \leq -s$  and  $Y_j = U \cdot \|(DT_\zeta)^j(\Theta)\|^d \frac{(DT_\zeta)^j(\Theta)}{\|(DT_\zeta)^j(\Theta)\|}$ , for all  $j > -s$ , where  $U$  is a uniformly distributed random variable on the interval  $[0, 1]$  independent of  $\Theta$ , which has a uniform distribution on  $\mathbb{S}^{d-1}$  and  $(DT)_\zeta^i$  denotes the  $i$ -fold product of the derivative of  $T$  at  $\zeta$  (which is invertible since we are assuming that  $\zeta$  is repelling).

This clearly satisfies (2)–(4) for the transformed anchored tail process, so we need to show (1), i.e.,

$$\mathcal{L}\left(\frac{1}{\tau} \mathbb{X}_n^{r_n+s, r_n+t} \mid \|\mathbf{X}_{r_n}\| > u_n(\tau)\right) \xrightarrow{n \rightarrow \infty} \mathcal{L}((Y_j)_{j=s, \dots, t}),$$

for all  $s < t \in \mathbb{Z}$  and all  $\tau > 0$ .

Without loss of generality we assume that  $g$  is positive in a neighbourhood of  $\zeta$ , so  $\|\mathbf{X}_{r_n}(x)\| > u_n(\tau)$  can be written  $g(\text{dist}(T^{r_n}(x), \zeta)) > u_n(\tau)$ . From (2.15) we see that asymptotically  $u_n(\tau) \sim g\left(\left(\frac{\tau}{Cn}\right)^{1/d}\right)$ , where  $C = DL_\zeta$  for  $L_\zeta = \lim_{r \rightarrow 0} \frac{\text{Leb}(B_r(\zeta))}{r}$  (this exists in the cases considered here). In this section we will assume that this is an equality since all our estimates are asymptotic in  $n$ , so similarly  $u_n^{-1}(z) = Cn(g^{-1}(z))^d$ , where we will also assume this is well-defined. Since we are assuming that  $\|\mathbf{X}_{r_n}(x)\| > u_n(\tau)$ , let  $v < 1$  be such that  $\|\mathbf{X}_{r_n}(x)\| = u_n(v\tau)$ , which translates as  $\text{dist}(T^{r_n}(x), \zeta) = \left(\frac{v\tau}{Cn}\right)^{1/d}$ . Using the linearisation domain around the fixed point  $\zeta$  from Hartman-Grobman theory,  $\text{dist}(T^{r_n+k}(x), \zeta) \sim \left(\frac{v\tau}{Cn}\right)^{1/d} \|(DT_\zeta)^k(w)\|$ , where  $w = \frac{\Phi_\zeta^{-1}(T^{r_n}(x))}{\|\Phi_\zeta^{-1}(T^{r_n}(x))\|}$  and  $\Phi_\zeta$  is as in (2.6). It follows that

$$\mathbf{X}_{r_n+k}(x) \sim g\left(\left(\frac{v\tau}{Cn}\right)^{1/d} \|(DT_\zeta)^k(w)\|\right) \frac{(DT_\zeta)^k(w)}{\|(DT_\zeta)^k(w)\|},$$

from which we find

$$\frac{u_n^{-1}(\|\mathbf{X}_{r_n+k}(x)\|)}{\tau} \frac{\mathbf{X}_{r_n+k}(x)}{\|\mathbf{X}_{r_n+k}(x)\|} \sim v \|(DT_\zeta)^k(w)\|^d \frac{(DT_\zeta)^k(w)}{\|(DT_\zeta)^k(w)\|}.$$

All of these asymptotics become equalities as  $n \rightarrow \infty$ . The fact that  $T^{r_n}(x)$  is chosen according to the invariant measure  $\mu$  on the small neighbourhood  $B_{g^{-1}(u_n(\tau))}(\zeta) = \{\|\mathbf{X}_{r_n}(x)\| > u_n(\tau)\}$  of  $\zeta$  and the fact that  $\mu$  behaves asymptotically like Lebesgue measure on small neighbourhoods of  $\zeta$  give that the finite  $Y_j$ s are indeed of the form required. The infinite terms appear because, due to the repelling nature of  $\zeta$ , the entrance of the orbit of  $x$  at time  $r_n$  in  $B_{g^{-1}(u_n(\tau))}(\zeta)$  can only be preceded by a finite number,  $s \in \mathbb{N}_0$ , of consecutive hits to  $B_{g^{-1}(u_n(\tau))}(\zeta)$ . But then  $T^{r_n-s-1}(x)$ , for example, must belong to neighbourhood of  $T^{-1}(\zeta) \setminus \{\zeta\}$ , which are at a fixed distance from  $\zeta$ : let  $c = \text{dist}(\zeta, T^{-1}(\zeta) \setminus \{\zeta\}) > 0$ . Then for  $n$  sufficiently large we have  $\text{dist}(T^{r_n-s-1}(x), \zeta) > c/2$  and hence  $u_n^{-1}(\|\mathbf{X}_{r_n-s-1}\|) \sim Cn(c/2)^d$ , which clearly diverges to  $\infty$  as  $n \rightarrow \infty$ .

In the invertible case, for the toral diffeomorphisms considered, we have  $T^{-1}(\zeta) = \zeta$  and hence for all  $j \in \mathbb{Z}$ , we have  $Y_j = U \cdot \|(DT_\zeta)^j(\Theta)\|^d \frac{(DT_\zeta)^j(\Theta)}{\|(DT_\zeta)^j(\Theta)\|}$ , where  $U$  and  $\Theta$  are as above.

We observe that if instead  $\zeta$  were a periodic point of period  $q > 1$  then elements of  $Y$  would be the same as for the fixed point case, but with  $q-1$   $\infty$  terms between the entries  $qk$  and  $q(k+1)$ , for all  $k \in \mathbb{Z}$ .

Now that we understand  $(Y_j)_j$  it is easy to see that the condition

$$\mathbb{P}(\text{all finite } Q_j \text{'s are mutually different}) = 1$$

in Theorem 2.12 is also trivially satisfied in these examples as it can be interpreted as  $\zeta$  being hyperbolic.

In the one-dimensional case, concerning most of the examples given in the noninvertible scenario in Section 2.2, if we consider  $\zeta$  as a repelling fixed point such that  $|DT_\zeta| = \chi > 1$  then, in this case,  $(Y_j)_j$  assumes the simpler form:

$$(\dots, \infty, \infty, U\chi^{-s}, \dots, U\chi^{-1}, U, U\chi, U\chi^2, \dots)$$

Finally, to obtain the transformed anchored tail process we need to condition the  $Y_j$  on the event  $\inf_{j \leq -1} \|Y_j\| \geq 1$ , which is to say that in the computation of the  $Y_j$ , the exceedance occurring at time  $r_n$  is indeed the first exceedance of the cluster, which means  $s = 0$  and therefore  $(Z_j)_j$  is of the form

$$(\dots, \infty, \infty, U, U\chi, U\chi^2, \dots)$$

**D.2. Particular application with overshooting.** We consider now the application in Example 2.7. We start by noting that  $\{|\psi(x)| > u\} = B_{1/8}(\varepsilon_u) \cup B_{3/8}(\tilde{\varepsilon}_u)$ , where  $\varepsilon_u, \tilde{\varepsilon}_u \sim u^{-1/2}$ , which means that  $\mu(|\psi(x)| > u) \sim 4u^{-1/2}$ . Having (2.15) in mind, we take  $u_n(\tau) = \frac{16n^2}{\tau^2}$  and consequently  $u_n^{-1}(z) = 4nz^{-1/2}$  (note that, as we did earlier, we write equalities since all our estimates are asymptotic in  $n$ ). If  $x \in B_{1/8}(\varepsilon_{u_n(\tau)})$  (recall that  $\varepsilon_{u_n(\tau)} \sim \tau/4n$ ) then  $u_n^{-1}(|\psi(x)|) \sim 4n|x-1/8|$ , while if  $x \in B_{3/8}(\varepsilon_{u_n(\tau)})$  then  $u_n^{-1}(|\psi(x)|) \sim 4n|x-3/8|$ . For simplicity, let  $0 < v < 1$  be such that  $|X_{r_n}| = u_n(v\tau)$ , which translates as  $|T^{r_n}(x) - 1/8| \sim v\tau/4n$  or  $|T^{r_n}(x) - 3/8| \sim v\tau/4n$ . Now, recalling that  $\{1/8, 3/8\}$  corresponds to a period two orbit,  $T$  triples distances and  $\psi$  has opposite signs on the neighbourhood of these points, then

$$\frac{u_n^{-1}(|X_{r_n+k}(x)|)}{\tau} \frac{X_{r_n+k}(x)}{|X_{r_n+k}(x)|} \sim v(-1)^s(-3)^k,$$

where  $s = 0$  if  $T^{r_n}(x) \in B_{1/8}(\varepsilon_{u_n(\tau)})$  and  $s = 1$  if  $T^{r_n}(x) \in B_{3/8}(\tilde{\varepsilon}_{u_n(\tau)})$ . All of these asymptotics become equalities as  $n \rightarrow \infty$ . The fact that  $T^{r_n}(x)$  is chosen according to the invariant measure  $\mu = \text{Leb}$  on the set  $B_{1/8}(\varepsilon_{u_n(\tau)}) \cup B_{3/8}(\tilde{\varepsilon}_{u_n(\tau)})$  and  $\lim_{n \rightarrow \infty} \frac{\mu(B_{1/8}(\varepsilon_{u_n(\tau)}))}{\mu(|X_0| > u_n(\tau))} = \frac{1}{2} = \lim_{n \rightarrow \infty} \frac{\mu(B_{3/8}(\tilde{\varepsilon}_{u_n(\tau)}))}{\mu(|X_0| > u_n(\tau))}$  gives that  $(Z_j)_{j \in \mathbb{Z}}$  is indeed of the form required, namely,

$$(\dots, \infty, \infty, UE, U(-3)E, U(-3)^2E, \dots),$$

where  $U$  is uniformly distributed on  $[0, 1]$  and  $\mathbb{P}(E = 1) = 1/2 = \mathbb{P}(E = -1)$ .

## REFERENCES

- [1] M. Abadi, A. C. M. Freitas, and J. M. Freitas. Dynamical counterexamples regarding the extremal index and the mean of the limiting cluster size distribution. *J. Lond. Math. Soc. (2)*, 102(2):670–694, 2020.
- [2] J. F. Alves. *Nonuniformly hyperbolic attractors*. Springer Monographs in Mathematics. Springer, Cham, 2020. Geometric and probabilistic aspects.
- [3] J. F. Alves, J. M. Freitas, S. Luzzatto, and S. Vaienti. From rates of mixing to recurrence times via large deviations. *Adv. Math.*, 228(2):1203–1236, 2011.
- [4] D. Azevedo, A. C. M. Freitas, J. M. Freitas, and F. B. Rodrigues. Clustering of extreme events created by multiple correlated maxima. *Phys. D*, 315:33–48, 2016.
- [5] D. Azevedo, A. C. M. Freitas, J. M. Freitas, and F. B. Rodrigues. Extreme Value Laws for Dynamical Systems with Countable Extremal Sets. *J. Stat. Phys.*, 167(5):1244–1261, 2017.
- [6] P. Bálint and I. Melbourne. Decay of correlations and invariance principles for dispersing billiards with cusps, and related planar billiard flows. *J. Stat. Phys.*, 133(3):435–447, 2008.
- [7] B. Basrak, D. Krizmanić, and J. Segers. A functional limit theorem for dependent sequences with infinite variance stable limits. *Ann. Probab.*, 40(5):2008–2033, 2012.
- [8] B. Basrak and H. Planinić. Compound Poisson approximation for regularly varying fields with application to sequence alignment. *Bernoulli*, 27(2):1371–1408, 2021.
- [9] B. Basrak, H. Planinić, and P. Soulier. An invariance principle for sums and record times of regularly varying stationary sequences. *Probab. Theory Related Fields*, 172(3-4):869–914, 2018.
- [10] B. Basrak and J. Segers. Regularly varying multivariate time series. *Stochastic Process. Appl.*, 119(4):1055–1080, 2009.
- [11] J. Beirlant, Y. Goegebeur, J. Teugels, and J. Segers. *Statistics of extremes*. Wiley Series in Probability and Statistics. John Wiley & Sons, Ltd., Chichester, 2004. Theory and applications, With contributions from Daniel De Waal and Chris Ferro.
- [12] P. Billingsley. *Convergence of probability measures*. Wiley Series in Probability and Statistics: Probability and Statistics. John Wiley & Sons Inc., New York, second edition, 1999. A Wiley-Interscience Publication.
- [13] T. Caby, D. Faranda, G. Mantica, S. Vaienti, and P. Yiou. Generalized dimensions, large deviations and the distribution of rare events. *Phys. D*, 400:132143, 15, 2019.
- [14] M. Carney, M. Holland, and M. Nicol. Extremes and extremal indices for level set observables on hyperbolic systems. *Nonlinearity*, 34(2):1136–1167, 2021.
- [15] M. Carvalho, A. C. M. Freitas, J. M. Freitas, M. Holland, and M. Nicol. Extremal dichotomy for uniformly hyperbolic systems. *Dyn. Syst.*, 30(4):383–403, 2015.



- [16] I. Chevyrev and P. K. Friz. Canonical RDEs and general semimartingales as rough paths. *Ann. Probab.*, 47(1):420–463, 2019.
- [17] D. J. Daley and D. Vere-Jones. *An introduction to the theory of point processes. Vol. I.* Probability and its Applications (New York). Springer-Verlag, New York, second edition, 2003. Elementary theory and methods.
- [18] D. J. Daley and D. Vere-Jones. *An introduction to the theory of point processes. Vol. II.* Probability and its Applications (New York). Springer, New York, second edition, 2008. General theory and structure.
- [19] R. A. Davis. Stable limits for partial sums of dependent random variables. *Ann. Probab.*, 11(2):262–269, 1983.
- [20] R. A. Davis and T. Hsing. Point process and partial sum convergence for weakly dependent random variables with infinite variance. *Ann. Probab.*, 23(2):879–917, 1995.
- [21] L. de Haan and A. Ferreira. *Extreme value theory.* Springer Series in Operations Research and Financial Engineering. Springer, New York, 2006. An introduction.
- [22] M. Denker and W. Philipp. Approximation by Brownian motion for Gibbs measures and flows under a function. *Ergodic Theory Dynam. Systems*, 4(4):541–552, 1984.
- [23] M. Dwass. Extremal processes. *Ann. Math. Statist.*, 35:1718–1725, 1964.
- [24] P. Embrechts, C. Klüppelberg, and T. Mikosch. *Modelling extremal events*, volume 33 of *Applications of Mathematics (New York)*. Springer-Verlag, Berlin, 1997. For insurance and finance.
- [25] M. Falk, J. Hüsler, and R.-D. Reiss. *Laws of small numbers: extremes and rare events.* Birkhäuser/Springer Basel AG, Basel, extended edition, 2011.
- [26] D. Faranda, H. Ghouidi, P. Guiraud, and S. Vaienti. Extreme value theory for synchronization of coupled map lattices. *Nonlinearity*, 31(7):3326–3358, 2018.
- [27] M. Field, I. Melbourne, and A. Török. Decay of correlations, central limit theorems and approximation by Brownian motion for compact Lie group extensions. *Ergodic Theory Dynam. Systems*, 23(1):87–110, 2003.
- [28] A. C. M. Freitas and J. M. Freitas. Extreme values for Benedicks-Carleson quadratic maps. *Ergodic Theory Dynam. Systems*, 28(4):1117–1133, 2008.
- [29] A. C. M. Freitas, J. M. Freitas, and M. Magalhães. Convergence of marked point processes of excesses for dynamical systems. *J. Eur. Math. Soc. (JEMS)*, 20(9):2131–2179, 2018.
- [30] A. C. M. Freitas, J. M. Freitas, and M. Magalhães. Complete convergence and records for dynamically generated stochastic processes. *Trans. Amer. Math. Soc.*, 373(1):435–478, 2020.
- [31] A. C. M. Freitas, J. M. Freitas, F. B. Rodrigues, and J. V. Soares. Rare events for Cantor target sets. *Comm. Math. Phys.*, 378(1):75–115, 2020.
- [32] A. C. M. Freitas, J. M. Freitas, and M. Todd. Hitting time statistics and extreme value theory. *Probab. Theory Related Fields*, 147(3-4):675–710, 2010.
- [33] A. C. M. Freitas, J. M. Freitas, and M. Todd. The extremal index, hitting time statistics and periodicity. *Adv. Math.*, 231(5):2626–2665, 2012.
- [34] A. C. M. Freitas, J. M. Freitas, and M. Todd. The compound Poisson limit ruling periodic extreme behaviour of non-uniformly hyperbolic dynamics. *Comm. Math. Phys.*, 321:483–527, 2013.
- [35] A. C. M. Freitas, J. M. Freitas, and M. Todd. Speed of convergence for laws of rare events and escape rates. *Stochastic Process. Appl.*, 125(4):1653–1687, 2015.
- [36] A. C. M. Freitas, J. M. Freitas, M. Todd, and S. Vaienti. Rare events for the Manneville-Pomeau map. *Stochastic Processes and their Applications*, 126(11):3463–3479, 2016.
- [37] S. Gouëzel. Central limit theorem and stable laws for intermittent maps. *Probab. Theory Related Fields*, 128(1):82–122, 2004.
- [38] S. Gouëzel. Stable laws for the doubling map. Unpublished notes available at <https://www.math.sciences.univ-nantes.fr/~gouezel/articles/DoublingStable.pdf>, 2008.
- [39] S. Gouëzel. Almost sure invariance principle for dynamical systems by spectral methods. *Ann. Probab.*, 38(4):1639–1671, 2010.
- [40] C. Gupta, M. Holland, and M. Nicol. Extreme value theory and return time statistics for dispersing billiard maps and flows, Lozi maps and Lorenz-like maps. *Ergodic Theory Dynam. Systems*, 31(5):1363–1390, 2011.
- [41] F. Hofbauer and G. Keller. Ergodic properties of invariant measures for piecewise monotonic transformations. *Math. Z.*, 180(1):119–140, 1982.
- [42] M. Holland and I. Melbourne. Central limit theorems and invariance principles for Lorenz attractors. *J. Lond. Math. Soc. (2)*, 76(2):345–364, 2007.
- [43] M. Holland, M. Nicol, and A. Török. Extreme value theory for non-uniformly expanding dynamical systems. *Trans. Amer. Math. Soc.*, 364(2):661–688, 2012.
- [44] M. Holland and M. Todd. Weak convergence to extremal processes and record events for non-uniformly hyperbolic dynamical systems. *Ergodic Theory Dynam. Systems*, 39(4):980–1001, 2019.

- [45] T. Hsing. On the characterization of certain point processes. *Stochastic Process. Appl.*, 26(2):297–316, 1987.
- [46] H. Hu. Decay of correlations for piecewise smooth maps with indifferent fixed points. *Ergodic Theory Dynam. Systems*, 24(2):495–524, 2004.
- [47] P. Jung, I. Melbourne, F. Pène, P. Varandas, and H.-K. Zhang. Necessary and sufficient condition for  $\mathcal{M}_2$ -convergence to a Lévy process for billiards with cusps at flat points. *Stoch. Dyn.*, 21(5):Paper No. 2150024, 8, 2021.
- [48] P. Jung, F. Pène, and H.-K. Zhang. Convergence to  $\alpha$ -stable Lévy motion for chaotic billiards with several cusps at flat points. *Nonlinearity*, 33(2):807–839, 2020.
- [49] O. Kallenberg. *Random measures, theory and applications*, volume 77 of *Probability Theory and Stochastic Modelling*. Springer, Cham, 2017.
- [50] R. Kulik and P. Soulier. *Heavy-tailed time series*. Springer Series in Operations Research and Financial Engineering. Springer, New York, 2020.
- [51] J. Lamperti. On extreme order statistics. *Ann. Math. Statist.*, 35:1726–1737, 1964.
- [52] M. R. Leadbetter. Extremes and local dependence in stationary sequences. *Z. Wahrsch. Verw. Gebiete*, 65(2):291–306, 1983.
- [53] M. R. Leadbetter, G. Lindgren, and H. Rootzén. *Extremes and related properties of random sequences and processes*. Springer Series in Statistics. Springer-Verlag, New York, 1983.
- [54] R. LePage, M. Woodroffe, and J. Zinn. Convergence to a stable distribution via order statistics. *Ann. Probab.*, 9(4):624–632, 1981.
- [55] C. Liverani, B. Saussol, and S. Vaienti. A probabilistic approach to intermittency. *Ergodic Theory Dynam. Systems*, 19(3):671–685, 1999.
- [56] V. Lucarini, D. Faranda, A. C. M. Freitas, J. M. Freitas, M. Holland, T. Kuna, M. Nicol, M. Todd, and S. Vaienti. *Extremes and Recurrence in Dynamical Systems*. Pure and Applied Mathematics: A Wiley Series of Texts, Monographs and Tracts. Wiley, Hoboken, NJ, 2016.
- [57] G. Mantica and L. Perotti. Extreme value laws for fractal intensity functions in dynamical systems: Minkowski analysis. *J. Phys. A*, 49(37):374001, 21, 2016.
- [58] I. Melbourne and M. Nicol. Almost sure invariance principle for nonuniformly hyperbolic systems. *Comm. Math. Phys.*, 260(1):131–146, 2005.
- [59] I. Melbourne and M. Nicol. A vector-valued almost sure invariance principle for hyperbolic dynamical systems. *Ann. Probab.*, 37(2):478–505, 2009.
- [60] I. Melbourne and A. Török. Central limit theorems and invariance principles for time-one maps of hyperbolic flows. *Comm. Math. Phys.*, 229(1):57–71, 2002.
- [61] I. Melbourne and P. Varandas. Convergence to a Lévy process in the Skorohod  $\mathcal{M}_1$  and  $\mathcal{M}_2$  topologies for nonuniformly hyperbolic systems, including billiards with cusps. *Comm. Math. Phys.*, 375(1):653–678, 2020.
- [62] I. Melbourne and R. Zweimüller. Weak convergence to stable Lévy processes for nonuniformly hyperbolic dynamical systems. *Ann. Inst. Henri Poincaré Probab. Stat.*, 51(2):545–556, 2015.
- [63] G. Messori, R. Caballero, F. Bouchet, D. Faranda, R. Grotjahn, N. Harnik, S. Jewson, J. G. Pinto, G. Rivière, T. Woollings, and P. Yiou. An interdisciplinary approach to the study of extreme weather events: Large-scale atmospheric controls and insights from dynamical systems theory and statistical mechanics. *Bulletin of the American Meteorological Society*, 99(5):ES81–ES85, 2018.
- [64] G. Messori, R. Caballero, and D. Faranda. A dynamical systems approach to studying midlatitude weather extremes. *Geophysical Research Letters*, 44(7):3346–3354, 2017.
- [65] T. Mori. Limit distributions of two-dimensional point processes generated by strong-mixing sequences. *Yokohama Math. J.*, 25(2):155–168, 1977.
- [66] S. Y. Novak. Multilevel clustering of extremes. *Stochastic Process. Appl.*, 97(1):59–75, 2002.
- [67] G. L. O’Brien. Extreme values for stationary and Markov sequences. *Ann. Probab.*, 15(1):281–291, 1987.
- [68] F. Pène and B. Saussol. Spatio-temporal Poisson processes for visits to small sets. *Israel J. Math.*, 240(2):625–665, 2020.
- [69] J. Pickands, III. The two-dimensional Poisson process and extremal processes. *J. Appl. Probability*, 8:745–756, 1971.
- [70] Y. Pomeau and P. Manneville. Intermittent transition to turbulence in dissipative dynamical systems. *Comm. Math. Phys.*, 74(2):189–197, 1980.
- [71] S. I. Resnick. Weak convergence to extremal processes. *Ann. Probability*, 3(6):951–960, 1975.
- [72] S. I. Resnick. *Extreme values, regular variation, and point processes*, volume 4 of *Applied Probability. A Series of the Applied Probability Trust*. Springer-Verlag, New York, 1987.
- [73] K.-i. Sato. *Lévy processes and infinitely divisible distributions*, volume 68 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 2013. Translated from the 1990 Japanese original, Revised edition of the 1999 English translation.

- [74] E. W. Saw, D. Kuzzay, D. Faranda, A. Guittonneau, F. Daviaud, C. Wiertel-Gasquet, V. Padilla, and B. Dubrulle. Experimental characterization of extreme events of inertial dissipation in a turbulent swirling flow. *Nature Communications*, 7:12466, 08 2016.
- [75] J. Segers. Approximate distributions of clusters of extremes. *Statist. Probab. Lett.*, 74(4):330–336, 2005.
- [76] R. L. Smith. The extremal index for a Markov chain. *J. Appl. Probab.*, 29(1):37–45, 1992.
- [77] M. Tyran-Kamińska. Convergence to Lévy stable processes under some weak dependence conditions. *Stochastic Process. Appl.*, 120(9):1629–1650, 2010.
- [78] M. Tyran-Kamińska. Weak convergence to Lévy stable processes in dynamical systems. *Stoch. Dyn.*, 10(2):263–289, 2010.
- [79] W. Whitt. *Stochastic-process limits*. Springer Series in Operations Research. Springer-Verlag, New York, 2002. An introduction to stochastic-process limits and their application to queues.
- [80] L.-S. Young. Statistical properties of dynamical systems with some hyperbolicity. *Ann. of Math. (2)*, 147(3):585–650, 1998.
- [81] L.-S. Young. Recurrence times and rates of mixing. *Israel J. Math.*, 110:153–188, 1999.

ANA CRISTINA MOREIRA FREITAS, CENTRO DE MATEMÁTICA & FACULDADE DE ECONOMIA DA UNIVERSIDADE DO PORTO, RUA DR. ROBERTO FRIAS, 4200-464 PORTO, PORTUGAL

*Email address:* amoreira@fep.up.pt

*URL:* <http://www.fep.up.pt/docentes/amoreira/>

JORGE MILHAZES FREITAS, CENTRO DE MATEMÁTICA & FACULDADE DE CIÊNCIAS DA UNIVERSIDADE DO PORTO, RUA DO CAMPO ALEGRE 687, 4169-007 PORTO, PORTUGAL

*Email address:* jmfreira@fc.up.pt

*URL:* <http://www.fc.up.pt/pessoas/jmfreira/>

MIKE TODD, MATHEMATICAL INSTITUTE, UNIVERSITY OF ST ANDREWS, NORTH HAUGH, ST ANDREWS, KY16 9SS, SCOTLAND,

*Email address:* m.todd@st-andrews.ac.uk

*URL:* <https://mtoddm.github.io/>