The Tail Measure on the Space of Càdlàg Processes

Felix Konrad

26th March 2023

Contents

1 Introduction

A classic problem of univariate extreme value theory is to find conditions on random variables X such that there exist sequences $(a_n)_{n\in\mathbb{N}}$ and $(b_n)_{n\in\mathbb{N}}$ for which the normalized maxima converge weakly, meaning

$$\mathbb{P}\left[\frac{\max\{X_1,\dots,X_n\}-b_n}{a_n}\leq t\right]\to G(t)$$

for a suitable distribution function G and i.i.d. copies X_i of X. In this case we say that X is in the domain of attraction of G.
If we assume that G is non-trivial, the Fisher-Tippett-Gnedenko theorem gives us three classes of possible limit distributions G; the Gumbel, Fréchet and Weibull distributions.
The domain of attraction of the Fréchét distribution consists of so-called *regularly varying* random variables: We call a function $f:(0,\infty)\to\mathbb{R}$ regularly varying with index $-\alpha$ if for all $x>0$ we have

$$\lim_{t\to\infty}\frac{f(tx)}{f(t)}=x^{-\alpha}$$

and call a random variable X regularly varying with index α if the function $\mathbb{P}[|X|>t]$ is a regularly varying function with index $-\alpha$ and additionally

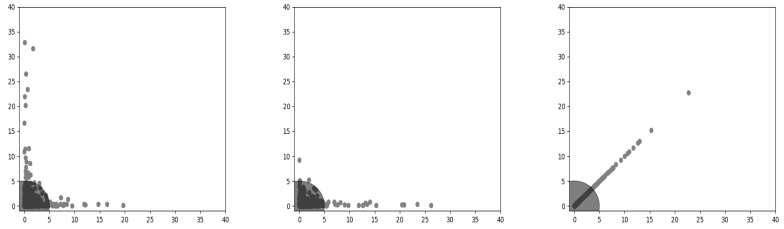
$$\lim_{t\to\infty}\frac{\mathbb{P}[X>tx]}{\mathbb{P}[|X|>t]}=p_X\in[0,1].$$

The main goal of this thesis is to present how the concept of regular variation is extended to multivariate stochastic processes in continuous time in order to study the extremal behaviour of such processes, which will be the topic of Section 4. The main problem when considering such processes is that it does not suffice to merely consider each time point t and each component $X_t^{(i)}$, since this would neglect two important factors: Firstly, there might be a spatial dependence structure between the components of the multivariate process, meaning $X^{(i)}$ and $X^{(j)}$ are not independent. Describing such a dependence structure for a multivariate random vector will be the topic of Section 2. Secondly, there might be a temporal dependence structure, meaning X_s and X_t might be dependent for $s\neq t$ and thus considering them seperately neglects this dependency.
In Section 3 we define a suitable topology for càdlàg processes, which allows us to define stochastic processes with càdlàg paths and study their convergence.
In Section 4, primarily Section 4.1, we will present how a single process, called a *tail process*, decribes both the temporal dependence structure of a process as well as the spatial dependence structure of the process's components. This chapter will primarily be based on the recent work of Soulier [Sou22].
Finally, in Section 5 we will apply some of the results of Section 4 to Markov

processes and extend known results of regularly varying Markov chains to the continuous-time setting. In particular, we show how to compute the forward tail process using the semigroup or generator of a regularly varying Markov process.

2 Multivariate Extreme Value Theory

In this section we will define the concepts motivated in the intoduction in the context of random vectors. One additional difficulty that we face when considering random vectors instead of univariate random variables is that we not only have to consider how heavy-tailed our variables are, but also whether the individual components are dependent on each other. The exponent measure, introduced in Definition 2.2, will both tell us how heavy-tailed the components are and what the dependence structure of extreme values looks like. To get some intuition about what one might expect such a measure to look like, we take a look at two simple cases:
In the figure below sample plots of three random vectors are shown. The first vector $X = (X_1, X_2)$ has identical and independent $\mathrm{Par}(\alpha)$-distributed components, meaning $\mathbb{P}[X_i > t] = t^{-\alpha}$ for $t > 1$.
The second vector $Y = (Y_1, Y_2)$ has independent components with $Y_1 \sim \mathrm{Par}(\alpha)$ and $Y_2 \sim \mathrm{Par}(\beta)$ where $\alpha < \beta$.
The third vector $Z = (Z_1, Z_1)$ has identical components with $Z_1 \sim \mathrm{Par}(\alpha)$.
As one can see, at most one component of X will be large in any sample. Heuristically, a large value of $\|X\|$ is caused by exactly one component being large, but rarely by multiple components being large simultaneously. This is known as the "single big jump heuristic", see [KS20], Example 2.1.4.
Note that while both X and Y have independent components, their extremal behaviour is different, since for Y the component Y_1 seems to dominate.
Comparing the extremal behaviour of X and Z shows why it does not suffice to merely consider the extremal behaviour of each marginal distribution of a random vector, since in the case of independent components at most one component will be large while in the latter case of identical components all components will be large simultaneously, even though their marginal distributions are the same.
As we will see in Example 2.4, the exponent measure of X is concentrated on the coordinate axis', the exponent measure of Y on the x_1-axis and the exponent measure of Z on the diagonal.



(a) X (b) Y (c) Z

2.1 Regular Variation

In one dimension, the concept of regular variation naturally arises when considering the domain of attraction of a Frechet distribution. We want to extend this to maxima of multivariate random vectors. There are many equivalent approaches to defining the concept of regular variation of multivariate random variables. We will mainly follow the approach using the so-called *vague convergence* of measures, which extends the concept of weak convergence to infinite measures, but on a restricted set system:

Definition 2.1. Vague Convergence ([KS20] Definition 2.1.1)
Let $\mathbb{R}^d$ be endowed with an arbitrary norm $\|\cdot\|$. We call a set $A \subseteq \mathbb{R}^d$ bounded away from 0 in $\mathbb{R}^d$ if there exists an open set $U \subseteq \mathbb{R}^d$ containing 0 such that $A \subseteq U^C$.
We call the set $\mathcal{B}_0$ of all sets seperated from 0 a boundedness.
A measure ν on $\mathbb{R}^d \setminus \{0\}$ is said to be $\mathcal{B}_0$-boundedly finite if for all measurable $A \in \mathcal{B}_0$ we have $\nu(A) < \infty$.
A sequence $(\nu_n)_{n\in\mathbb{N}}$ of $\mathcal{B}_0$-boundedly finite measures is said to converge vaguely to a $\mathcal{B}_0$-boundedly finite measure ν on $\mathbb{R}^d \setminus \{0\}$ if for all ν-continuity sets A in $\mathcal{B}_0$, i.e. sets with $\nu(\partial A) = 0$, we have

$$\lim_{n\to\infty} \nu_n(A) = \nu(A).$$

Essentially, vague convergence means weak convergence on $\mathcal{B}_0$, though a boundedness $\mathcal{B}_0$ is not a σ-algebra and therefore the term "weak convergence" is technically incorrect.
The choice of the norm $\|\cdot\|$ is arbitrary since all norms on $\mathbb{R}^d$ are equivalent and thus induce the same topology and the same boundedness. Throughout this Section we let $\|\cdot\|$ be an arbitrary norm on $\mathbb{R}^d$.

To motivate the Definition of regular variation for multivariate random variables, consider again the definition of regular variation of univariate random variables X: We say X is regularly varying if its tail function satisfies

$$\frac{\mathbb{P}[|X| > tx]}{\mathbb{P}[|X| > x]} \to t^{-\alpha}, \qquad x \to \infty$$

for all $t > 0$. If we choose a sequence a_n such that $\mathbb{P}[|X| > a_n] \sim n^{-1}$, assuming such a sequence exists, we obtain

$$n \cdot \mathbb{P}[X > a_n t] \to t^{-\alpha}, \qquad n \to \infty.$$

This leads us to the following definition:

Definition 2.2. Multivariate Regular Variation ([KS20] Definition 2.1.2)
Let X be a d-dimensional random vector. We call X multivariate regularly

varying if

$$n\mathbb{P}[a_n^{-1}X \in A] \to \nu(A)$$

converges $\mathcal{B}_0$-vaguely to a non-zero $\mathcal{B}_0$-boundedly finite measure ν on $\mathbb{R}^d \setminus \{0\}$.

By this definition we can immediately derive a limit theorem for the maximum of i.i.d. copies of X, see [KS20], Theorem 2.1.6:
Let $(X^{(i)})_{i\in\mathbb{N}}$ be a sequence of i.i.d. copies of X where X is non-negative and regularly varying with exponent measure ν and scaling sequence $(a_n)_{n\in\mathbb{N}}$. Define the maximum of two vectors $x, y \in \mathbb{R}^d$ as the vector of componentwise maxima, meaning $(x \vee y)_i := (x_i \vee y_i)$. Additionally we define the partial order $\preceq$ by saying $x \preceq y$ if and only if $x_i \leq y_i$ for all $i = 1, \ldots, d$. Then for $u \in \mathbb{R}^d$ with $u \succeq 0$ we have

$$\begin{aligned}\mathbb{P}\left[a_n^{-1}\bigvee_{i=1}^{n} X^{(i)} \preceq u\right] &= \mathbb{P}[a_n^{-1}X^{(1)} \preceq u, \ldots, a_n^{-1}X^{(n)} \preceq u] = \mathbb{P}[a_n^{-1}X^{(1)} \preceq u]^n \\ &= \left(1 - \frac{1}{n}n\mathbb{P}\big[X^{(1)} \in [0,u]^C\big]\right)^n \xrightarrow[n\to\infty]{} \exp(-\nu([0,u]^C))\end{aligned}$$

This limit theorem also justifies the term "exponent measure". Note that in the one-dimensional case the limiting distribution in the Fréchet case was given by $\exp(-x^{-\alpha})$, so in this case the exponent measure was given by $\nu([x,\infty)) = x^{-\alpha}$ for $x > 0$. Therefore we can see that in one dimension we have a single index α fully determining the limit distribution of the maxima, namely the Fréchet distribution. Using the definition above, it is not immediately clear whether we have an analogous index α for the tail behaviour of the vector X. Yet there is an equivalent definition of regular variation which involves the tail index α:

Proposition 2.3. *([KS20] Proposition 2.1.3)*
A random vector X is multivariate regularly varying with exponent measure ν, as in Definition 2.2, if and only if there exists a constant $\alpha > 0$ and a function $g : (0,\infty) \to (0,\infty)$ which is regularly varying with index α such that

$$g(t)\mathbb{P}[t^{-1}X \in A] \to \nu(A)$$

converges $\mathcal{B}_0$-vaguely.
The constant α is called the tail index of X.
Furthermore the sequence (a_n) in definition 2.2 is regularly varying with index $1/\alpha$ and the limit measure ν is homogeneous:

$$\forall t > 0 : \nu(tA) = t^{-\alpha}\nu(A)$$

for all Borel-measurable $A \subseteq \mathbb{R}^d$.
If ν and ν' are exponent measures of X with scaling sequences (a_n) and (a'_n) as in Definition 2.2 we have $\nu = \zeta\nu'$ where $\zeta = (\lim a'_n/a_n)^\alpha$.

In other sources, regular variation of random vectors is often defined using this equivalent characterization with the specific choice of $g(t) = \mathbb{P}[\|X\| > t]$. This choice is justified by the fact that if X is multivariate regularly varying with index α, then $\|X\|$ is univariate regularly varying with index α: Let $B_r(x) := \{z \in \mathbb{R}^d : \|z - x\| \leq r\}$, then

$$\frac{\mathbb{P}[\|X\| > t \cdot a_n]}{\mathbb{P}[\|X\| > a_n]} = \frac{n\mathbb{P}\big[a_n^{-1}X \in \overline{B_t(0)}^C\big]}{n\mathbb{P}\big[a_n^{-1}X \in \overline{B_1(0)}^C\big]} \to \frac{\nu\big(\overline{B_t(0)}^C\big)}{\nu\big(\overline{B_1(0)}^C\big)} = t^{-\alpha}$$

due to the homogeneity of ν and the fact that $t \cdot B_1(0) = B_t(0)$.
Here we have implicitely used the fact that $\nu(\overline{B_1(0)}^C)$ is non-zero, which follows by homogeneity of ν together with the fact that ν is non-zero. The convergence is justified by the assumption that X is regularly varying and that $S_t(0) := \partial B_t(0)$ must have ν-measure 0 since if $\nu(S_t(0))$ were strictly greater than 0 for some t then by homogeneity $\nu(S_t(0))$ is strictly greater than 0 for all t. In particular, since ν is $\mathcal{B}_0$-boundedly finite we have:

$$\infty > \nu(B_1(0)^C) \geq \nu\bigg(\bigcup_{n\in\mathbb{N}} S_{n^{1/\alpha}}(0)\bigg) = \sum_{n\in\mathbb{N}} \nu(S_{n^{1/\alpha}}) = \sum_{n\in\mathbb{N}} n^{-1}\nu(S_1(0))$$

This gives a contradiction if $\nu(S_1(0)) > 0$.

While two different choices of the norm $\|\cdot\|$ induce the same topology, they may not lead to the same exponent measure. In particular, for two norms $\|\cdot\|_1$ and $\|\cdot\|_2$ on $\mathbb{R}^d$ and their corresponding functions $g_1(t) = \mathbb{P}[\|X\|_1 > t]^{-1}$ and $g_2(t) = \mathbb{P}[\|X\|_2 > t]^{-1}$ we have that the limit $\lim_t g_1(t)\mathbb{P}[t^{-1}X \in A]$ exists and defines a $\mathcal{B}_0$-boundedly finite measure ν_1 if and only if the limit $\lim_t g_2(t)\mathbb{P}[t^{-1}X \in A]$ exists and defines a $\mathcal{B}_0$-boundedly finite measure ν_2. The corresponding measures are constant multiples of each other, more specifically

$$\nu_1(A) = \nu_1(\{x \in \mathbb{R}^d : \|x\|_2 > 1\}) \cdot \nu_2(A).$$

The homogeneity of ν is reminiscent of the power law distributions $x^{-\alpha}$ in the one dimensional case, as power laws are the only non-constant measurable $(-\alpha)$-homogeneous functions on $(0, \infty)$ for $\alpha > 0$.
However, unlike the one dimensional case, the extremal behaviour of a random vector X is not fully described by the index α, as this quantity alone cannot describe the dependencies between the components of X. To show this, we will now revisit the examples given in the introduction which show very different extremal behaviour, while having the same tail index:

Example 2.4. Let $X = (X_1, X_2)$ be a vector with i.i.d. components following a Pareto distribution with index α. Let $Y = (Y_1, Y_2)$ be a random vector

with independent components with $Y_1 \sim Par(\alpha)$ and $Y_2 \sim Par(\beta)$, $\beta > \alpha$. Let $Z = (Z_1, Z_1)$ be a vector with identical components following a Pareto distribution with index α. Then by choosing $a_n = n^{1/\alpha}$ in each case we get:

$$\nu_X([t_1,\infty) \times [t_2,\infty)) = \begin{cases} 0 & t_1 > 0, t_2 > 0 \\ t_1^{-\alpha} & t_1 > 0, t_2 = 0 \\ t_2^{-\alpha} & t_1 = 0, t_2 > 0 \end{cases}$$

$$\nu_Y([t_1,\infty) \times [t_2,\infty)) = \begin{cases} 0 & t_1 > 0, t_2 > 0 \\ t_1^{-\alpha} & t_1 > 0, t_2 = 0 \\ 0 & t_1 = 0, t_2 > 0 \end{cases}$$

$$\nu_Z([t_1,\infty) \times [t_2,\infty)) = \max(t_1,t_2)^{-\alpha}, t_1 > 0, t_2 > 0$$

Thus, just as indicated in Figure 1, ν_X and ν_Y are concentrated on the coordinate axis'. The difference between ν_X and ν_Y is that ν_Y vanishes on the x_2-axis, which is the axis corresponding to the less heavy-tailed component Y_2. Intuitively, this means that if we know an extreme value occurs, it is likely caused solely by the first component Y_1 being large, which is what we would expect as it is more heavy-tailed than Y_2. This example also shows that not every component needs to be regularly varying with index α. In fact we could have also chosen Y_2 to have an exponential distribution, or any other random variable Y_2 satisfying $\mathbb{E}|Y_2|^{\alpha+\epsilon} < \infty$ for some $\epsilon > 0$, and we would have gotten the same limit measure ν_Y. Therefore some components of a regularly varying random vector with tail index α might not be regularly varying with tail index α, or not even regularly varying at all. This also shows that exponent measures are not unique, meaning two random vectors with different distributions can have the same exponent measure.
Finally, ν_Z is concentrated on the diagonal, which is also where the values of Z live.

2.2 Exponent Measures

Now that we have seen how exponent measures allow us to derive limit theorems for the maxima of a regularly varying random vector X we want to find practical ways to describe the exponent measure ν. It can be very difficult to determine ν based on the expressions given in Definition 2.2 and Proposition 2.3 and hence we want to find different representations of the exponent measure ν. One convenient representation of ν is given by the so-called *spectral decomposition* (Theorem 2.5) of ν into a parametric measure on $(0,\infty)$ and a measure on the unit sphere $\mathbb{S}^{d-1} := \{x \in \mathbb{R}^d : \|x\| = 1\}$. This decomposition can then be associated to a pair of random variables R and Θ, living on $(0,\infty)$ and $\mathbb{S}^{d-1}$ respectively, which again turn out to have a very simple connection to a random

vector X having exponent measure ν (Corollary 2.6).

Choose an arbitrary vector norm $\|\cdot\|$ on $\mathbb{R}^d$ and consider the unit sphere $\mathbb{S}^{d-1} = \{x \in \mathbb{R}^d : \|x\| = 1\}$. Then we can express any non-zero vector x using polar coordinates $(r, \theta) \in (0, \infty) \times \mathbb{S}^{d-1}$ as described by the transformation

$$T : \mathbb{R}^d \setminus \{0\} \to (0, \infty) \times \mathbb{S}^{d-1}, T(x) = \left(\|x\|, \frac{x}{\|x\|} \right).$$

Now we can express the corresponding pushforward measure $\nu^T(A) = \nu(T^{-1}A)$ as a product measure:

Theorem 2.5. *Spectral Representation*
A random vector X on $\mathbb{R}^d$ is multivariate regularly varying if and only if there exists a probability measure Λ, called spectral measure*, on $\mathbb{S}^{d-1}$ and a sequence a_n such that*

$$n\mathbb{P}\left[\left(a_n^{-1}\|X\|, \frac{X}{\|X\|} \right) \in \cdot \right] \to c \cdot \mu_\alpha \otimes \Lambda$$

converges vaguely for a constant $c > 0$, where μ_α is a measure with Lebesgue density $\mathbb{1}_{(0,\infty)} \alpha u^{-(\alpha+1)} du$.
Hence $\nu^T = c \cdot \mu_\alpha \otimes \Lambda$.

Proof. See [KS20] Theorem 2.2.1. □

Now the spectral measure Λ offers a way to describe the dependency structure of extremes which the tail index α alone could not. To come back to the examples given in 2.4, we can see that for the vector X with i.i.d. components we have Λ_X having mass $1/2$ at the unit vectors $(1, 0)$ and $(0, 1)$. For the vector Y the measure Λ_Y is concentrated on the unit vector $(1, 0)$. Finally for the vector Z with identical components Λ_Z is concentrated on $2^{-1/2}(1, 1)$.
In general, we call the case of Λ being concentrated on the unit vectors *extremal independence*, as the exponent measure is identical to that of a vector with i.i.d. components and thus the extremal behaviour is similar to that of a vector with i.i.d. components.

The ambiguous constant c in Theorem 2.5 comes from the fact that there can be different exponent measures depending on the scaling sequence chosen. From now on, we choose the scaling function $g(t) = \mathbb{P}[\|X\| > t]^{-1}$ so that $c = 1$. Theorem 2.5 then suggests that we can describe the measure ν on $B_1(0)^C$ by a pair of random variables R and Θ having laws $\mu_\alpha|_{[1,\infty)}$ and Λ respectively. By homogeneity, ν is fully determined on $\mathbb{R}^d \setminus \{0\}$ once ν is determined on $B_1(0)^C$.

We can make the connection between regular variation of X and the pair (R, Θ)

more concrete by the following statement which follows immediately from Theorem 2.5:

Corollary 2.6. *A random vector X is regularly varying with exponent measure ν if and only if we have*

$$\mathcal{L}\left(\frac{\|X\|}{t}\middle|\|X\|>t\right) \xrightarrow[t\to\infty]{} \mathcal{L}(R)$$
$$\mathcal{L}\left(\frac{X}{\|X\|}\middle|\|X\|>t\right) \xrightarrow[t\to\infty]{} \mathcal{L}(\Theta)$$

For independent random variables $R \sim Par(\alpha)$ and $\Theta \in \mathbb{S}^{d-1}$.
Here $\mathcal{L}(Z)$ denotes the law of a random variable Z and $\mathcal{L}(Z|A)$ denotes the law of Z conditioned on an event A.

Theorem 2.5 also classifies all possible exponent measures ν which can arise from Definition 2.2: Pick any probability measure Λ on $\mathbb{S}^{d-1}$, and any $\alpha > 0$. Then the measure defined by $\nu^T = \mu_\alpha \otimes \Lambda$ is an exponent measure of a random variable $Y = R\Theta$ with $R \sim \mathrm{Par}(\alpha)$ and $\Theta \sim \Lambda$ independent, see [KS20] Lemma 2.2.2.

Another consequence is the following representation of ν, see [KS20] Proposition 2.2.3:

$$\nu(A) = c \cdot \int_0^\infty \mathbb{P}[u\Theta \in A]\alpha u^{-(\alpha+1)} du.$$

In Section 4 we will derive a similar decomposition for analogues of exponent measures for stochastic processes, see Theorem 4.7.

In summary, knowing the exponent measure of X helps us determine limit laws for the maxima of X. This exponent measure can, by means of Theorem 2.5 and Corollary 2.6, be computed using the tail index α and the spectral measure Λ, or equivalently by using the variables R and Θ.

3 Càdlàg Processes

Now that we have extended the concept of regular variation to multivariate random variables, we want to do the same for stochastic processes in continuous time. Many properties of a stochastic process, for example its running maximum or its path integral, are not fully determined by the marginal distributions of the process. Hence it does not suffice to merely consider the exponent measures of the finite-dimensional projections of a stochastic process.
Therefore we want to develope a concept of regular variation of a stochastic process as a whole, not just of its finite-dimensional projections. For this we first need to determine what a measurable space of processes looks like. For practical reasons, we need to make certain continuity assumptions on the underlying process. One such assumption would be to require the underlying process to be continuous. However, this excludes many important classes of stochastic processes, like Poisson processes or Lévy processes. Such processes can have jumps and are therefore not continuous. Instead, they are so-called *càdlàg* processes:
For $I = [a, b] \subseteq \mathbb{R}$, we call a function $x : I \to \mathbb{R}^d$ càdlàg (short for "continu à droite, limites à gauche" which roughly translates to "right continuous with left limits") if

1. for all $a \leq t < b$ we have $x(t^+) = \lim_{s \searrow t} x(s) = x(t)$ (right-continuous)
2. for all $a < t \leq b$ we have that $x(t^-) = \lim_{s \nearrow t} x(s)$ exists and is finite (left limits)

We write $D(I, \mathbb{R}^d)$ for the space of all càdlàg functions from I to $\mathbb{R}^d$ and x_t for $x(t)$. If the dimension d is clear from the context, we also write $D_I = D(I, \mathbb{R}^d)$.

Now to define a measure on the space of càdlàg functions we need a σ-algebra, ideally a σ-algebra compatible with a topology on $D(\mathbb{R}, \mathbb{R}^d)$, that is a Borel-$\sigma$-algebra. In Section 3.1 we will introduce a suitable topology for càdlàg processes on compact time intervals, which will have a very natural extension to non-compact time intervals. Using this topology we then study the weak and vague convergence of measures on D in Section 3.2.

3.1 Skorokhod's Topology

Let $I = [a, b] \subset \mathbb{R}$ be a compact interval in $\mathbb{R}$. Since for any càdlàg function x on I we can define the function $y(t) = x\big(\frac{t-a}{b-a}\big)$ on $[0, 1]$, which is in $D_{[0,1]}$ if and only if x is in D_I, we can without loss of generality only consider $D_{[0,1]}$.

Since for any $x \in D_{[0,1]}$ we have

$$\|x\|_\infty := \sup_{t\in[0,1]} |x(t)| < \infty$$

(see [Bil99] Eq. 12.5) one might think of defining convergence in $D_{[0,1]}$ as convergence with respect to the norm $\|\cdot\|_\infty$, as one does for the space of continuous functions C. The main problem with this approach is that while D is *complete* under this metric, it is not *separable*.
We call a metric space (M, d) separable if it contains a countable and dense subset Q, meaning for every $\epsilon > 0$ and any $x \in M$ we can find a $q \in Q$ such that $d(x, q) < \epsilon$.
We say that (M, d) is complete if every Cauchy sequence in M converges in M. The reason why separability and completeness are important is that it makes verifying weak convergence in such a space much easier. In short, for complete and separable metric spaces, *relative compactness* ([Bil99], page 57) and *tightness* ([Bil99], page 59) of a sequence of probability measures are equivalent, see Prokhorov's theorem [Bil99], Theorems 5.1 and 5.2. As a consequence, we can show weak convergence of probability measures by proving tightness, which results in a convenient characterization of weak convergence for the space D, see Theorem 3.11.
A metric for which $D_{[0,1]}$ is separable was introduced by Skorokhod:

Definition 3.1. ([Bil99], Equation 12.13)
Let $D_{[0,1]}$ be the space of càdlàg functions on $[0, 1]$. Let Λ be the set of all increasing, bijective and continuous functions with continuous inverse from $[0, 1]$ to itself. For $x, y \in D_{[0,1]}$ we define the function:

$$d_{J_1}(x, y) = \inf_{\lambda\in\Lambda} \left[\|\lambda - \mathrm{id}\,\|_\infty \vee \|x \circ \lambda - y\|_\infty \right]$$

where id denotes the identity map and $x \vee y$ denotes the maximum of x and y.

The function $d_{J_1}(\cdot, \cdot) : D_{[0,1]} \times D_{[0,1]} \to [0, \infty)$ defines a metric on $D_{[0,1]}$ (see [Bil99] page 124). We will refer to d_{J_1} as the J_1-metric on $D_{[0,1]}$.

Since $\lambda = \mathrm{id} \in \Lambda$ we can see that $d_{J_1}(x, y) \leq \|x - y\|_\infty$ and therefore uniform convergence implies convergence with respect to the metric d_{J_1}.

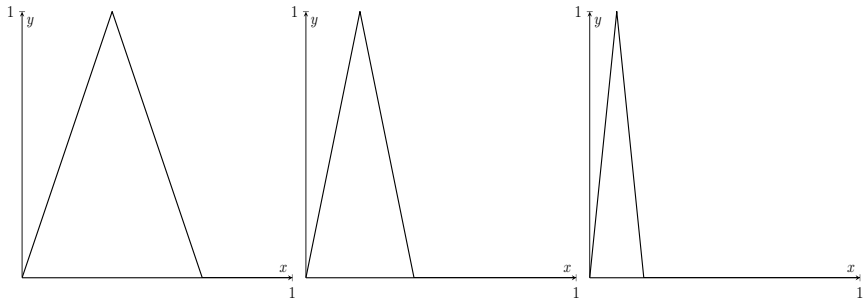
Example 3.2. Projection Maps
If we are dealing with càdlàg processes x we might be interested in the value of x at a certain time point t, meaning we want to determine x_t. This can be interpreted as a map

$$\pi_t : D \to \mathbb{R}^d, x \mapsto x_t.$$

The natural question to ask is: Is this map continuous?
If we were to use the uniform metric $\|\cdot\|_\infty$ this would clearly be the case, since uniform convergence implies pointwise convergence. However, this does not hold in (D, d_{J_1}), meaning functions x_n can converge to x with respect to d_{J_1} but $x_n(t) \not\to x(t)$ for certain t. As an example, consider the sequence $x_n(t) = \mathbb{1}_{[0,1/2+1/n)}(t)$ and $x(t) = \mathbb{1}_{[0,1/2)}(t)$. Clearly $x_n(1/2) \not\to x(1/2)$, but x_n converges to x with respect to d_{J_1} (see also [Bil99] page 124): Choose λ_n to be equal to $1/2 + 1/n$ at $t = 1/2$ and let λ_n be linear on $[0, 1/2)$ as well as on $[1/2, 1]$. Now we have $\|\lambda_n - \mathrm{id}\|_\infty = 1/n$ as well as $\|x_n \circ \lambda_n - x\|_\infty = 0$. Hence $d_{J_1}(x_n, x) \leq 1/n \to 0$.
Therefore projection maps are not continuous. However, note that in this example $x_n(t) \to x(t)$ for every continuity point t of x. This is always the case, meaning π_t is continuous at x if and only if t is a continuity point of x, see [Bil99] Theorem 12.5.
On the other hand, pointwise convergence $\pi_t(x_n) \to \pi_t(x)$ for all t does not imply convergence with respect to d_{J_1}: Consider the sequence z_n given in [Bil99] Example 1.3 defined by

$$z_n(t) = \begin{cases} nt & t \in [0, 1/n) \\ 2 - nt & t \in [1/n, 2/n) \\ 0 & t \in [2/n, 1] \end{cases}$$

then $z_n(t) \to 0$ for all t, but $d_{J_1}(z_n, 0) = 1$ for all n. Hence $z_n \not\to 0$ in $D_{[0,1]}$.



(a) $n = 3$ (b) $n = 5$ (c) $n = 10$

Figure 2: z_n

Fortunately, projection maps are still measurable maps for the Borel-σ-algebra induced by the metric d_{J_1}, see [Bil99] Theorem 12.5.

Definition 3.3. Skorokhod's J_1 topology

The topology generated by open sets of the metric d_{J_1}, as defined in Definition 3.1, is called Skorokhod's J_1 topology on $D_{[0,1]}$.

Since uniform convergence implies convergence with respect to d_{J_1} (but not the other way around), the topology generated by the norm $\|\cdot\|_\infty$ is coarser than the J_1 topology and thus more sequences converge in J_1 than in the uniform topology. Unlike $(D_{[0,1]}, \|\cdot\|_\infty)$, the space $(D_{[0,1]}, d_{J_1})$ is now separable (see [Bil99] Theorem 12.2).
Unfortunately, it is not a complete metric space, see [Bil99] Example 12.2. However with a slight modification of d_{J_1} we can find a metric $d^\circ_{J_1}$ equivalent to d_{J_1} (and thus inducing the same topology J_1), but for which $(D, d^\circ_{J_1})$ is separable and complete!

Theorem 3.4. *For $\lambda \in \Lambda$ define the function*

$$\|\lambda\|^\circ = \sup_{s<t} \left| \log \frac{\lambda(t) - \lambda(s)}{t - s} \right|$$

and define the metric

$$d^\circ_{J_1}(x, y) = \inf_{\lambda \in \Lambda} \left[\|\lambda\|^\circ \vee \|x \circ \lambda - y\|_\infty \right]$$

Then the space $(D_{[0,1]}, d^\circ_{J_1})$ is separable and complete.
Furthermore, $d^\circ_{J_1}$ and d_{J_1} are equivalent on $D_{[0,1]}$.

Proof. See [Bil99] Theorem 12.1 and Theorem 12.2. □

One important concept related to càdlàg functions is the *modulus of continuity*. It is a function which characterizes whether a function x is càdlàg. To motivate this, note that a function x is continuous on $[0, 1]$ if and only if

$$w(x, 0, 1, \delta) := \sup_{s,t \in [0,1]; |s-t| \le \delta} |x_s - x_t| \to 0, \qquad \delta \to 0.$$

We would like to have to have an analogue of such a function for D. This will play an important role in Chapter 3.2 for weak convergence of measures in D as well as in Chapter 4.1 for verifying regular variation in D:

Lemma 3.5. *Let $\mathcal{P}([0, 1], \delta)$ be the set of all increasing time points $t_0, t_1, \ldots, t_k$ in $[0, 1]$ such that $t_0 = 0, t_k = 1$ and for all $i = 1, \ldots, k$ we have $|t_i - t_{i-1}| \ge \delta$. Define the function*

$$w'(x, 0, 1, \delta) = \inf_{(t_0,\ldots,t_k) \in \mathcal{P}([0,1],\delta)} \max_{1 \le i \le k} \sup_{t_{i-1} \le s,t < t_i} |x_s - x_t|$$

We call w' the modulus of continuity and we have

$$x \in D([0, 1], \mathbb{R}^d) \Leftrightarrow \lim_{\delta \to 0} w'(x, 0, 1, \delta) = 0$$

Proof. See [Bil99] Section 12 Lemma 1. □

Skorokhod's Topology on $D(\mathbb{R}, \mathbb{R}^d)$

We now naturally extend the topology of $D([0,1], \mathbb{R}^d)$ to the space of càdlàg functions over the real line.

As already noted, we can define the metric d_{J_1} in Definition 3.1 for every compact interval I. Denote by $d_{[a,b]}$ the metric as in Definition 3.1 for càdlàg functions defined on $[a,b]$. We can then extend this metric to all of $D_{\mathbb{R}}$:

Definition 3.6. ([Sou22], Appendix B)
The J_1 metric on $D(\mathbb{R}, \mathbb{R}^d)$ is defined by

$$d_{J_1}(f,g) = \int_0^\infty [d_{[-t,t]}(f,g) \wedge 1] e^{-t} dt$$

where $d_{[-t,t]}(f,g)$ is interpreted as the distance of f and g restricted to the interval $[-t,t]$.

The topology generated by this metric is called the J_1 topology on $D(\mathbb{R}, \mathbb{R}^d)$.

The fact that this function defines a metric follows immediately from the fact that d_I defines a metric for compact I and that $d_{J_1}(f,g) \leq \int_0^\infty e^{-t} dt < \infty$. An analogous metric on $D([0,\infty), \mathbb{R}^d)$ is defined by integrating $d_{[0,t]}$ instead of $d_{[-t,t]}$.

We can extend $d^\circ_{J_1}$ to $D(\mathbb{R}, \mathbb{R}^d)$ in just the same way: For $f, g \in D(\mathbb{R}, \mathbb{R}^d)$ define

$$d^\circ_{J_1}(f,g) = \int_0^\infty \left[d^\circ_{[-t,t]}(f,g) \vee 1\right] e^{-t} dt$$

Again, $d^\circ_{J_1}$ is a metric because every $d^\circ_{[-t,t]}$ is a metric. Under this metric, $D(\mathbb{R}, \mathbb{R}^d)$ is again a complete and separable metric space (see [Bil99] Theorem 16.3).

Notice that in order for $d^\circ_{J_1}(f_n, f)$ to go to 0 almost all $d^\circ_{[-t,t]}(f_n, f)$ must go to 0. Therefore convergence in $D(\mathbb{R}, \mathbb{R}^d)$ is equivalent to convergence in $D([-t,t], \mathbb{R}^d)$ for almost all t. This brings us to the following Theorem connecting convergence in $D(I, \mathbb{R}^d)$ to convergence in $D(\mathbb{R}, \mathbb{R}^d)$:

Theorem 3.7. *A sequence $(x_n)_{n\in\mathbb{N}}$ in $D(\mathbb{R}, \mathbb{R}^d)$ converges to x in $D(\mathbb{R}, \mathbb{R}^d)$ if and only if $d^\circ_{[-t,t]}(x_n, x) \to 0$ for all continuity points t of x.*

Proof. See [Bil99] Theorem 16.2. □

3.2 Weak Convergence and Vague Convergence in D

Theorem 3.7 essentially says that convergence in $D(\mathbb{R}, \mathbb{R}^d)$ is equivalent to convergence in $D([-t,t], \mathbb{R}^d)$ for almost all t. This concept extends to weak convergence, in the sense that a sequence of probability measure $\mathbb{P}^{(n)}$ on $D(\mathbb{R}, \mathbb{R}^d)$

converges weakly to $\mathbb{P}$ if and only if $\mathbb{P}^{(n)} \circ \pi_{[-t,t]}^{-1}$ converges weakly to $\mathbb{P} \circ \pi_{[-t,t]}^{-1}$ in $D([-t,t], \mathbb{R}^d)$ for every t, where

$$\pi_{[-t,t]} : D(\mathbb{R}, \mathbb{R}^d) \to D([-t,t], \mathbb{R}^d), \left(\pi_{[-t,t]}(x)\right)_{s \in [-t,t]} = (x_s)_{s \in [-t,t]},$$

see [Bil99] Section 16 Lemma 3. Therefore, in this section we will only consider weak convergence on $D := D(I, \mathbb{R}^d)$ for compact $I = [a, b]$.

Now that we have defined a topology on D, we can generate a σ-algebra using the sets in J_1. In this section every measure on D will be defined on the σ-algebra $\sigma(J_1)$, meaning the smallest σ-algebra containing all sets in J_1.
A random variable $\mathbf{X}$ is called a càdlàg stochastic process or a D-valued random variable if it is a measurable map of the form

$$\mathbf{X} : (\Omega, \mathcal{F}) \to (D, \sigma(J_1))$$

where $(\Omega, \mathcal{F}, \mathbb{P})$ is some probability space. The law of $\mathbf{X}$ is then defined as usual by $\mathbb{P}^{\mathbf{X}} := \mathbb{P} \circ \mathbf{X}^{-1}$, which defines a probability measure on D.

For real-valued random variables convergence in distribution, or weak convergence, is often defined by convergence of distribution functions F_n to a distribution function F at continuity points of F. For a D-valued random variable we have no obvious analogue of a distribution function on D. Instead, notice that if a is a continuity point of F then for the measure defined by $\mathbb{P}[(-\infty, a]] := F(a)$ we have $\mathbb{P}[\{a\}] = \mathbb{P}[\partial(-\infty, a]] = 0$. Thus weak convergence of random variables can be reformulated as $\mathbb{P}^{(n)}[A] \to \mathbb{P}[A]$ for every A with $\mathbb{P}[\partial A] = 0$, where $\mathbb{P}^{(n)}[(-\infty, a]] = F_n(a)$. This formulation can be extended to general metric spaces as follows:

Definition 3.8. Weak Convergence Let (M, d) be a metric space and make M into a measurable space by taking $\mathcal{F}$ to be the σ-algebra generated by open sets of d. Then finite measures μ_n on $(M, \mathcal{F})$ are said to converge weakly to a finite measure μ on $(M, \mathcal{F})$ if for all measurable sets A with $\mu(\partial A) = 0$ we have $\mu_n(A) \to \mu(A)$ as $n \to \infty$.
We then write $\mu_n \Rightarrow \mu$.
Here ∂A denotes the topological boundary of A, meaning the set of points x such that every open neighbourhood around x intersects with both A and A^C.

Similarly we say a sequence of random variables $\mathbf{X}^{(n)}$ on M converge weakly to a random variable $\mathbf{X}$ on M if their laws converge weakly, i.e. $\mathbb{P}^{\mathbf{X}^{(n)}} \Rightarrow \mathbb{P}^{\mathbf{X}}$. We often write $\mathcal{L}(\mathbf{X})$ for the law of a random variable $\mathbf{X}$ and write $\mathcal{L}(\mathbf{X}^{(n)}) \to \mathcal{L}(\mathbf{X})$ if $\mathbf{X}^{(n)}$ converges weakly to $\mathbf{X}$.

Using this definition alone, it can be very hard to verify weak convergence on D, so instead we search for sufficient conditions for weak convergence which

are easier to verify than the condition given in Definition 3.9.
A first idea for such a condition would be to require weak convergence of the finite-dimensional projections $X_t = \pi_t(\mathbf{X})$. Since the projection maps are measurable, X_t is a well-defined random variable on $\mathbb{R}^d$. This is indeed *almost* a necessary condition, in the sense that this is true on a certain subset of time points $T \subset \mathbb{R}$.
To show this, we need a version of the continuous-mapping theorem on D:

Theorem 3.9. *([Bil99], Theorem 2.7)*
Let (M, d) and (M', d') be two metric spaces endowed with their Borel-σ-algebras. Let h be a measurable map from M to M'. Define the set D_h as the set of discontinuity points of h. If for probability measures $\mathbb{P}^{(n)}$ on M we have $\mathbb{P}^{(n)} \Rightarrow \mathbb{P}$ on M and $\mathbb{P}[D_h] = 0$ then $\mathbb{P}^{(n)} \circ h^{-1} \Rightarrow \mathbb{P} \circ h^{-1}$ on M'.

Example 3.10. Finite-dimensional weak convergence
Let $\mathbb{P}$ be a probability measure on D and let $T_{\mathbb{P}}$ be the set of time points t in $\mathbb{R}$ for which π_t is $\mathbb{P}$-almost-surely continuous. For time points $t_1, ..., t_k \in T_{\mathbb{P}}$ we then can apply the continuous mapping theorem to the map $\pi = (\pi_{t_1}, ..., \pi_{t_k})$ and get that if $\mathbf{X}^{(n)}$ are D-valued random variables converging weakly to $\mathbf{X}$ with law $\mathbb{P}$, then $\pi(\mathbf{X}^{(n)}) = (X^{(n)}_{t_1}, ..., X^{(n)}_{t_k})$ converges weakly to $\pi(\mathbf{X}) = (X_{t_1}, ..., X_{t_k})$. Hence weak convergence in D implies finite-dimensional weak convergence in $T_{\mathbb{P}}$.
However, it does not imply finite-dimensional weak convergence outside of $T_{\mathbb{P}}$: Consider the functions x_n defined in Example 3.2 and define point measures $\mathbb{P}^{(n)} = \delta_{x_n}$ as well as $\mathbb{P} = \delta_x$, then $\mathbb{P}^{(n)} \Rightarrow \mathbb{P}$ in D but $\mathbb{P}^{(n)} \circ \pi^{-1}_{1/2} \not\Rightarrow \mathbb{P} \circ \pi^{-1}_{1/2}$. Hence although we have weak convergence in D, we might not have finite-dimensional weak convergence on *all* finite-dimensional projections.
Similarly, define the point measures $\mathbb{P}^{(n)} = \delta_{z_n}$ as well as $\mathbb{P} = \delta_0$, where z_n are the functions in Example 3.2, then clearly $\mathbb{P}^{(n)} \not\Rightarrow \mathbb{P}$ in D but $\mathbb{P}^{(n)} \circ \pi_t^{-1} \Rightarrow \mathbb{P} \circ \pi_t^{-1}$ for all $t \in [0, 1]$. Hence weak convergence of finite-dimensional projections is not sufficient for probability measures to converge weakly in D.

Since finite-dimensional weak convergence is not sufficient for weak convergence in D, we need to look for additional conditions on the sequence $(\mathbb{P}^{(n)})_{n\in\mathbb{N}}$ such that weak convergence in D holds. The following Theorem gives an example of such a condition. It also serves as motivation for Theorem 4.11, which will give a similar suffiecient condition for regular variation in D.

Theorem 3.11. *([Bil99], Theorem 13.2)*
Let $(\mathbb{P}^{(n)})_{n\in\mathbb{N}}$ and $\mathbb{P}$ be probability measures on D and let $T_{\mathbb{P}}$ be the set of all t such that π_t is $\mathbb{P}$-almost-surely continuous. Suppose that for all $t_1, \ldots, t_k \in T_{\mathbb{P}}$ we have $\mathbb{P}^{(n)} \circ \pi^{-1}_{t_1,...,t_k} \Rightarrow \mathbb{P} \circ \pi^{-1}_{t_1,...,t_k}$. Then $\mathbb{P}^{(n)} \Rightarrow \mathbb{P}$ in D if and only if the following two conditions hold:

1.

$$\lim_{a\to\infty} \limsup_{n\to\infty} \mathbb{P}^{(n)}[\{x \in D : \|x\| \geq a\}] = 0$$

2. For all $\epsilon > 0$ we have

$$\lim_{\delta\to 0} \limsup_{n\to\infty} \mathbb{P}^{(n)}[\{x \in D : w'(x, 0, 1, \delta) \geq \epsilon\}] = 0$$

Vague Convergence

Just as in Section 2, denote by $\mathcal{B}_0$ the set of all measurable sets bounded away from 0 in D, meaning $\inf_{y\in A} d_{J_1}(y, 0) > \epsilon$ for some $\epsilon > 0$. Just as exponent measures were not finite and thus required a modification of weak convergence, the measures we will consider in Section 4 are also infinite, but finite on $\mathcal{B}_0$. Therefore we now introduce the concept of vague convergence in D.

Definition 3.12. Vague Convergence ([Sou22], Appendix A)
A measure ν on D is called $\mathcal{B}_0$-boundedly finite if for every $A \in \mathcal{B}_0$ we have $\nu(A) < \infty$.
A sequence $(\nu_n)_{n\in\mathbb{N}}$ is said to converge vaguely to a $\mathcal{B}_0$-boundedly finite measure ν if for every $A \in \mathcal{B}_0$ with $\nu(\partial A) = 0$ we have

$$\lim_{n\to\infty} \nu_n(A) = \nu(A).$$

The following Lemma, taken from [Sou22] Appendix B, characterizes sets in $\mathcal{B}_0$ in a more practical way:

Lemma 3.13. *A measurable set A is bounded away from 0 in $D(\mathbb{R}, \mathbb{R}^d)$ if and only if there exists an $\epsilon > 0$ and an $a > 0$ such that for all $y \in A$ we have $\|y\|_{[-a,a]} > \epsilon$.*

4 Tail Measures

Now that we have introduced stochastic processes on $D := D(\mathbb{R}, \mathbb{R}^d)$ we will try to extend the results from multivariate extreme value theory discussed in Section 2 to càdlàg processes. First we will discuss the two main approaches of defining the concept of regular variation for a stochastic process $\mathbf{X}$:
Since a multivariate stochastic process $X : \Omega \to D(\mathbb{R}, \mathbb{R}^d)$ can be seen as a collection of random vectors $\pi_t(\mathbf{X}) = X_t : \Omega \to \mathbb{R}^d$ indexed by $t \in \mathbb{R}$, one might think to define a stochastic process $\mathbf{X}$ to be regularly varying if and only if all of its finite dimensional projections $(X_{t_1}, ..., X_{t_k})$ for $t_1, ..., t_k \in \mathbb{R}$ are multivariate regularly varying. If this is the case and all finite dimensional projections have the same index of regular variation α as well as the same scaling sequence $\{a_n\}$, we call the process $\mathbf{X}$ *finite dimensional* regularly varying, see [Sou22], Equation 3.1. On the other hand, one might think to define regular variation of $\mathbf{X}$ by formulating the limit as in Definition 2.2, meaning $\mathbf{X}$ is regularly varying if and only if there exists a $\mathcal{B}_0$-boundedly finite non-zero measure ν such that

$$n\mathbb{P}[a_n^{-1}\mathbf{X} \in A] \to \nu(A)$$

for all $A \in \mathcal{B}_0$ with $\nu(\partial A) = 0$.
If $\mathbf{X}$ is a stationary time series, meaning a random element of $(\mathbb{R}^d)^{\mathbb{Z}}$, then ν is a measure on $(\mathbb{R}^d)^{\mathbb{Z}}$ and these two definitions turn out to be equivalent, meaning the exponent measures of the finite-dimensional projections can be extended to a tail measure on $(\mathbb{R}^d)^{\mathbb{Z}}$, as shown in [KS20] Appendix B.2.3.
In [SO12] Theorem 2.1 it was shown that the exponent measures of finite-dimensional projections of a finite dimensional regularly varying continuous-time process can also be extended to a measure on $(\mathbb{R}^d)^{\mathbb{R}}$. However, since our original process $\mathbf{X}$ is assumed to be càdlàg, we would want our measure to be defined on D. Such an extension does not always exist, meaning for continuous-time processes, finite-dimensional regular variation is a weaker notion than regular variation in D. This will further be explained by Theorem 4.11.

Definition 4.1. Regular Variation in D ([Sou22], Definition 3.1)
We call a stationary càdlàg stochastic process $\mathbf{X}$ regularly varying in D if there exists a non-zero $\mathcal{B}_0$-boundedly finite measure ν on D such that

$$\frac{\mathbb{P}[\mathbf{X} \in u \cdot]}{\mathbb{P}[|X_0| > u]} \xrightarrow[u\to\infty]{} \nu(\cdot)$$

vaguely.
We call ν the *tail measure* of $\mathbf{X}$.

Note that if we choose a sequence a_n such that $\mathbb{P}[|X_0| > a_n] \sim n^{-1}$ we see that Definition 4.1 implies $n\mathbb{P}[a_n^{-1}\mathbf{X} \in u\cdot] \to \nu(\cdot)$, which is reminiscent of

Definition 2.2. This formulation is often used for defining regular variation of non-stationary processes, see for example [Lin04], Equation 3.1.

Just like exponent measures are useful to derive limit theorems for maxima of regularly varying random vectors, tail measures can be used to do the same for càdlàg processes:
Define the maximum of two processes $x, y \in D$ by $(x \vee y)_t = (x_t \vee y_t)$, where in the case that x_t and y_t are multivariate the maximum is taken componentwise as was done in Section 2. For $u \in D$ write $x \preceq u$ if for every $t \in \mathbb{R}$ we have $x_t \preceq u_t$, where again $x_t \preceq u_t$ is the same partial order we used in Section 2. For $l, u \in D$ define $[l, u] := \{y \in D : l \preceq y \preceq u\}$ and $(-\infty, u]$ analogously.
Then for a regularly varying process $\mathbf{X}$ with non-negative components we have for the sequence a_n of Definition 4.1, i.i.d. copies $(\mathbf{X}^{(i)})_{i\in\mathbb{N}}$ of $\mathbf{X}$ and $0 \preceq u$, $u \neq 0$:

$$\lim_{n\to\infty} \mathbb{P}\left[a_n^{-1} \bigvee_{i=1}^{n} \mathbf{X}^{(i)} \preceq u\right] = \exp\big(-\nu([0,u]^C)\big)$$

The proof of this result is exactly the same as the proof given in Section 2.

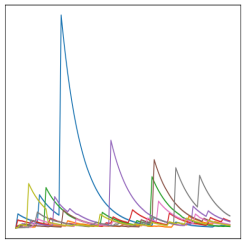


Figure 3: 5 samples of the maxima of 100 Lévy-driven Ornstein-Uhlenbeck processes (Example 4.7)

Another important application of regular variation of processes is that it allows us to study the extremal behaviour of functionals H of our process $\mathbf{X}$. In particular, we would hope to get a "continuous mapping" result like

$$n \cdot \mathbb{P}[H(a_n^{-1}\mathbf{X}) \in \cdot] \to \nu \circ H^{-1}(\cdot) \qquad \text{vaguely},$$

see for example [HL05] Theorem 6 for such a result in $D([0,1], \mathbb{R}^d)$.
One functional that is of particular interest in extreme value theory is the running maximum $\sup_{0\leq s\leq T} |X_s|$. In Section 4.3 we will see that under certain conditions we can derive a single index ϑ, called (candidate) extremal index,

using the tail measure such that

$$\mathbb{P}\left[a_T^{-1} \sup_{0\le s\le T} |X_s| \le t\right] \to \exp(-\vartheta t^{-\alpha}).$$

The aim of this section is the same as the one in Section 2: We want to derive basic properties of the limit measure ν, find more convenient ways of computing ν and find sufficient conditions for regular variation of $\mathbf{X}$.

The first significant difference between the theory of regular variation of stochastic processes in D and the theory of regular variation of time series in $(\mathbb{R}^d)^{\mathbb{Z}}$ is that finite-dimensional regular variation and regular variation in D are not equivalent. Regular variation of cádlág processes is a stronger condition than finite-dimensional regular variation, since the former implies the latter:

Proposition 4.2. *Suppose $\boldsymbol{X}$ is stationary and regularly varying in D with tail measure ν.*
Then $\boldsymbol{X}$ is finite-dimensional regularly varying. In particular, the finite-dimensional projections of ν are exponent measure of the finite-dimensional projections of $\boldsymbol{X}$; meaning $\nu \circ \pi_{t_1,\dots,t_k} = \nu_{t_1,\dots,t_k}$ where $\nu_{t_1,\dots,t_k}$ is an exponent measure of $(X_{t_1},\dots,X_{t_k})$.

Proof. Pick any $A \subseteq (\mathbb{R}^d)^k$ which is bounded away from 0 and any sequence $t_1 < \cdots < t_k \in \mathbb{R}$, then since the projection $\pi_{t_1,\dots,t_k}$ is measurable the set $\pi^{-1}_{t_1,\dots,t_k}(A)$ is measurable in D. Since for all $x \in A$ we have $\|x\|_\infty > \epsilon$ for some fixed $\epsilon > 0$ and therefore $\|y\|_{[t_1,t_k]} > \epsilon$ for all $y \in \pi^{-1}_{t_1,\dots,t_k} A$, it follows by Lemma 3.13 that $\pi^{-1}_{t_1,\dots,t_k} A$ is bounded away from 0 in D.
Now by definition of regular variation we have

$$n\mathbb{P}[a_n^{-1}(X_{t_1},...,X_{t_k}) \in A] = n\mathbb{P}[\mathbf{X} \in a_n \pi^{-1}_{t_1,\dots,t_k}(A)] \to \nu(\pi^{-1}_{t_1,\dots,t_k}(A))$$

for every ν-continuity set $\pi^{-1}_{t_1,\dots,t_k}(A)$ or equivalently for every $\nu \circ \pi^{-1}_{t_1,\dots,t_k}$-continuity set A.
If $\nu(\pi^{-1}_{t_1,\dots,t_k}(A))$ were 0 for all A then by shift-invariance of ν we get

$$\nu(\pi^{-1}_{t_1-T,\dots,t_k-T}(A)) = 0, \qquad \text{for all } T \in \mathbb{R}.$$

This would imply that ν were zero everywhere, which contradicts the definition of regular variation. Thus $\nu \circ \pi_{t_1,\dots,t_k}$ is non-zero and hence the projection $(X_{t_1},...,X_{t_k})$ is regularly varying. It follows that $(X_{t_1},\dots,X_{t_k})$ is regularly varying with exponent measure $\nu \circ \pi^{-1}_{t_1,\dots,t_k}$.
Note that the scaling sequence a_n did not depend on A nor on $t_1,\dots,t_k$ and hence is the same for all finite-dimensional projections. Therefore $\mathbf{X}$ is finite-dimensional regularly varying. □

Using Definition 4.1 alone, one can immediately deduce several properties of the limit measure ν if $\mathbf{X}$ is stationary:

1. If $\mathbb{P}[\mathbf{X} = 0] = 0$ (which we may assume since the process $\mathbf{X} = 0$ is not regularly varying) then it is natural to assume $\nu(\{0\}) = 0$, even though this does not strictly follow from Definition 4.1 as $\{0\}$ is not bounded away from 0 in D.

2. Just like exponent measures, tail measures are homogeneous. For any measurable set A in D and any scalar $t > 0$ we have the homogeneity property $\nu(tA) = t^{-\alpha}\nu(A)$: By Proposition 4.2 X_0 must be regularly varying and hence

$$\nu(tA) = \lim_{u\to\infty} \frac{\mathbb{P}[X \in utA]}{\mathbb{P}[|X_0| > ut]} \cdot \frac{\mathbb{P}[|X_0| > ut]}{\mathbb{P}[|X_0| > u]} = \nu(A)t^{-\alpha}$$

3. Defining the set $A = \{y \in D : |y_0| > 1\}$ we get by homogeneity of ν that $\nu(\partial A)$ must be equal to 0 and thus applying the limit definition we get

$$\nu(A) = \lim_{u\to\infty} \frac{\mathbb{P}[X \in uA]}{\mathbb{P}[|X_0| > u]} = \lim_{u\to\infty} \frac{\mathbb{P}[|X_0| > u]}{\mathbb{P}[|X_0| > u]} = 1$$

   and thus we have the normalization $\nu(\{y \in D : |y_0| > 1\}) = 1$.

4. By stationarity of $\mathbf{X}$, we get for the shift-operator B^t defined by $(B^t x)(s) = x(s-t)$ that

$$\nu(B^t A) \lim_{u\to\infty} \frac{\mathbb{P}[X \in uB^t A]}{\mathbb{P}[|X_0| > u]} = \frac{\mathbb{P}[X \in uA]}{\mathbb{P}[|X_0| > u]} = \nu(A).$$

   This property is called *shift-invariance*.

This leads to the following abstract definition of a *tail measure*:

Definition 4.3. Tail Measure ([Sou22], Definition 2.1)
A measure ν on $(D, \sigma(J_1))$ is called a tail measure if it is a $\mathcal{B}_0$-boundedly finite Borel measure with

1. $\nu(\{0\}) = 0$;

2. There exists some $\alpha > 0$ such that for all Borel sets A in D we have $\nu(tA) = t^{-\alpha}\nu(A)$;

3. $\nu(\{y \in D : |y_0| > 1\}) = 1$.

If furthermore ν has the property $\nu(B^t A) = \nu(A)$ for all Borel sets A, we call ν a shift-invariant tail measure.

Due to the homogeneity of ν and the fact that $\nu(\{0\}) = 0$ we get the following useful property of tail measures, which was proven for tail measures of time series in [DHS18] Lemma 2.3:

Lemma 4.4. *A tail measure ν on D is σ-finite and uniquely determined by its restrictions $\nu(\cdot \cap \{y \in D : |y_t| > 1\})$ for $t \in \mathbb{R}$.*

Proof. The proof is very similar to the one given in [DHS18] Lemma 2.3:
We can write $D \setminus \{0\}$ as

$$D \setminus \{0\} = \bigcup_{q \in \mathbb{Q}, n \geq 1} A_{q,n}$$

with $A_{q,n} = \{y \in D : |y_t| > n^{-1}\}$, since if $y \neq 0$ in D there must exist a $t \in \mathbb{R}$ and $n \geq 1$ such that $|y_t| > n^{-1}$ and since $\mathbb{Q}$ is dense in $\mathbb{R}$ and y is right-continuous there must be a $q \in \mathbb{Q}$ such that $|y_q| > n^{-1}$ meaning $y \in A_{q,n}$. Since each set $A_{q,n}$ is bounded away from 0 in D we have $\nu(A_{q,n}) < \infty$. Pick an arbitrary enumeration $(B_i)_{i \in \mathbb{N}}$ of $(A_{q,n})_{q \in \mathbb{Q}, n \geq 1}$ and define the sequence $C_1 = B_1$, $C_i = B_i \setminus (C_{i-1} \cup \cdots \cup C_1)$. Now we have pairwise disjoint sets C_i of finite measure with

$$D \setminus \{0\} = \bigcup_{i \in \mathbb{N}} C_i$$

proving that ν is σ-finite as $\nu(\{0\}) = 0$.
To prove that ν is determined by its restrictions $\nu(\cdot \cap \{y \in D : |y_t| > 1\})$ pick any $A \subset D \setminus \{0\}$ and note that we can write

$$\nu(A) = \nu\left(\dot{\bigcup_{i \in \mathbb{N}}} C_i \cap A \right) = \sum_{i \in \mathbb{N}} \nu(C_i \cap A)$$

which shows that ν is completely determined by its restrictions to the sets C_i. Since each C_i is a finite combination of the sets $(A_{h,n})_{q \in \mathbb{Q}, n \in \mathbb{N}}$ and for each set $A_{h,n}$ we have $\nu(A_{h,n}) = n^{-\alpha} \nu(A_{h,1}) = \nu(A_{h,1} \cap \{y \in D : |y_q| > 1\})$ it follows that ν is uniquely determined on $(A_{h,n})_{h \in \mathbb{Q}, n \in \mathbb{N}}$ and hence on $(C_i)_{i \in \mathbb{N}}$ by the restrictions $(\nu(\cdot \cap \{y \in D : |y_h| > 1\}))_{h \in \mathbb{Q}}$. □

One corollary of Lemma 4.4 is that a shift-invariant tail measure ν is determined by the single restriction $\nu(\cdot \cap \{y \in D : |y_0| > 1\})$. Due to property 3 of Definition 4.3 we can associate this restriction to a probability measure and a corresponding stochastic process, called a *tail process*, originally introduced for time series in [BS09]. In Section 4.1 we will introduce the tail process in continuous time, as was done in [Sou22].
Unfortunately, Lemma 4.4 does not offer any practical way on how to reconstruct a shift-invariant ν by just knowing $\nu(\cdot \cap \{|y_0| > 1\})$. The purpose of Sections 4.2 and 4.3 will be to show how to compute ν given its tail process.

4.1 Tail Process and Spectral Tail Process

In Theorem 2.5 we saw how to associate an exponent measure to a pair $(R, \Theta) \in (0,\infty)\times\mathbb{S}^{d-1}$ of random variables completely determining the exponent measure. In this section, we will see that there is an extension of the pair (R, Θ) to tail measures, meaning there is a stochastic process $\mathbf{Y} = R\Theta$ completely determining a tail measure. First, it is unclear what a sphere $\mathbb{S}$ should be in D as we do not have a natural choice for a norm on D. Instead we consider the very simple pseudo-norm $y \mapsto |y_0|$ and define $\mathbb{S} := \{y \in D : |y_0| = 1\}$. Since for multivariate regularly varying random variables X the pair (R, Θ) could be seen as the distributional limit of $(\|X\|/u, X/\|X\|)$ conditioned on $\|X\| > u$ (see Theorem 2.5 and Corollary 2.6), we might now replace $\|\cdot\|$ by our pseudo-norm $y \mapsto |y_0|$ and define a pair of random variables $|Y_0|$ and Θ (in distribution) as the weak limit of $(|X_0|/u, \mathbf{X}/|X_0|)$ conditioned on $|X_0| > u$ as $u \to \infty$, assuming it exists. Note that $|Y_0|$ is an $\mathbb{R}$-valued random variable and Θ is a stochastic process. By regular variation of $|X_0|$ it is clear that $\mathcal{L}(|X_0|/u \big| |X_0| > u)$ converges to a Pareto distribution with index α, therefore the random variable $|Y_0|$ has the same distribution as the random variable R in Corollary 2.6.
Now define the process $\mathbf{Y} = |Y_0| \cdot \Theta$ and by definition of $\mathcal{L}(|Y_0|)$ and $\mathcal{L}(\Theta)$ we get the relation

$$\mathcal{L}(\mathbf{X}/u \big| |X_0| > u) \to \mathcal{L}(\mathbf{Y}) \qquad u \to \infty$$

which we can express using the tail measure ν of $\mathbf{X}$ in the following way:

$$\begin{aligned}
\mathbb{P}[\mathbf{X}/u \in A \big| |X_0| > u] &= \frac{\mathbb{P}[\mathbf{X} \in uA, |X_0| > u]}{\mathbb{P}[|X_0| > u]} \\
&= \frac{\mathbb{P}[\mathbf{X} \in u \cdot (A \cap \{y \in D : |y_0| > 1\})]}{\mathbb{P}[|X_0| > u]} \\
&\to \nu(A \cap \{y \in D : |y_0| > 1\}) \qquad u \to \infty
\end{aligned}$$

assuming $A \cap \{y \in D : |y_0| > 1\}$ is a ν-continuity set.
Therefore it is natural to define the law of $\mathbf{Y}$ as $\nu(\cdot \cap \{y \in D : |y_0| > 1\})$ which is a well-defined probability measure on D for any tail measure ν by Property 3 of Definition 4.3 and uniquely determines the shift-invariant ν by Lemma 4.4. This leads us to the following definition (see [Sou22] Definition 2.1):

Definition 4.5. Tail Process and Spectral Tail Process
Let ν be a tail measure on D in the sense of Definition 4.3. Define the probability measure

$$\mathbb{P}^{\mathbf{Y}}(\cdot) = \nu(\cdot \cap \{y \in D : |y_0| > 1\})$$

We call a D-valued random variable $\mathbf{Y}$ with distribution $\mathbb{P}^{\mathbf{Y}}$ the tail process of ν and the process $\Theta = \mathbf{Y}/|Y_0|$ the spectral tail process of ν.

If $\mathbf{X}$ is regularly varying with tail measure ν, we say that $\mathbf{Y}$ is the tail process and Θ is the spectral tail process of $\mathbf{X}$.

The forward spectral tail process $(\Theta_t)_{t\geq 0}$ describes how $\mathbf{X}$ behaves after an extreme value at $t = 0$ occured, while the backward spectral tail process $(\Theta_{-t})_{t\geq 0}$ describes what lead to the extreme value of $\mathbf{X}$ at $t = 0$. The value of $|Y_0|$ gives the magnitude of the extreme value at $t = 0$. Specifically, if we condition on $|X_0| > u$ then $|X_0|$ will be approximately $u \cdot |Y_0|$ for large u.

Note that we could have also chosen $\mathbb{P}^{\mathbf{Y}}$ to be $\nu(\cdot \cap \{y \in D : |y_T| > 1\})$, meaning we condition on an extreme value at another time point T. If ν is shift-invariant, choosing $T = 0$ has the same effect as any other T, in the following sense: Let $\mathbf{Y}^T$ be a random variable with law $\nu(\cdot \cap \{y \in D : |y_T| > 1\})$. Then by shift-invariance of ν we have

$$\mathbb{P}[\mathbf{Y}^T \in A] = \nu(A \cap \{|y_T| > 1\}) = \nu(B^{-T}A \cap \{|y_0| > 1\}) = \mathbb{P}[B^T\mathbf{Y} \in A]$$

and hence $\mathbf{Y}^T \overset{d}{=} B^T\mathbf{Y}$.
If ν is not shift-invariant, such a relationship between $\mathbf{Y}^T$ and $\mathbf{Y}$ cannot be established. Hence in this case one does not define a single tail process $\mathbf{Y}$ but rather a collection of so-called *local* tail processes $\mathbf{Y}^T$ with laws $p_T\nu(\cdot \cap \{|y_T| > 1\})$ for some $p_T > 0$. (see [BHS22] Section 3). However, we will not follow this approach here and will only consider shift-invariant tail measures ν and hence restrict ourselves to one single tail process $\mathbf{Y}$.

As already noted, $|Y_0|$ has a Pareto distribution with index α, where α is the index of ν. More precisely, Y_0 is exactly the random variable generating the exponent measure of X_0 since

$$\begin{aligned}\mathbb{P}[Y_0 \in A] &= \lim_{u\to\infty} \frac{\mathbb{P}[X \in u \cdot \pi_0^{-1}(A), |X_0| > u]}{\mathbb{P}[|X_0| > u]} \\ &= \nu(\pi_0^{-1}(A \cap \{x \in \mathbb{R}^d : |x| > 1\})) = \nu_0(A \cap \{x \in \mathbb{R}^d : |x| > 1\})\end{aligned}$$

where ν_0 is the exponent measure of X_0 and A is a $\mathbb{P}^{Y_0}$-continuity set and thus $\pi_0^{-1}(A \cap \{x \in \mathbb{R}^d : |x| > 1\})$ is a ν-continuity set.
It follows that Θ_0 is concentrated on the unit sphere $\mathbb{S}^{d-1}$ and its distribution is the spectral measure Λ of X_0 as defined in Section 2.2.

Example 4.6. AR(1) process
Consider the time series $(M_n)_{n\in\mathbb{N}}$ given by

$$M_{n+1} = \rho M_n + \epsilon_{n+1}$$

for $\rho \in (0,1)$ and i.i.d. real-valued $(\epsilon_n)_{n\in\mathbb{N}}$. Suppose ϵ_0 is regularly varying. If $\mathbb{E}[\log(|\epsilon_0|) \vee 0] < \infty$, there exists a stationary distribution of $\mathbf{M}$, see [KS20] Chapter 14.3.4, and we may extend $\mathbf{M}$ to be defined for all $n \in \mathbb{Z}$.
Now define a homogeneous Poisson process $(N_t)_{t\geq 0}$ with intensity $\lambda > 0$ and define $X_t = M_{N_t}$. Similarly, $\mathbf{X}$ admits a stationary distribution and we can extend $\mathbf{X}$ to be defined for all $t \in \mathbb{R}$, meaning $X_t = M_{\tilde{N}_t}$ for some process $(\tilde{N}_t)_{t\in\mathbb{R}}$.
The process $\mathbf{X}$ is regularly varying in D, which we will prove as a consequence of Theorem 4.11. Its tail process will be determined by means of Corollary 4.12 together with Theorem 5.5. However, we can already conjecture what the tail process might look like based on the single big jump heuristic:
What might cause a large value of $|X_0|$?
Heuristically, the summand ρM_n cannot cause M_{n+1} to be large if M_n was not already large. Hence, any large jump must be caused solely by the noise terms $(\epsilon_n)_{n\in\mathbb{N}}$. By the single big jump heuristic we know that at most one ϵ_K will be large at any sample, for some random integer K. If we condition on $|M_0|$ being large, this large jump must have taken place before $t = 0$, meaning $K \leq 0$. Before and after the jump has occured, the noise terms $(\epsilon_{K-n})_{n\geq 1}$ and $(\epsilon_{K+n})_{n\geq 1}$ will be comparatively small and therefore for $n < K$ the process will be approximately equal to 0 and for $n > K$ will decay towards 0 as $M_{n+1} \approx \rho M_n$. Hence we conjecture the tail process $\hat{\mathbf{Y}}$ of $\mathbf{M}$ to be of the form

$$\hat{Y}_k = \rho^k Y_0 \cdot \mathbb{1}\{k \geq K\}.$$

Since the process $\mathbf{M}$ and the Poisson process $\tilde{N}$ are independent, the tail process of $\mathbf{X}$ should be given by

$$Y_t = \rho^{\tilde{N}_t} Y_0 \cdot \mathbb{1}\{\tilde{N}_t \geq K\}.$$

To make these arguments rigourous, we will need Theorem 4.11 together with Corollary 4.12.

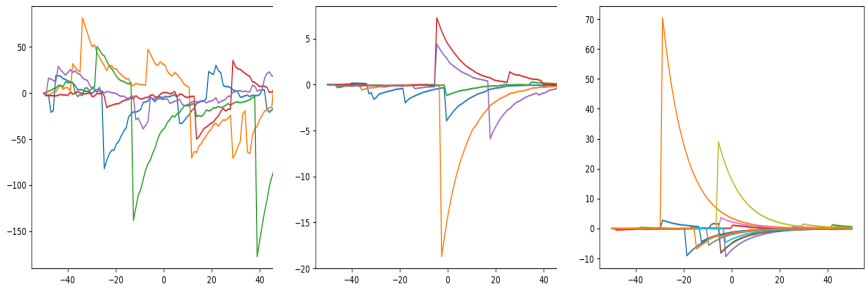
Example 4.7. Let $\mathbf{X}$ be a Lévy-driven Ornstein-Uhlenbeck process for a symmetric α-stable Lévy process with $\alpha < 2$, meaning it satisfies the stochastic differential equation

$$dX_t = -\theta X_t dt + \sigma dL_t.$$

Heuristically, the term $-\theta X_t dt$ makes $\mathbf{X}$ revert to 0 while the term σdL_t is an additive noise term. The process $\mathbf{X}$ admits a stationary distribution, see [RAS19] Theorem 2.17.
Its tail process, as a concequence of Theorem 5.10, is given by

$$Y_t = Y_0 e^{-\theta t} \mathbb{1}\{t \geq -U\}$$

where $Y_0 = R\Theta_0$ with $R \sim \text{Par}(\alpha)$ and $\mathbb{P}[\Theta_0 = 1] = \mathbb{P}[\Theta_0 = -1] = 1/2$ independent of R. The variable $-U$ is independent of Y_0 and follows an $\text{Exp}(\alpha\theta)$-distribution.
The tail process tells us that if $|X_0|$ is large, there must have been a time point $-U$ such that $\mathbf{X}$, and equivalently the Lévy process L, had a large jump. The variable Θ_0 tells us whether there was an upward jump or a downward jump. Since the Lévy process was assumed to be symmetric, both options are equally likely, hence $\mathbb{P}[\Theta_0 = 1] = \mathbb{P}[\Theta_0 = -1]$. Before $t = -U$ no large jump occured; and for $t > -U$ the process $\mathbf{X}$ decays exponentially.
Figure 4 illustrates the convergence $\mathcal{L}(\mathbf{X}/u \big| |X_0| > u) \to \mathcal{L}(\mathbf{Y})$. Note also the similar shape of the sample paths of the tail process $\mathbf{Y}$ and the maxima of $\mathbf{X}$ shown in Figure 3. This similarity will be explained by Theorem 4.19 together with Corollary 4.20



(a) $\mathbf{X}$ (b) $\mathbf{X}/100 \big| |X_0| > 100$ (c) $\mathbf{Y}$

Figure 4: Samples of the Lévy-driven Ornstein-Uhlenbeck process $\mathbf{X}$ (Example 4.7) and its tail process $\mathbf{Y}$

Using the homogeneity of ν, we can find an analogue of the spectral decomposition of exponent measures, given in Theorem 2.5, for tail measures. The following Theorem was given in [Sou22] Equation (2.2) for tail measures on D. See also [KS20] Proposition 5.2.1 for the same statement for tail measures on $(\mathbb{R}^d)^{\mathbb{Z}}$.

Theorem 4.8. *Let ν be a tail measure and Θ be the corresponding spectral tail process. Then for any bounded or non-negative measurable map H we have*

$$\int_D H(y)\mathbb{1}\{|y_0| > 0\}\nu(dy) = \int_0^\infty \mathbb{E}[H(r\Theta)]\alpha r^{-(\alpha+1)}dr.$$

In particular, for the tail process $\boldsymbol{Y}$ we have

$$\mathbb{E}[H(\boldsymbol{Y})] = \int_{\mathcal{D}} H(y)\mathbb{1}\{|y_0| > 1\}\nu(dy) = \int_1^\infty \mathbb{E}[H(u\Theta)]\alpha u^{-(\alpha+1)}du$$

from which it follows that $|Y_0|$ and Θ are independent.

Proof. The proof follows the arguments given in [KS20] Proposition 5.2.1.
Define $\mathbb{S} := \{y \in D : |y_0| = 1\}$ and $D^* = \{y \in D : |y_0| > 0\}$. For the bijective map

$$\psi : (0, \infty) \times \mathbb{S} \to D^*, \psi(r, \theta) = r \cdot \theta$$

define the measure σ on $\mathbb{S}$ as $\sigma(B) = \nu \circ \psi((1, \infty) \times B)$. Then by homogeneity of ν we get for every $r > 0$:

$$\nu\left(\left\{y \in D : |y_0| > r, \frac{y}{|y_0|} \in B\right\}\right) = r^{-\alpha}\nu\left(\left\{y \in D : |y_0| > 1, \frac{y}{|y_0|} \in B\right\}\right)$$

and hence ν can be seen a product measure, i.e. $\nu \circ \psi((r, \infty) \times B) = r^{-\alpha}\sigma(B)$. Therefore for any bounded measurable $H : D \to \mathbb{R}$ we get

$$\int_D H(y)\mathbb{1}\{|y_0| > 0\}\nu(dy) = \int_0^\infty \int_{\mathbb{S}} H(r\theta)\sigma(d\theta)\alpha r^{-(\alpha+1)}dr$$

and for H of the form $H(y) = f(|y_0|) \cdot g(y/|y_0|)$ we get

$$\begin{aligned}
\mathbb{E}[f(|Y_0|) \cdot g(\Theta)] &= \int_D H(y)\mathbb{1}\{|y_0| > 1\}\nu(dy) \\
&= \int_1^\infty \int_{\mathbb{S}} f(r)g(\theta)\sigma(d\theta)\alpha r^{-(\alpha+1)}dr \\
&= \int_1^\infty f(r)\alpha r^{-(\alpha+1)}dr \cdot \int_{\mathbb{S}} g(\theta)\sigma(d\theta) = \mathbb{E}[f(|Y_0|)] \cdot \mathbb{E}[g(\Theta)]
\end{aligned}$$

This proves independence of $|Y_0|$ and Θ.
Therefore for any bounded or non-negative measurable H we get:

$$\mathbb{E}[H(\mathbf{Y})] = \int_1^\infty \int_{\mathbb{S}} H(r\theta)\sigma(d\theta)\alpha r^{-(\alpha+1)}dr = \int_1^\infty \mathbb{E}[H(r\Theta)]\alpha r^{-(\alpha+1)}dr$$

□

The shift-invariance of ν allows us to establish a simple relationship between the forward tail process $(Y_t)_{t\geq 0}$ and the backward tail process $(Y_{-t})_{t\geq 0}$ and similarly between the forward and backward spectral tail processes. This relationship is called the *Time-change formula* and was originally shown in [BS09] Theorem 2.2 for the spectral tail process Θ and extended in [PS18] to the tail process $\mathbf{Y}$ for time series. In [Sou22] Lemma 2.2 the following version of a Time-change formula for tail processes in D was given:

Theorem 4.9. *Time-change formula*
Let ν be a shift-invariant tail measure on D with tail index α and associated tail

process $\boldsymbol{Y}$ *and spectral tail process* Θ*. For every non-negative measurable map* $g : D \to \mathbb{R}$ *and* $x > 0$ *we have:*

$$\mathbb{E}\big[g(\boldsymbol{Y})\mathbb{1}\{|Y_t| > x\}\big] = x^{-\alpha}\mathbb{E}\big[g(xB^t\boldsymbol{Y})\mathbb{1}\{|xY_{-t}| > 1\}\big]$$
$$\mathbb{E}\left[g\left(\frac{\Theta}{|\Theta_t|}\right)\cdot|\Theta_t|^\alpha\right] = \mathbb{E}\big[g(B^t\Theta)\mathbb{1}\{|\Theta_{-t}| \neq 0\}\big]$$

The expression $g(\Theta/|\Theta_t|)\cdot|\Theta_t|^\alpha$ *is supposed to be taken as* 0 *whenever* $|\Theta_t| = 0$*.*

The proof of the former equation is given in [Sou22] Lemma 2.2, however the proof of the latter equation is omitted. For completeness we show the whole proof here:

Proof. The following proof is analogous to the one given in [KS20], where the result was proven for time series.
First, we prove the Time-change formula for $\mathbf{Y}$. Let $x > 0$ and $t \in \mathbb{R}$ be arbitrary, then

$$\begin{aligned}
\mathbb{E}[g(\mathbf{Y})\mathbb{1}\{|Y_t| > x\}] &= \int_D g(y)\mathbb{1}\{|y_t| > x\}\mathbb{1}\{|y_0| > 1\}\nu(dy)\\
\text{(shift-invariance)} &= \int_D g(B^t y)\mathbb{1}\{|y_0| > x\}\mathbb{1}\{|y_{-t}| > 1\}\nu(dy)\\
\text{(homogeneity)} &= x^{-\alpha}\int_D g(xB^t y)\mathbb{1}\{|y_0| > 1\}\mathbb{1}\{|y_{-t}| > x^{-1}\}\nu(dy)\\
&= x^{-\alpha}\mathbb{E}[g(xB^t\mathbf{Y})\mathbb{1}\{|Y_{-t}| > x^{-1}\}]
\end{aligned}$$

Now for the Time-change formula for Θ we first show the result for 0-homogeneous maps g, that is a map with $g(xy) = x^0 g(y) = g(y)$ for every $x \in \mathbb{R}\setminus\{0\}$. Let g be bounded measurable and 0-homogeneous, then

$$\begin{aligned}
\mathbb{E}[g(xB^t\mathbf{Y})\mathbb{1}\{|Y_{-t}| > 1/x\}] &= \mathbb{E}[g(x|Y_0|B^t\Theta)\mathbb{1}\{|Y_{-t}| > 1/x\}]\\
&= \mathbb{E}[g(B^t\Theta)\mathbb{1}\{|Y_{-t}| > 1/x\}]\\
&\xrightarrow[x\to\infty]{} \mathbb{E}\big[g(B^t\Theta)\mathbb{1}\{|\Theta_{-t}| \neq 0\}\big]
\end{aligned}$$

where the convergence is justified by the monotone convergence theorem.
Now applying the representation from Theorem 4.8 to the function $H(y) =$

$g(y)\mathbb{1}\{|y_t| > x\}$ we get:

$$\begin{aligned}
x^\alpha \mathbb{E}[g(\mathbf{Y})\mathbb{1}\{|Y_t| > x\}] &= x^\alpha \int_1^\infty \mathbb{E}\Big[g(r\Theta)\mathbb{1}\{|r\Theta_t| > x\}\Big]\alpha r^{-(\alpha+1)}dr \\
&= x^\alpha \mathbb{E}\Big[g(\Theta)\int_1^\infty \mathbb{1}\{r > x/|\Theta_t|\}\alpha r^{-(\alpha+1)}dr\Big] \\
&= x^\alpha \mathbb{E}[g(\Theta)(1 \wedge |\Theta_t|/x)^\alpha] \\
&= \mathbb{E}[g(\Theta)(x \wedge |\Theta_t|)^\alpha] \\
&\xrightarrow[x\to\infty]{} \mathbb{E}[g(\Theta)\cdot|\Theta_t|^\alpha] = \mathbb{E}\Big[g\Big(\frac{\Theta}{|\Theta_t|}\Big)\cdot|\Theta_t|^\alpha\Big]
\end{aligned}$$

Using the first equality we can see that these two limits must be the same, proving the second equality for 0-homgeneous maps.
Now we can show the result for general non-negative measurable maps g using 0-homogeneous maps: Any map g can be transformed into a 0-homogeneous map h by defining $h(y) = g(y/|y_t|)\mathbb{1}\{y_t \neq 0\}$ for any $t \in \mathbb{R}$. Note that $g(B^t(\Theta/|\Theta_t|)) = g(|\Theta_0|^{-1}B^t\Theta)$. Now we get that

$$\begin{aligned}
\mathbb{E}\Big[g\Big(\frac{\Theta}{|\Theta_t|}\Big)\cdot|\Theta_t|^\alpha\Big] &= \mathbb{E}\Big[h(\Theta)\cdot|\Theta_t|^\alpha\Big] \\
&= \mathbb{E}\Big[h(B^t\Theta)\mathbb{1}\{|\Theta_{-t} \neq 0\}\Big] \\
&= \mathbb{E}\Big[g(|\Theta_0|^{-1}g(B^t\Theta)\mathbb{1}\{\Theta_{-t} \neq 0\}\Big] \\
&= \mathbb{E}\Big[g(B^t\Theta)\mathbb{1}\{\Theta_{-t} \neq 0\}\Big]
\end{aligned}$$

where we used the result of step 1 for the second equality and the fact that $|\Theta_0| = 1$ for the last equality. □

The Time-change formula can be used to derive a few simple properties of the tail and spectral tail process. We will list two here:

Example 4.10.

- Pick $g = 1$, then the time change formula for Θ reduces to

$$\mathbb{E}[|\Theta_t|^\alpha] = \mathbb{E}[\mathbb{1}\{|\Theta_{-t} \neq 0\}] = \mathbb{P}[\Theta_{-t} \neq 0]$$

  This property has a simple explanation ([KS20] Corollary 5.3.5): Since the spectral tail process describes the extremal dependence of the process $\mathbf{X}$ given an extreme value at time 0, Θ_t being equal to 0 means that X_t and X_0 are extremally independent, meaning an extreme value of X_0 cannot cause any extreme value of X_t. But by stationarity of $\mathbf{X}$ this means that an extreme value of X_0 cannot be caused by an extreme value of X_{-t} and hence Θ_{-t} must also be equal to 0.

- To establish a relationship between the forward and backward spectral tail processes, pick any function $g(\theta)$ which only depends on θ_0 and θ_t, then

$$\mathbb{E}\left[g\left(\frac{\Theta_t}{|\Theta_t|}, \frac{\Theta_0}{|\Theta_t|}\right)|\Theta_t|^\alpha\right] = \mathbb{E}\big[g(\Theta_0, \Theta_{-t})\mathbb{1}\{\Theta_{-t} \neq 0\}\big]$$

  We will use this relation between (Θ_t, Θ_0) and (Θ_0, Θ_{-t}) in Section 5 to derive a transition function of the backward (spectral) tail process knowing the transition function of the forward (spectral) tail process, assuming both processes are Markov processes.

The time change formula will be an essential tool used to reconstruct ν given its tail process $\mathbf{Y}$. We will show how to do this in Section 4.2 and Section 4.3. Before we do this, we present a necessary and sufficient condition for regular variation in D, one which will make verifying regular variation of a process easier than using Definition 4.1 alone. To motivate this, remember that in Section 3 we saw that finite-dimensional weak convergence of processes does not suffice for the processes to weakly converge in D. Instead, an additional tightness condition using the modulus of continuity w' was necessary to show weak convergence in D, see Theorem 3.11.
We want to extend this concept to regular variation in D. This was done in [HL05] Theorem 10 for processes defined on compact time intervals and extended in [Sou22] Theorem 3.2 for *stochastically continuous* processes. We call a càdlàg process $\mathbf{X}$ stochastically continuous if for every $t_0 \in \mathbb{R}$ we have $\lim_{t\to t_0} \mathbb{P}[|X_t - X_{t_0}| \geq \epsilon] = 0$ for all $\epsilon > 0$. Then regular variation for such processes can be proven in the following way: 0

Theorem 4.11. *Let $\boldsymbol{X}$ be a stationary, stochastically continuous càdlàg process. Then the following are equivalent:*

1. *$\boldsymbol{X}$ is regularly varying in D.*
2. *$\boldsymbol{X}$ is finite-dimensional regularly varying and for all $a < b$ and $\epsilon > 0$ we have*

$$\lim_{\delta\to 0} \limsup_{x\to\infty} \frac{\mathbb{P}\big[w'(\boldsymbol{X}, a, b, \delta) > x\epsilon\big]}{\mathbb{P}[|X_0| > x]} = 0$$

For the proof, see [Sou22] Theorem 3.2.

Note that this again proves that $\mathcal{L}(\mathbf{X}/u\big||X_0| > u) \to \mathcal{L}(\mathbf{Y})$ since condition 2 of Theorem 4.11 clearly implies condition 2 of Theorem 3.11 for a sequence of random variables $\mathbf{X}^{(n)}$ with law $\mathcal{L}(\mathbf{X}/n\big||X_0| > n)$. Condition 1 of Theorem 3.11 is satisfied since if $\mathbf{X}$ is regularly varying we have

$$\lim_{a\to\infty} \limsup_{n\to\infty} \frac{\mathbb{P}[\|\mathbf{X}\|_{[-t,t]} > na]}{\mathbb{P}[|X_0| > n]} = \lim_{a\to\infty} a^{-\alpha}\nu(\{y \in D : \|y\|_{[-t,t]} > 1\}) = 0$$

since $\nu(\{y \in D : \|y\|_{[-t,t]} > 1\}) < \infty$ because ν is $\mathcal{B}_0$-boundedly finite.

Now we briefly present a simple application of this Theorem to show how it can be utilized to verify regular variation:

Corollary 4.12. *Let $\boldsymbol{M} = (M_n)_{n\in\mathbb{Z}}$ be a regularly varying stationary time series with discrete-time tail process $\hat{\boldsymbol{Y}}$. Let $(N_t)_{t\geq 0}$ be a homogeneous Poisson process with intensity $\lambda > 0$. Define the process $\boldsymbol{X}$ by $X_t = M_{N_t}$ for $t > 0$. Then $X_t = M_{N_t}$ can be extended to be defined for all $t \in \mathbb{R}$ and $\boldsymbol{X}$ is regularly varying in D.*
Furthermore $\boldsymbol{X}$ has tail process $(\hat{Y}_{N_t})_{t\in\mathbb{R}}$.

Proof. If $\mathbf{X}$ is stationary, it can be extended to $t \in \mathbb{R}$ If $\mathbf{M}$ is regularly varying, it is finite-dimensional regularly varying. Since the distribution of X_t is a mixture of the identically distributed regularly varying vectors $(\dots, M_{-1}, M_0, M_1, \dots)$ it follows that $\mathbf{X}$ must be finite-dimensional regularly varying.
Note that

$$\mathbb{P}[|X_t - X_{t_0}| \geq \epsilon] \leq \mathbb{P}[|N_t - N_{t_0}| \geq 1] \to 0$$

as $t \to t_0$. Hence $\mathbf{X}$ is stochastically continuous.
Let $a < b$ be arbitrary and define Z_δ as the number of jumps (T_i, T_{i+1}) of N in the interval $[a, b]$ with time difference at most δ. The modulus of continuity $w'(\mathbf{X}, a, b, \delta)$ is greater than $x\epsilon$ only if $Z_\delta = k > 0$ and there are $m, n \leq k$ such that $|M_m - M_n| > x\epsilon$. By Markov's inequality we have the upper bound

$$\mathbb{P}[Z_\delta \geq k] \leq \frac{\mathbb{E}|Z_\delta|^{2+\eta}}{k^{2+\eta}} \leq \frac{\mathbb{E}\big[|N_b - N_a|^{2+\eta}\big]}{k^{2+\eta}} < \infty$$

for any $\eta > 0$ and thus since $Z_\delta \to 0$ almost-surely as $\delta \to 0$ it follows by dominated convergence that $\mathbb{E}Z_\delta^{2+\eta} \to 0$ as $\delta \to 0$.
Finally, we get

$$\begin{aligned}
&\mathbb{P}[w'(\mathbf{X}, a, b, \delta) > x\epsilon] \\
&= \sum_{k=0}^{\infty} \mathbb{P}[w'(\mathbf{X}, a, b, \delta) > x\epsilon | Z_\delta = k]\mathbb{P}[Z_\delta = k] \\
&\leq \sum_{k=1}^{\infty} \Big(\mathbb{P}[|M_{k+1} - M_k| > x\epsilon] + \cdots + \mathbb{P}[|M_1 - M_0| > x\epsilon]\Big)\mathbb{P}[Z_d \geq k] \\
&\leq \sum_{k=0}^{\infty} (k+1) \cdot \mathbb{P}[|M_0| > x\epsilon/2] \frac{\mathbb{E}Z_\delta^{2+\eta}}{k^{2+\eta}}
\end{aligned}$$

and therefore

$$\lim_{\delta\to 0}\limsup_{x\to\infty}\frac{\mathbb{P}[w'(\mathbf{X},a,b,\delta)>x\epsilon]}{\mathbb{P}[|X_0|>x]}$$
$$\leq \lim_{\delta\to 0}\limsup_{x\to\infty}\sum_{k=1}^{\infty}(k+1)\frac{\mathbb{P}[|M_0|>x\epsilon/2]}{\mathbb{P}[|M_0|>x]}\cdot\frac{\mathbb{E}Z_\delta^{2+\eta}}{k^{2+\eta}} \leq c\cdot\lim_{\delta\to 0}\mathbb{E}Z_\delta^{2+\eta}=0$$

for some positive constant c. By Theorem 4.10 $\mathbf{X}$ is regularly varying.
Now the tail process can be computed by

$$\mathbb{P}[u^{-1}X_t\in A\big||X_0|>u]=\sum_{k\in\mathbb{Z}}\mathbb{P}[u^{-1}M_k\in A\big||M_0|>u]\cdot\mathbb{P}[N_t=k]$$
$$\to\sum_{k\in\mathbb{Z}}\mathbb{P}[\hat{Y}_k\in A]\cdot\mathbb{P}[N_t=k]=\mathbb{P}[\hat{Y}_{N_t}\in A]$$

due to independence of $\mathbf{M}$ and N. □

This confirms that the process defined in Example 4.6 is regularly varying in D and its tail process is the one we conjectured based on the single big jump heuristic.
More importantly, Corollary 4.12 essentially tells us that the regular variation of the continuous-time process $\mathbf{X}$ is entirely determined by the regular variation of the time series $\mathbf{M}$. As a consequence, for such processes $\mathbf{X}$ as given in Corollary 4.12 finite-dimensional regular variation is equivalent to regular variation in D.

4.2 Spectral Representation

In Lemma 4.4 we have seen that the tail process uniquely determines its tail measure. In this section, we want to make this connection more concrete, meaning we want to find a way to construct ν given its spectral tail process Θ. Since Theorem 4.8 showed how to compute $\nu(\cdot\cap\{|y_0|>0\})$ given its spectral tail process Θ, and similarly how to compute $\nu(\cdot\cap\{|y_T|>0\})$ using $B^T\Theta$, the idea is to consider all shifted versions $(B^T\Theta)_{T\in\mathbb{R}}$ of the spectral tail process and hope that by summing all $\nu(\cdot\cap\{|y_T|>0\})$ we can compute $\nu(\cdot)$. Such a sum and the processes $B^T\Theta$ need to be normalized in an appropriate way to make this work. The shifted processes $B^T\Theta$ will be combined into one process $\mathbf{Z}$, which will then determine ν completely.
This leads us to the following Theorem, which was given in [Sou22] Theorem 2.3:

Theorem 4.13. *Spectral Representation*
A Borel measure ν on D is a shift-invariant tail measure with index $\alpha>0$ if and only if there exists a stochastic process $\boldsymbol{Z}$ in D, called a spectral process *for*

ν, such that $\mathbb{P}[\boldsymbol{Z} = 0] = 0$, $\mathbb{E}[|Z_0|^\alpha] = 1$,

$$\nu(A) = \int_0^\infty \mathbb{P}[u\boldsymbol{Z} \in A]\alpha u^{-(\alpha+1)} du \tag{Spec}$$

for all $a < b, t \in \mathbb{R}$

$$0 < \mathbb{E}\left[\sup_{a \le s \le b} |Z_s|^\alpha\right] < \infty$$

and for bounded measurable 0-homogeneous maps $g : D \to \mathbb{R}$:

$$\mathbb{E}[|Z_t|^\alpha g(\boldsymbol{Z})] = \mathbb{E}[|Z_0|^\alpha g(B^t \boldsymbol{Z})] \tag{TSF}$$

The tail process $\boldsymbol{Y}$ of ν is given by

$$\mathbb{P}^{\boldsymbol{Y}} = \mathbb{E}\left[|Z_0|^\alpha \delta_{\frac{RZ}{|Z_0|^\alpha}}\right]$$

where R is a random variable following a Pareto distribution with index α, independent of $\boldsymbol{Z}$.

Remark:
If we exclude the "tilt shift formula" (TSF) we still obtain a tail measure with tail index α; however shift-invariance of ν is not guaranteed anymore, see [BHS22] Section 3.4 for a version of Theorem 4.13 for general tail measures based on local tail processes.
As already shown in [DHS18], the condition (TSF), which is a weaker condition than $\mathbf{Z}$ being stationary, is equivalent to the following conditions:

1. For all α-homogeneous maps $g : D \to \mathbb{R}$ and $t \in \mathbb{R}$ we have

$$\mathbb{E}[g(\mathbf{Z})] = \mathbb{E}[g(B^t\mathbf{Z})]$$

2. For all non-negative measurable maps $k : D \to \mathbb{R}$ and $t \in \mathbb{R}$ we have

$$\mathbb{E}[|Z_0|^\alpha k(|Z_0|^{-1}B^t\mathbf{Z})] = \mathbb{E}[|Z_t|^\alpha k(|Z_t|^{-1}Z)]$$

where $k(|Z_t|^{-1}Z)$ is taken to be 0 if $|Z_t| = 0$.

Proof. Suppose ν is given by (Spec). We have to prove that ν is a shift-invariant tail measure. The set function ν satisfies the standard properties of a measure; meaning non-negative and countably additive; because $\mathbb{P}[u\mathbf{Z} \in \cdot]$ does. We only need to prove (1) $\nu(\{0\}) = 0$, (2) homogeneity, (3) the normalization $\nu(\{|y_0| > 1\}) = 1$, (4) shift-invariance and (5) $\mathcal{B}_0$-finiteness and $\nu \neq 0$.

Of course, (1) is satisfied because $\mathbb{P}[\mathbf{Z} = 0] = 0$. Now let $A \subseteq D$ and $t > 0$ then by a simple subsitution we get

$$\nu(tA) = \int_0^\infty \mathbb{P}[t^{-1}u\mathbf{Z} \in A]\alpha u^{-(\alpha+1)}du$$
$$(u \leftarrow u/t) = t^{-\alpha}\int_0^\infty \mathbb{P}[u\mathbf{Z} \in A]\alpha u^{-(\alpha+1)}du = t^{-\alpha}\nu(A)$$

which shows (2).
Now since $\mathbb{E}|Z_0|^\alpha = 1$ we can show (3) by

$$1 = \mathbb{E}|Z_0|^\alpha = \int_0^\infty \mathbb{P}[|Z_0|^\alpha > s]ds = \int_0^\infty \mathbb{P}[|Z_0| > s^{1/\alpha}]ds$$
$$(u = s^{-1/\alpha}) = \int_0^\infty \mathbb{P}[|Z_0| > u^{-1}]\alpha u^{-(\alpha+1)}du = \nu(\{|y_0| > 1\})$$

The proof of shift-invariance relies on (TSF). As mentioned before, (TSF) is equivalent to $\mathbb{E}g(\mathbf{Z}) = \mathbb{E}g(B^t\mathbf{Z})$ for all α-homogeneous maps g and thus ν is shift-invariant.
Finally, we prove $\mathcal{B}_0$-finiteness (5) by utilizing $0 < \mathbb{E}\sup_{a\le s\le b}|Z_s|^\alpha < \infty$: Let A be bounded away from 0, meaning there exist $a, \epsilon > 0$ such that $\|y\|_{[-a,a]} > \epsilon$ for all $y \in A$. Therefore

$$\begin{aligned}\nu(A) &\le \nu(\{\|y\|_{[-a,a]} > \epsilon\}) = \epsilon^\alpha\nu(\{\|y\| > 1\}) \\ &= \epsilon^\alpha\int_0^\infty \mathbb{P}[\|\mathbf{Z}\|_{[-a,a]} > u^{-1}]\alpha u^{-(\alpha+1)}du \\ (s = u^{-\alpha}) &= \epsilon^\alpha\int_0^\infty \mathbb{P}[\|\mathbf{Z}\|_{[-a,a]} > s^{1/\alpha}]ds = \epsilon^\alpha\mathbb{E}\left[\sup_{-a\le s\le a}|Z_s|^\alpha\right] \in (0,\infty)\end{aligned}$$

which proves (5).
Therefore ν is a shift-invariant tail measure.

The proof of the converse is given in [Sou22] Theorem 2.3. For completeness, we repeat the arguments here:
Let ν be a shift-invariant tail measure and let Θ be the spectral tail process of ν. Let q be an arbitrary strictly positive Riemann density on $\mathbb{R}$. We define the norm

$$\|\theta\|_{q,\alpha}^\alpha := \int_{\mathbb{R}} q(t)|\theta_t|^\alpha dt$$

Then since $\mathbb{E}|\Theta_t|^\alpha \le 1$ (see Example 4.9) we get that $\mathbb{E}\|B^t\Theta\|_{q,\alpha}^\alpha \le 1$ and therefore $\|B^t\Theta\|_{q,\alpha} < \infty$ almost-surely for all t. Now if θ is càdlàg and $|\theta_s| = 1$ there must exist a $\delta > 0$ such that $|\theta_t|^\alpha \ge \epsilon$ for all $|t - s| \le \delta$ and thus

$$\int_{\mathbb{R}} q(t)|\theta_{t-s}|^\alpha dt \ge \int_{-\delta}^{\delta} q(t)\epsilon > 0$$

from which it follows that $0 < \|B^t\Theta\|_{q,\alpha}$ almost surely for all t. This allows us to construct the random variables

$$Z^{(t)} = \frac{B^t\mathbf{Y}}{\|B^t\mathbf{Y}\|_{q,\alpha}} = \frac{B^t\Theta}{\|B^t\Theta\|_{q,\alpha}}$$

and the random variable $\mathbf{Z} = Z^{(\tau)}$ for a real-valued random variable τ with density q. We now need to verify that Z is a spectral process and show that the measure defined by

$$\nu_q(A) = \int_0^\infty \mathbb{P}[u\mathbf{Z} \in A]\alpha u^{-(\alpha+1)}du$$

is equal to ν. We have already proven that ν_q is again a shift-invariant tail measure.
Let g be a bounded or non-negative measurable map on D and let $\epsilon > 0$ then we get

$$\begin{aligned}
&\int_D g(y)\mathbb{1}\{|y_t| > \epsilon\}\nu_q(dy) \\
&\quad \text{(definition } \nu_q) = \int_0^\infty \mathbb{E}\big[g(uZ)\mathbb{1}\{u|Z_t| > \epsilon\}\big]\alpha u^{-(\alpha+1)}du \\
&\quad \text{(definition } Z) = \int_{\mathbb{R}}\int_0^\infty \mathbb{E}\left[g\left(\frac{uB^s\mathbf{Y}}{\|B^s\mathbf{Y}\|_{q,\alpha}}\right)\mathbb{1}\left\{\frac{u|Y_{t-s}|}{\|B^s\mathbf{Y}\|_{q,\alpha}} > \epsilon\right\}\right]\alpha u^{-(\alpha+1)}q(s)duds \\
&(y \leftarrow \epsilon^{-1}\|B^s y\|_{q,\alpha}y) = \epsilon^{-\alpha}\int_{\mathbb{R}}\int_0^\infty \mathbb{E}\left[\frac{g(\epsilon uB^{s-t}B^t\mathbf{Y})\mathbb{1}\{u|Y_{s-t}| > 1\}}{\|B^{s-t}B^s\mathbf{Y}\|_{q,\alpha}^\alpha}\right]\alpha u^{-(\alpha+1)}q(s)duds \\
&\quad \text{(TCF)} = \epsilon^{-\alpha}\int_{\mathbb{R}}\int_0^\infty \mathbb{E}\left[\frac{g(\epsilon B^t\mathbf{Y})\mathbb{1}\{|Y_{s-t}| > u\}}{\|B^t\mathbf{Y}\|_{q,\alpha}^\alpha}\right]\alpha u^{\alpha-1}q(s)duds \\
&\quad \text{(Fubini + } (*)) = \epsilon^{-\alpha}\int_{\mathbb{R}} \mathbb{E}\left[\frac{g(\epsilon B^t\mathbf{Y})|Y_{s-t}|^\alpha}{\|B^t\mathbf{Y}\|_{q,\alpha}^\alpha}\right]q(s)ds \\
&\quad \text{(Fubini + } (**)) = \epsilon^{-\alpha}\mathbb{E}[g(\epsilon B^t\mathbf{Y})] \\
&\qquad\qquad = \epsilon^{-\alpha}\int_D g(\epsilon B^t y)\mathbb{1}\{|y_0| > 1\}\nu(dy) = \int_D g(y)\mathbb{1}\{|y_t| > \epsilon\}\nu(dy)
\end{aligned}$$

where in $(*)$ we used the fact that $|y|^\alpha = \int_0^\infty \mathbb{1}\{|y| > u\}\alpha u^{-(\alpha+1)}du$ and in $(**)$ we used $\|B^t\mathbf{Y}\|_{q,\alpha}^{-\alpha} \cdot \int_{\mathbb{R}} |Y_{s-t}|^\alpha q(s)ds = 1$ by definition.

Now define $\mathcal{E}_q(y) = \int_{\mathbb{R}} \mathbb{1}\{|y_t| > 1\}q(t)dt$ and let g be a bounded or non-negative measurable map on D with $g(y) = 0$ if $y^* = \sup|y_t| \le \epsilon$. By Lemma 4.4 such maps are measure determining and we have

$$\begin{aligned}
\nu_q(g) &= \int_D \frac{g(y)\mathcal{E}_q(\epsilon^{-1}y)}{\mathcal{E}_q(\epsilon^{-1}y)}\nu_q(dy) = \int_{\mathbb{R}}\int_D \frac{g(y)}{\mathcal{E}_q(\epsilon^{-1}y)}\mathbb{1}\{|y_t| > \epsilon\}q(t)\nu_q(dy)dt \\
&= \int_{\mathbb{R}}\int_D \frac{g(y)}{\mathcal{E}_q(\epsilon^{-1}y)}\mathbb{1}\{|y_t| > \epsilon\}q(t)\nu(dy)dt = \int_D \frac{g(y)\mathcal{E}_q(\epsilon^{-1}y)}{\mathcal{E}_q(\epsilon^{-1}y)}\nu(dy) = \nu(g)
\end{aligned}$$

from which it follows that $\nu_q = \nu$ for all densities q.

Note that this also implies that for any bounded or non-negative measurable function g we have

$$\mathbb{E}g(\mathbf{Y}) = \int_D g(y)\mathbb{1}\{|y_0| > 1\}\nu(dy) = \int_0^\infty \mathbb{E}[g(u\mathbf{Z})\mathbb{1}\{u|Z_0| > 1\}]\alpha u^{-(\alpha+1)}du$$

from which it follows that

$$\mathbb{P}^{\mathbf{Y}} = \mathbb{E}\left[|Z_0|^\alpha \delta_{\frac{R\mathbf{Z}}{|Z_0|^\alpha}}\right]$$

for $R \sim \text{Par}(\alpha)$ independent of $\mathbf{Z}$.

We have already shown in the first part of the proof that $\mathbb{P}[\mathbf{Z} = 0] = 0$ is equivalent to $\nu(\{0\}) = 0$, $\mathbb{E}|Z_0|^\alpha = 1$ is equivalent to $\nu(\{|y_0| > 1\}) = 1$, $0 < \mathbb{E}\sup_{a\le s\le b}|Z_s|^\alpha < \infty$ is equivalent to ν being $\mathcal{B}_0$-boundedly finite and non-zero and the tilt-shift formula is equivalent to ν being shift-invariant. All of these are true since, by assumption, $\nu = \nu_q$ is a tail measure. □

Remark
As we have seen in the proof, the construction of $\mathbf{Z}$ is based on a strictly positive Riemann density q. Hence, spectral processes $\mathbf{Z}$ are not unique in distribution, as every strictly positive Riemann density can be used to define a spectral process.
Note also that a spectral tail process Θ is not necessarily a spectral process $\mathbf{Z}$. However, if $\mathbb{P}[\Theta_t = 0] = 0$ for all t then the time change formula reduces to

$$\mathbb{E}[g(\Theta)|\Theta_t|^\alpha] = \mathbb{E}[g(B^t\Theta)] = \mathbb{E}[g(B^t\Theta)\cdot|\Theta_0|^\alpha]$$

which is the tilt-shift formula. Hence in this case a spectral tail process also defines a spectral process.

Theorem 4.13 shows how to reconstruct ν from its tail process $\mathbf{Y}$. We know that such a tail process $\mathbf{Y}$ must necessarily satisfy $|Y_0| > 1$ and the time change formula (Theorem 4.8). But is every càdlàg process $\mathbf{Y}$ satisfying these two conditions a tail process for a shift-invariant tail measure ν? We can still construct a process $\mathbf{Z}$ as we did in the proof of Theorem 4.13, however we need an additional assumption to ensure that $\mathbf{Z}$ is indeed a spectral process:

Corollary 4.14. *([Sou22] Corollary 2.4)*
Let $\mathbf{Y}$ be a process such that $|Y_0| > 1$ almost-surely and $\mathbf{Y}$ satisfies the time change formula. Additionally, assume that for all $a < b$ we have

$$\int_a^b \mathbb{E}\left[\frac{1}{\int_a^b \mathbb{1}\{|Y_{t-s}| > 1\}dt}\right]ds < \infty$$

Then there exists a shift-invariant tail measure ν such that $\mathbf{Y}$ is the tail process of ν.

For the proof we refer to [Sou22], Corollary 2.4. In short, we construct a spectral process $\mathbf{Z}$ using $\mathbf{Y}$ as we did in Theorem 4.13. Then $\mathbf{Z}$ uniquely defines a tail measure ν. The extra boundedness condition of $\mathbf{Y}$ is necessary to ensure that $\mathbb{E}[\sup_{a\leq s\leq b}|Z_s|^\alpha] < \infty$, which is necessary for $\mathbf{Z}$ to be a spectral process. The boundedness condition of $\mathbf{Y}$ is also necessary, meaning for any tail measure the tail process must satisfy this condition, see [Sou22], Corollary 2.4.

Theorem 4.13 also offers a way to construct a max-stable process $\mathbf{W}$ which is regularly varying with tail measure ν. The spectral representation of ν entails that the Poisson point process N with mean measure ν on D can be expressed as

$$N = \sum_{i=1}^{\infty} \delta_{P_i \mathbf{Z}^{(i)}}$$

where $(P_i, \mathbf{Z}^{(i)})_{i\in\mathbb{N}}$ is an enumeration of points of a Poisson point process on $(0,\infty)\otimes D$ with mean measure $\mu_\alpha \otimes \mathbb{P}^{\mathbf{Z}}$, see [Sou22], page 126. Using this, we can define a process

$$\mathbf{W} = \bigvee_{i=1}^{\infty} P_i \mathbf{Z}^{(i)}$$

for which we see that for $u \succeq 0$ we have

$$\mathbb{P}[\mathbf{W} \preceq u] = \mathbb{P}\big[N\big([0,u]\big) = 0\big] = \exp(-\nu([0,u]^C))$$

which is exactly the limit distribution of the maxima of a process $\mathbf{X}$ with tail measure ν, as described in the introduction of Section 4.

Example 4.15. Brown-Resnick Max-stable Process
Let $(B_t)_{t\in\mathbb{R}}$ be a two-sided Brownian motion, meaning $(B_t)_{t\geq 0}$ and $(B_{-t})_{t\geq 0}$ are independent Brownian motions. Define $W_t = B_t - \alpha|t|/2$ and let $Z = \exp(W_t) = \exp(B_t - \alpha|t|/2)$.
Then $\mathbf{Z}$ is a spectral process, see [BHS22] Example 4.14 and [KS20] Theorem 13.5.1.
Therefore $\mathbf{Z}$ defines a shift-invariant tail measure ν and a max-stable process

$$\mathbf{W} = \bigvee_{i=1}^{\infty} P_i \exp(B_t - \alpha|t|/2)$$

for a Poisson point process $(P_i, Z^{(i)})_{i\in\mathbb{N}}$ with mean measure $\mu_\alpha \otimes \mathbb{P}^{\mathbf{Z}}$. The max-stable process $\mathbf{W}$ is regularly varying with tail process

$$Y_t = R \cdot \exp(B_t - \alpha|t|/2)$$

for $R \sim \text{Par}(\alpha)$ independent of B.
This example is a special case of a Brown-Resnick max-stable process, see also [BHS22], Example 4.14.

4.3 Mixed Moving Average Representation

If we take a look at the tail process $Y_t = Y_0 e^{-\theta t} \mathbb{1}\{t \geq -U\}$ given in Example 4.7, we might recognize that every sample path is just a shifted and scaled version of one single process $\mathbf{Q}$, namely $Q_t = \Theta_0 e^{-\theta t} \mathbb{1}\{t \geq 0\}$. Therefore we let $\mathcal{I} : D \to [-\infty, \infty]$ be the map which takes a process y and gives the first time it reaches its maximum:

$$\mathcal{I}(y) = \inf \left\{ t \in \mathbb{R} : y^* := \sup_{s \in \mathbb{R}} |y_s| \in \{|y_t|, |y_{t^-}|\} \right\}$$

where $y_{t^-} = \lim_{s \nearrow t} y_s$.
Then we can interpret $\mathbf{Q}$ as $\mathbf{Y}/\mathbf{Y}^*$ conditioned on $\mathcal{I}(\mathbf{Y}) = 0$, meaning conditioned on the jump occuring at time $t = 0$. This suggests we might be able to express ν using this even simpler process $\mathbf{Q}$ by simply shifting and scaling $\mathbf{Q}$ accordingly.
This leads us to the following definition:

Definition 4.16. ([Sou22], Equation 2.12)
We say that a tail measure ν has a *mixed moving average representation* if there exists a constant $\vartheta \in (0, \infty)$ and a process $\mathbf{Q} \in D$ such that $\mathbb{P}[\mathbf{Q}^* = 1] = 1$ and

$$\nu(A) = \vartheta \int_{\mathbb{R}} \int_0^\infty \mathbb{P}[rB^t \mathbf{Q} \in A] \alpha r^{-(\alpha+1)} dr dt.$$

One application of such a representation is immeadiate: Note that if A is homogeneous (meaning $tA = A$ for all $t > 0$) and shift-invariant, $\mathbb{P}[rB^t \mathbf{Q} \in A]$ is independent of r and t. Hence $\nu(A) = 0$ if and only if $\mathbb{P}[\mathbf{Q} \in A] = 0$ and $\nu(A) = \infty$ otherwise. We can go further with this relation:

Lemma 4.17. *([Sou22], Lemma 2.8)*
Let ν be a shift-invariant tail measure with tail process $\boldsymbol{Y}$. Let $A \subseteq D$ be homogeneous and shift-invariant. Then $\nu(A) = 0$ if and only if $\mathbb{P}[\boldsymbol{Y} \in A] = 0$.
If ν admits a mixed moving average representation with process $\boldsymbol{Q}$ we have $\nu(A) = 0$ if and only if $\mathbb{P}[\boldsymbol{Q} \in A] = 0$.

The question we will now try to answer is when a tail measure ν admits a mixed moving average representation and how one might be able to compute the constant ϑ and the process $\mathbf{Q}$. The constant ϑ will turn out to be related

to the tail process $\mathbf{Y}$ in the following way:

Definition 4.18. Exceedance Functional, Candidate Extremal Index ([Sou22], Definition 2.6)
The exceedance functional $\mathcal{E}$ on D is a map $D \to [0, \infty]$ defined by

$$\mathcal{E}(y) = \int_{-\infty}^{\infty} \mathbb{1}\{|y_t| > 1\} dt.$$

For a shift-invariant tail measure ν on D with tail process $\mathbf{Y}$ we call the constant

$$\vartheta = \mathbb{E}\left[\frac{1}{\mathcal{E}(\mathbf{Y})}\right]$$

the candidate extremal index of ν.

The case $\vartheta = 0$:
As a simple example for a process $\mathbf{X}$ with tail process $\mathbf{Y}$ having candidate extremal index 0, consider the constant process $X_t = X_0$ for a regularly varying random vector X_0. Its tail process is given by $Y_t = Y_0$ for all t. Since $|Y_0| > 1$ by definition, we immediately see that $\mathcal{E}(Y) = \infty$ and therefore $\vartheta = 0$.
The case $\vartheta = 0$ is called long-range dependence, since, as we can see in this example, an extreme value of X_0 continues to affect the process $\mathbf{X}$ in a non-negligible way.

In the case that $\vartheta \in (0, \infty)$, it can be seen as a measure of how fast the tail process $\mathbf{Y}$ vanishes, meaning if ϑ is very large, $\mathcal{E}(\mathbf{Y})$ must on average be very small, which in turn means $\mathbf{Y}$ does not spend much time outside of the unit disk. We will see in Theorem 4.19 that this is equivalent to $\mathbf{Y}$ vanishing at infinity, meaning $\lim_{|t|\to\infty} Y_t = 0$ almost-surely.

The candidate extremal index is closely related to the classical *extremal index* of univariate extreme value theory (see [HF06] Definition 5.5.9): For a sequence of i.i.d. regularly varying univariate random variables, the rescaled maximum $M_n = \max(|X_1|, \dots, |X_n|)$ converges weakly to a Fréchet distribution, meaning $\mathbb{P}[a_n^{-1} M_n \leq t] \to \exp(-t^{-\alpha})$. However, if we consider a stationary sequence $(X_n)_{n\in\mathbb{N}}$ which may not be independent, we can still get a variant of this limit theorem. Under an additional assumption (see [Lea82], Condition $D(u_n)$ and Corollary 2.3) there exists $\theta \in [0, 1]$, called the *extremal index*, such that

$$\mathbb{P}[a_n^{-1} M_n \leq t] \to \exp(-\theta t^{-\alpha}).$$

The candidate extremal index ϑ can be used to derive a similar limit theorem for the running maximum $a_T^{-1} \sup_{0\leq s\leq T} |X_s|$ of a stationary regularly varying

stochastic process $\mathbf{X}$. In the case that $\vartheta \in (0, \infty)$ and under an additional assumption (mainly the *anticlustering condition*, see [Sou22] Assumption 3.3, and [Sou22] Assumption 3.9) we have

$$\mathbb{P}\left[a_T^{-1} \sup_{0 \le s \le T} |X_s| \le t\right] \to \exp(-\vartheta t^{-\alpha}).$$

Here $\alpha > 0$ is the index of regular variation and $(a_T)_{T \ge 0}$ is chosen such that $T\mathbb{P}[|X_0| > a_T] \to 1$ as $T \to \infty$ (see [Sou22] Corollary 3.11).

Now we are ready to present the main theorem of [Sou22] regarding mixed moving average representations. See also [KS20] Theorem 5.4.9 and Theorem 5.4.10 for the discrete-time version. Denote by D_0 the set of all càdlàg functions vanishing at infinity and let D_α be the set of càdlàg functions y such that $\|y\|_\alpha^\alpha := \int_{\mathbb{R}} |y_t|^\alpha dt < \infty$.

Theorem 4.19. *Let ν be a shift-invariant tail measure with tail process $\boldsymbol{Y}$. The following are equivalent:*

1. *There exists a càdlàg process $\boldsymbol{Q}$ such that $\mathbb{P}[\boldsymbol{Q}^* = 1] = 1$ and ν has the mixed moving average representation (MMA)*
2. *The tail process almost-surely vanishes at infinity, $\mathbb{P}[\boldsymbol{Y} \in D_0] = 1$*
3. *$\mathbb{P}[\boldsymbol{Y} \in D_\alpha] = 1$*
4. *$\mathbb{P}[\mathcal{E}(\boldsymbol{Y}) < \infty] = 1$ and therefore $\vartheta > 0$*
5. *$\nu(D \setminus D_0) = 0$ meaning ν is supported on D_0*
6. *$\nu(D \setminus D_\alpha) = 0$ meaning ν is supported on D_α*

In this case ϑ of Definition 4.16 is given by the candidate extremal index $\vartheta = \mathbb{E}[\mathcal{E}(\boldsymbol{Y})^{-1}] > 0$ and the law of $\boldsymbol{Q}$ is given by

$$\mathbb{P}^{\boldsymbol{Q}} = \vartheta^{-1}\mathbb{E}\left[\frac{\delta_{\boldsymbol{Y}/\boldsymbol{Y}^*}}{\mathcal{E}(\boldsymbol{Y})}\right].$$

Proof. We will repeat the arguments given in [Sou22], Theorem 2.9 here:

By Lemma 4.17 we immediately have $2 \Leftrightarrow 5$ as well as $3 \Leftrightarrow 6$.

Assume that 1 holds. Then by ν being $\mathcal{B}_0$-boundedly finite we get

$$\begin{aligned}
\infty > \nu(\{y \in D : \sup_{a \le s \le b} |y_s| > 1\} \\
&= \vartheta \int_{\mathbb{R}} \int_0^\infty \mathbb{P}\Big[u \sup_{a \le s+t \le b} |Q_s| > 1\Big] \alpha u^{-(\alpha+1)} du dt \\
(x = u^{-\alpha}) &= \vartheta \int_{\mathbb{R}} \int_0^\infty \mathbb{P}\Big[\Big(\sup_{a \le s+t \le b} |Q_s|\Big)^\alpha > x\Big] dx dt \\
&= \vartheta \int_{\mathbb{R}} \mathbb{E}\Big[\Big(\sup_{a \le s+t \le b} |Q_s|\Big)^\alpha\Big] dt
\end{aligned}$$

which implies $\mathbf{Q} \in D_\alpha$ almost-surely. Assume that $\mathbb{P}[\mathbf{Q} \in D_0] < 1$, meaning that with positive probability $\mathbf{Q}$ does not go to 0. Then there must exist an $\epsilon > 0$ such that there are infinitely many intervals of length at least $b - a$ such that $\sup_{a \le s+t \le b} |Q_s|^\alpha > (b-a)\epsilon^\alpha$ and hence the above integral equals ∞, leading to a contradiction. Therefore we also have $\mathbf{Q} \in D_0$. By Lemma 4.17 it follows that $\mathbf{Y} \in D_0$ and $\mathbf{Y} \in D_\alpha$ almost-surely.
Now assume 3 holds and let g be a non-negative measurable map on D. Then

$$\begin{aligned}
\int_D g d\nu &= \int_D g(y) \frac{\|y\|_\alpha^\alpha}{\|y\|_\alpha^\alpha} \nu(dy) = \int_D \int_{\mathbb{R}} g(y) \frac{|y_t|^\alpha}{\|y\|_\alpha^\alpha} dt \nu(dy) \\
&= \int_D \int_{\mathbb{R}} \int_0^\infty g(y) \frac{\mathbb{1}\{|y_t| > u\}}{\|y\|_\alpha^\alpha} \alpha u^{\alpha-1} du dt \nu(dy) \\
(\text{shift-invariance}) &= \int_D \int_{\mathbb{R}} \int_0^\infty g(B^t y) \frac{\mathbb{1}\{|y_0| > u\}}{\|y\|_\alpha^\alpha} \alpha u^{\alpha-1} du dt \nu(dy) \\
(y \leftarrow y/u) &= \int_D \int_{\mathbb{R}} \int_0^\infty g(uB^t y) \frac{\mathbb{1}\{|y_0| > 1\}}{\|y\|_\alpha^\alpha} \alpha u^{-(\alpha+1)} du dt \nu(dy) \\
&= \int_0^\infty \int_{\mathbb{R}} \mathbb{E}\Big[\frac{g(uB^t \mathbf{Y})}{\|\mathbf{Y}\|_\alpha^\alpha}\Big] \alpha u^{-(\alpha+1)} dt du \\
&= \int_0^\infty \int_{\mathbb{R}} \mathbb{E}\Big[\frac{(\mathbf{Y}^*)^\alpha g\big(\frac{uB^t \mathbf{Y}}{\mathbf{Y}^*}\big)}{\|\mathbf{Y}\|_\alpha^\alpha}\Big] \alpha u^{-(\alpha+1)} dt du
\end{aligned}$$

This almost has the form of a mixed moving average representation. The only problem is that $\mathbb{E}[(\mathbf{Y}^*)^\alpha \delta(uB^t \mathbf{Y}/\mathbf{Y}^*)\|\mathbf{Y}\|_\alpha^{-\alpha}]$ does not define a probability measure, hence it cannot define the law of $\mathbf{Q}$. However, note that by choosing

$g(y) = \mathcal{E}(y)^{-1}\mathbb{1}\{|y_0| > 1\}$ we get

$$\begin{aligned}\vartheta &= \mathbb{E}[\mathcal{E}(\mathbf{Y})^{-1}] = \int_D g d\nu \\ &= \int_0^\infty \int_{\mathbb{R}} \mathbb{E}\left[\frac{(\mathbf{Y}^*)^\alpha \mathbb{1}\{u|Y_{-t}| > \mathbf{Y}^*\}}{\|\mathbf{Y}\|_\alpha^\alpha \mathcal{E}(u\mathbf{Y}/(\mathbf{Y}^*))}\right] \alpha u^{-(\alpha+1)} dt du \\ &= \int_0^\infty \mathbb{E}\left[\frac{(\mathbf{Y}^*)^\alpha}{\|\mathbf{Y}\|_\alpha^\alpha} \mathcal{E}(u\mathbf{Y}/(\mathbf{Y}^*))^{-1} \int_{\mathbb{R}} \mathbb{1}\{u|Y_{-t}| > \mathbf{Y}^*\} dt\right] \alpha u^{-(\alpha+1)} du \\ &= \int_1^\infty \mathbb{E}\left[\frac{(\mathbf{Y}^*)^\alpha}{\|\mathbf{Y}\|_\alpha^\alpha}\right] \alpha u^{-(\alpha+1)} du = \mathbb{E}\left[\frac{(\mathbf{Y}^*)^\alpha}{\|\mathbf{Y}\|_\alpha^\alpha}\right]\end{aligned}$$

with the convention that if $\mathcal{E}(y) = 0$ then $\mathcal{E}(y)/\mathcal{E}(y) = 0$ and therefore $\mathcal{E}(uy/y^*)/\mathcal{E}(uy/y^*) = 0$ for $u < 1$.
Since 3 holds we have $\|\mathbf{Y}\|_\alpha^\alpha < \infty$ and it follows that $\vartheta > 0$ and therefore

$$\mathbb{P}^{\mathbf{Q}} = \vartheta^{-1}\mathbb{E}\left[\frac{(\mathbf{Y}^*)^\alpha \delta_{\mathbf{Y}/\mathbf{Y}^*}}{\|\mathbf{Y}\|_\alpha^\alpha}\right] = \vartheta^{-1}\mathbb{E}\left[\frac{\delta_{\mathbf{Y}/\mathbf{Y}^*}}{\mathcal{E}(\mathbf{Y})}\right]$$

defines a probability measure on D and gives a process $\mathbf{Q}$ with law $\mathbb{P}^{\mathbf{Q}}$ and $\mathbf{Q}^* = 1$ almost-surely. By the previous computations, this process defines a mixed moving average representation for ν, hence 3 implies 1.
Together with Lemma 4.17 we have now proven that 1, 2, 3, 5 and 6 are equivalent.
Clearly $\mathbf{Y} \in D_0$ almost-surely implies $\mathcal{E}(\mathbf{Y}) < \infty$ almost-surely and hence 2 implies 4. For sake of brevity we refer to [Sou22], Theorem 2.9 for the proof that 4 implies 1. □

In the proof of Theorem 4.19 we have seen that

$$\nu\left(\left\{y \in D : \sup_{a \le s \le b} |y_s| > 1\right\}\right) = \vartheta \int_{\mathbb{R}} \mathbb{E}\left[\left(\sup_{a \le s+t \le b} |Q_s|\right)^\alpha\right] dt$$

hence in order for a process $\mathbf{Q}$ to define a shift-invariant tail measure ν by a mixed moving average representation, it is necessary for this quantity to be finite. Under this assumption, it is possible to construct a tail measure and a corresponding tail process by only knowing $\mathbf{Q}$, as shown in [Sou22], Corollary 2.11:

Corollary 4.20. *Let $\boldsymbol{Q}$ be a càdlàg stochastic process with $\mathbb{P}[\boldsymbol{Q}^* = 1] = 1$ and*

$$0 < \int_{\mathbb{R}} \mathbb{E}\left[\left(\sup_{a \le s+t \le b} |Q_s|\right)^\alpha\right] dt < \infty$$

for all $a < b$.
Then there exists a unique shift-invariant tail measure ν such that

$$\nu(A) = \vartheta \int_{\mathbb{R}} \int_0^\infty \mathbb{P}[uB^t \boldsymbol{Q} \in A] \alpha u^{-(\alpha+1)} du dt.$$

The tail measure is supported on D_0 and the tail process $\boldsymbol{Y}$ is given by

$$\boldsymbol{Y} = R\frac{B^T\boldsymbol{Q}}{|Q_{-T}|}$$

where $R \sim Par(\alpha)$ is independent of $\boldsymbol{Y}$ and T. The real-valued random variable T has a joint distribution with $\boldsymbol{Q}$ given by

$$\vartheta \int_{\mathbb{R}} \mathbb{E}[\delta_{\boldsymbol{Q},t}|Q_{-t}|^\alpha]dt$$

where $\vartheta = \left(\int_{\mathbb{R}} \mathbb{E}[|Q_t|^\alpha]dt\right)^{-1}$.

Example 4.7 continued
The claim in the introduction of this section was that for the tail process $\mathbf{Y} = Y_0 e^{-\theta t}\mathbb{1}\{t \geq -U\}$ the process $\mathbf{Q}$ is given by $Q_t = \Theta_0 e^{-\theta t}\mathbb{1}\{t \geq 0\}$.
Clearly $\mathbf{Q}$ satisfies the assumptions of Corollary 4.20 and T has a distribution given by

$$\begin{aligned}\mathbb{P}[T \leq x] &= \vartheta \int_{\mathbb{R}} \mathbb{E}[\delta_{\mathbf{Q},t}(D \times (-\infty, x])|Q_{-t}|^\alpha]dt \\ &= \vartheta \int_{-\infty}^{x\wedge 0} e^{\alpha\theta t}dt = \frac{\vartheta}{\alpha\theta}e^{\alpha\theta\cdot(x\wedge 0)}.\end{aligned}$$

Since $\vartheta = (\int_0^\infty e^{-\alpha\theta t}dt)^{-1} = \alpha\theta$ we can see that T follows an exponential distribution on the negative numbers with parameter $\alpha\theta$. Hence

$$\mathbf{Y} = R\frac{B^T(\Theta_0 e^{-\theta t}\mathbb{1}\{t \geq 0\})}{e^{-\theta T}} = Y_0 e^{-\theta t}\mathbb{1}\{t \geq T\}$$

as claimed.

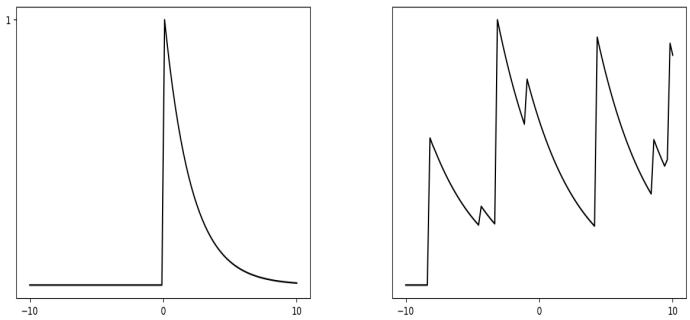
Just like Theorem 4.13, the mixed moving average representation can be used to construct a max-stable process $\mathbf{W}$ with tail measure ν. If ν satisfies the assumptions of Theorem 4.19 and N is a Poisson point process on D with mean measure ν, then we can represent N as

$$N = \sum_{i=1}^{\infty} \delta_{P_i B^{T_i}\mathbf{Q}^{(i)}}$$

where $(P_i, T_i, \mathbf{Q}^{(i)})$ is an enumeration of points of a Poisson point process on $(0,\infty)\times\mathbb{R}\times D$ with mean measure $\mu_\alpha \otimes \vartheta\lambda\lambda \otimes \mathbb{P}^{\mathbf{Q}}$, where $\vartheta, \mathbb{P}^{\mathbf{Q}}$ are as in Theorem 4.19 and μ_α has density $\alpha u^{-(\alpha+1)}du$.
Therefore we can construct a max-stable process $\mathbf{W}$ by

$$\mathbf{W} = \bigvee_{i=1}^{\infty} P_i B^{T_i}\mathbf{Q}^{(i)}$$

and $\mathbf{W}$ has tail measure ν.



(a) $\mathbf{Q}$ (b) $\mathbf{W}$

Figure 5: The processes $\mathbf{Q}$ and $\mathbf{W}$ for $|\mathbf{X}|$ of Example 4.7

The change of measure given in Theorem 4.19 might be hard to compute, hence we would like some alternative ways of determining $\mathbf{Q}$. In the introduction of this section we have already seen how one can interpret the process $\mathbf{Q}$ for Example 4.6 as $\mathbf{Y}/\mathbf{Y}^*$ conditioned on $\mathcal{I}(\mathbf{Y}) = 0$. We want to extend this construction of $\mathbf{Q}$. For this, we introduce a class of functionals on D, called *anchoring maps*.

Definition 4.21. Anchoring Maps ([Sou22], Section 2.3)
A map $\mathcal{I} : D \to [-\infty, \infty]$ called an anchoring map if it is measurable and for all $t \in \mathbb{R}$ we have

$$\mathcal{I}(B^t y) = \mathcal{I}(y) + t.$$

We already saw an example given by the map $\mathcal{I}$ which returns the first time $t \in \mathbb{R}$ the tail process $\mathbf{Y}$ reaches its maximum, meaning $\mathcal{I}(y) = \inf\{t \in \mathbb{R} : y^* \in \{|y_t|, |y_{t^-}|\}\}$. This map is called the *infargmax functional*. Another interesting example is given by the first exceedance over 1, meaning $\mathcal{I}(y) = \inf\{t \in \mathbb{R} : |y_t| > 1\}$.

To conclude this section, we state, without proof, a few other practical ways to compute the candidate extremal index ϑ:

Lemma 4.22. *Let ν be a shift-invariant tail measure with tail process $\boldsymbol{Y}$, spectral tail process Θ and candidate extremal index $\vartheta \neq 0$. Then we have:*

- $\vartheta = \lim_{T\to\infty} T^{-1}\nu\big(\{y \in D : \sup_{0\leq t\leq T} |y_t| > 1\}\big)$ *([Sou22], Lemma 2.5)*

- $\vartheta = \mathbb{E}\left[\frac{(\Theta^*)^\alpha}{\|\Theta\|_\alpha^\alpha}\right]$ *([Sou22], Corollary 2.10)*
- *If $\mathcal{I}$ is an anchoring map satisfying $\mathbb{P}[R|Q_{\mathcal{I}(R\boldsymbol{Q})}| > 1] = 1$ for $R \sim Par(\alpha)$ independent of $\boldsymbol{Q}$ and $\mathcal{I}(\boldsymbol{Y}) \in \mathbb{R}$ almost-surely, then ([Sou22], Equation 2.26)*

  $$\vartheta = \frac{1}{\mathbb{E}[\mathcal{E}(\boldsymbol{Y})|\mathcal{I}(\boldsymbol{Y}) = 0]}.$$

  Furthermore $\mathcal{I}(\boldsymbol{Y})$ has a Riemann density $f_{\mathcal{I}(\boldsymbol{Y})}$ which is left-continuous at 0 and we have $\vartheta = f_{\mathcal{I}(\boldsymbol{Y})}(0)$ ([Sou22], Equation 2.23).

Example 4.7 continued
Let $\mathcal{I}$ be the infargmax functional. For the tail process $\mathbf{Y} = Y_0 e^{-\theta t}\mathbb{1}\{t \geq -U\}$ we have $\mathcal{I}(\mathbf{Y}) = -U$. Therefore $\mathcal{I}(\mathbf{Y})$ follows an exponential distribution on the negative real numbers with parameter $\alpha\theta$. It follows that

$$\vartheta = f_{\mathcal{I}(\mathbf{Y})}(0) = \alpha\theta.$$

As one might expect, if θ is large, so is ϑ, since θ describes how fast the process $\mathbf{X}$ reverts back to 0 and hence how fast the tail process $\mathbf{Y}$ decays towards 0.

5 Regularly Varying Markov Processes

In this section we will consider an important class of stochastic processes in D, called Markov processes. Assuming that a Markov process is regularly varying in D, we show how to compute its forward and backward tail process by using the transition kernel (Proposition 5.6), semigroup (Theorem 5.10) or generator (Theorem 5.11) of the regularly varying Markov process. Notably, in many cases both the forward and backward tail process of a regularly varying Markov process will again be a Markov process.

5.1 Definitions

Definition 5.1. Markov Process ([Kal21], Chapter 11, Equation (1))
A stochastic process $(X_t)_{t\geq 0}$ is called a Markov process on $\mathbb{R}^d$ with respect to a filtration $(\mathcal{F}_t)_{t\geq 0}$ if for all $s, t \geq 0$ we have

$$\mathbb{E}\big[f(X_{t+s})|\mathcal{F}_t\big] = \mathbb{E}\big[f(X_{t+s})|X_t\big]$$

A stochastic process $(X_n)_{n\in\mathbb{N}_0}$ is called a Markov chain on $\mathbb{R}^d$ with respect to a filtration $(\mathcal{F}_n)_{n\in\mathbb{N}_0}$ if for all $n \in \mathbb{N}$ we have

$$\mathbb{E}\big[f(X_{n+1})|\mathcal{F}_n\big] = \mathbb{E}\big[f(X_{n+1})|X_n\big].$$

Definition 5.2. Transition Kernel ([Kal21], Chapter 11, Equation (3))
Let $(X_t)_{t\geq 0}$ be a Markov process. We call the family $(P_{s,t})_{t\geq s\geq 0}$ of kernels defined by

$$P_{s,t}(x, A) = \mathbb{P}[X_t \in A|X_s = x]$$

the transition kernels of $\mathbf{X}$. We say $\mathbf{X}$ is time-homogeneous if there exist transition kernels $(P_t)_{t\geq 0}$ such that $P_{s,t} = P_{t-s}$ for all $s \leq t$.
Analogously, for a Markov chain $(X_n)_{n\in\mathbb{N}}$ we call

$$\Pi(x, A) = \mathbb{P}[X_1 \in A|X_0 = x]$$

the transition kernel of $\mathbf{X}$.

Transition kernels allow us to more easily describe the joint distribution of $(X_{t_1}, \ldots, X_{t_k})$: For a distribution μ on $\mathbb{R}^d$ and a transition kernel K define the measure $\mu \otimes K$ on $(\mathbb{R}^d)^2$ by

$$\mu \otimes K(h) = \int_{(\mathbb{R}^d)^2} h(x, y)K(x, dy)\mu(dx)$$

for each bounded continuous function h on $(\mathbb{R}^d)^2$.
Similarly, for two transition kernels K and K' define the transition kernel

$$K \otimes K'(x, A \times B) = \int_A K'(y, B)K(x, dy)$$

Now we have for a Markov process $\mathbf{X}$ with transition kernels $P_{s,t}$ (see [Kal21] Proposition 11.2)

$$\mathcal{L}(X_0, X_{t_1}, \ldots, X_{t_k}) = \mu \otimes P_{t_1} \otimes P_{t_1,t_2} \otimes \cdots \otimes P_{t_{k-1},t_k}$$

where μ is the initial distribution of X_0.
In particular, we call μ a stationary distribution of a time-homogeneous Markov process $\mathbf{X}$ if $\mu(B) = \int_{\mathbb{R}^d} P_t(x, B)\mu(dx)$ for all $t \geq 0$. By [Kal21] Lemma 11.7 this is equivalent to the Markov process $\mathbf{X}$ having transition kernels P_t and starting with initial distribution $\mathcal{L}(X_0) = \mu$ to be stationary.

Definition 5.3. Feller Process ([Kal21] page 369)
Define the function space

$$C^0 := \{f : \mathbb{R}^d \to \mathbb{R}, f \text{ is continuous and vanishes at infinity}\}$$

where vanishing at infinity means for all $\epsilon > 0$ there exists an $R > 0$ such that $|f(x)| \leq \epsilon$ for all $|x| > R$.
Let $\mathbf{X}$ be a Markov process with transition kernels $(P_t)_{t\geq 0}$. Define the operators $(T_t)_{t\geq 0}$ on C^0 by

$$T_t f(x) := \mathbb{E}\big[f(X_t)|X_0 = x\big] = \int_{\mathbb{R}^d} f(y)P_t(x, dy)$$

If for each $f \in C^0$ we have $T_t f \in C^0$ and $T_t f \to f$ uniformly as $t \to 0$, we call $(T_t)_{t\geq 0}$ a Feller semigroup and $\mathbf{X}$ a Feller process.

Functions in C^0 can be used to approximate more general measurable functions, which allows us to recover the transition kernels P_t from the operators T_t.

Definition 5.4. Generator ([EK09], Chapter 1)
Let $(T_t)_{t\geq 0}$ be a Feller semigroup. We define the operator

$$\mathcal{A}f(x) = \lim_{t\searrow 0} \frac{1}{t}\big[T_t f(x) - f(x)\big]$$

for each function $f \in C^0$ for which the limit exists. We call $\mathcal{A}$ the generator of $\mathbf{X}$ and denote by $\mathcal{D}(\mathcal{A})$ the domain of $\mathcal{A}$ consisting of functions for which the above limit exists.

5.2 Transition Function

Our first approach is to find conditions on the transition kernels P_t which are sufficient for regular variation of the associated Markov process. For Markov chains, this has been investigated in [RZ13] by studying the asymptotic behaviour of their transition kernels and in [JS14] by studying the asymptotic

behaviour of their functional representations. In this section we will adopt the former approach, as it generalizes more easily to Markov processes in continuous time. See [KS20] Section 14.2 for an overview of the main results of both approaches.
The basic idea is that since for each u the process $(X_t)_{t\geq 0}/u$ conditioned on $|X_0| > u$ is a Markov process, we might think that the limit $(Y_t)_{t\geq 0}$ again is a Markov process. In particular, we would hope that the transition kernels $(P_t)_{t\geq 0}$ of $\mathbf{X}$ and the transition kernels $(Q_t)_{t\geq 0}$ of $\mathbf{Y}$ are related by

$$P_t(ux, uA) = \mathbb{P}[X_t \in uA \big| X_0 = ux] \xrightarrow[u\to\infty]{} \mathbb{P}[Y_t \in A \big| Y_0 = x] = Q_t(x, A).$$

This will be the working assumption throughout this chapter. We give one of the main theorems regarding regular variation of Markov chains here:

Theorem 5.5. *([KS20] Theorem 14.2.3)*
Let $\boldsymbol{M}$ be a Markov chain with transition kernel Π. Suppose there exists a transition kernel K such that

$$\Pi(ux, uA) \to K(x, A), \qquad u \to \infty, x \neq 0, K(x, \partial A) = 0$$

locally uniformly with respect to x, meaning for every compact $K \subset \mathbb{R}^d \setminus \{0\}$ we have $\sup_{x\in K} |\Pi(ux, uA) - K(x, A)| \to 0$.
Supppose that the initial distribution $\mathcal{L}(M_0)$ is regularly varying with spectral measure Λ_0.
Further assume one of the following conditions:

$$K(u, \{0\}) = 0 \qquad \forall u \neq 0$$
$$\lim_{\epsilon\to 0} \limsup_{x\to\infty} \sup_{|u|\leq\epsilon} \Pi\big(xu, B_x(0)^C\big) = 0$$

Extend K by setting $K(0, A) = \mathbb{1}_A(0)$. Then $\boldsymbol{M}$ is regularly varying in $(\mathbb{R}^d)^{\mathbb{N}_0}$. If furthermore $\boldsymbol{M}$ is stationary, its forward spectral tail process $(\Theta_n)_{n\in\mathbb{N}_0}$ has transition kernel K and initial distribution Λ_0.

The proof is given in [KS20] Theorem 14.2.3, though we will repeat their arguments here in the case of a continuous-time Markov process $\mathbf{X}$. Replace the transition kernel Π with the kernels $(P_t)_{t\geq 0}$ and K with kernels $(Q_t)_{t\geq 0}$ and we can formulate the exact assumptions for Markov processes $\mathbf{X} = (X_t)_{t\geq 0}$. However, in continuous time the arguments given in [KS20] Section 14.2 only suffice to prove *finite-dimensional* regular variation of $\mathbf{X}$ and *finite-dimensional* weak convergence to a process Θ.

Proposition 5.6. *Suppose $\boldsymbol{X} = (X_t)_{t\geq 0}$ is a Markov process with transition kernels $(P_t)_{t\geq 0}$. Assume each P_t satisfies the assumptions given in Theorem*

5.5 for a transition kernel Q_t. Then $\boldsymbol{X}$ is finite-dimensional regularly varying. The kernels Q_t can be extended by setting $Q_t(0, A) = \mathbb{1}_A(0)$. Then if $\boldsymbol{X}$ is stationary and $\boldsymbol{Y}$ is a Markov process with transition kernels $(Q_t)_{t\geq 0}$, the finite-dimensional projections of $\boldsymbol{X}/u$ converge weakly to the finite-dimensional projections of $\boldsymbol{Y}$ conditionally on $|X_0| > u$.

Remark:
Note that such a transition function Q_t necessarily has the homegeneity property $Q_t(sx, sA) = Q_t(x, A)$.

For the proof of Proposition 5.6, we need the following Lemma:

Lemma 5.7. *(see [KS20], Theorem A.2.6)*
Let E be a metric space endowed with its Borel-σ-field. Let $\{\mu_n, n \in \mathbb{N}\}$ be a sequence of probability measures converging weakly to a probability measure μ. Let $F \subset E$ be a Borel set with $\mu(F) = 1$. If $\{f_n, n \in \mathbb{N}\}$ is a sequence of functions which converges locally uniformly on F to a function f, then $\lim_{n\to\infty} \int f_n d\mu_n = \int f d\mu$.

Now we are ready to give the proof of Proposition 5.6. Note that, as already mentioned, the same arguments are used to prove Theorem 5.5, switching P_t with Π.

Proof. The proof is mostly the same as in [KS20], Section 14.2.2.
Let $h : B^C(0,1) \times (\mathbb{R}^d)^k \to \mathbb{R}$ be a bounded and continuous function. Pick any $t_1, ..., t_k \in \mathbb{R}$. For $x > 0$ define

$$\lambda_x(A) = \frac{\lambda(xA)}{\lambda(B^C(0,x))}$$

where λ is the stationary distribution of X. By regular variation of λ we know that λ_x converges vaguely to $\mu_\alpha \otimes \Lambda_0$ where $\mu_\alpha(du) = \alpha u^{-(\alpha+1)}du$ and Λ_0 is the spectral measure of X_0. Let $\nu(\cdot) = \mu_\alpha \otimes \Lambda_0(\cdot \cap B^C(0,1))$. Because of [Sou22], Proposition A.1, finite-dimensional regular variation is now equivalent to

$$\lambda_x \otimes P_x^{(t_1,...,t_k)} \overset{w}{\Rightarrow} \nu \otimes Q^{(t_1,...,t_k)}$$

on $B^C(0,1) \times (\mathbb{R}^d)^k$, where $P_t^x(u, A) = P_t(xu, xA)$ and $P_x^{(t_1,...,t_k)}$ analogously. That means for the chosen bounded continuous function h that

$$\begin{aligned}\lim_{x\to\infty} &\int_{B^C(0,1)\times(\mathbb{R}^d)^k} h(x_0, ..., x_k) P^x_{t_k - t_{k-1}}(x_{k-1}, dx_k) \dots P^x_{t_1}(x_0, dx_1)\lambda_x(dx_0) \\ &= \int_{B^C(0,1)\times(\mathbb{R}^d)^k} h(x_0, ..., x_k) Q_{t_k - t_{k-1}}(x_{k-1}, dx_k) \dots Q_{t_1}(x_0, dx_1)\nu(dx_0)\end{aligned}$$

This will be proven by induction over k. The base case $k = 0$ follows by regular variation of λ.
Now define the functions

$$f_x(x_0, ..., x_{k-1}) = \int_{\mathbb{R}^d} h(x_0, ..., x_{k-1}, x_k) P^x_{t_k - t_{k-1}}(x_{k-1}, dx_k)$$
$$f(x_0, ..., x_{k-1}) = \int_{\mathbb{R}^d} h(x_0, ..., x_{k-1}, x_k) Q_{t_k - t_{k-1}}(x_{k-1}, dx_k)$$

The induction assumption states that $\lambda_x \otimes P_x^{(t_1,...,t_{k-1})}$ converges weakly to $\nu \otimes Q^{(t_1,...,t_{k-1})}$, so we simply need to prove that f_x converges locally uniformly to f on a subset F of $B^C(0,1) \times (\mathbb{R}^d)^{k-1}$ with $\nu \otimes Q^{(t_1,...,t_{k-1})}(F) = 1$ and the induction step follows by Lemma 3 .
Since $P^x_t \to Q_t$ locally uniformly, $\int_{\mathbb{R}^d} h(x_0, ..., x_k) P^x_{t_k - t_{k-1}}(x_{k-1}, dx_k)$ converges locally uniformly (w.r.t. x_{k-1}) on $\mathbb{R}^d \backslash \{0\}$ to $\int_{\mathbb{R}^d} h(x_0, ..., x_k) Q_{t_k - t_{k-1}}(x_{k-1}, dx_k)$.
Case 1: If we assume $Q_t(u, \{0\}) = 0$ for all $u \neq 0$ then for $F = B^C(0,1) \times (\mathbb{R}^d \setminus \{0\})^{k-1}$ it follows that f_x converges locally uniformly to f on $B^C(0,1) \times (\mathbb{R}^d \setminus \{0\})^{k-1}$.
Case 2: If we assume the latter condition, then if u_x is a sequence converging to 0 and g is a bounded continuous function on $\mathbb{R}^d$ with $g(0) = 0$. Then for every $\epsilon > 0$, η small enough and x large enough,

$$\begin{aligned} |P^x_t g(u_x)| &\leq \int_{B_\eta(0)} |g(v)| P_t(xu_x, xdv) + \int_{B_\eta(0)^C} |g(v)| P_t(xu_x, xdv) \\ &\leq \sup_{|v| \leq \eta} |g(v) + \|g\|_\infty \sup_{|u| \leq \epsilon x} P_t(u, B_{\eta x}(0)^C) \\ &\leq \epsilon + \|g\|_\infty P_t(u, B_{\eta x}(0)^C) \end{aligned}$$

Therefore

$$\limsup_{x \to \infty} |P^x_t g(u_x)| \leq \epsilon + \|g\|_\infty \limsup_{x \to \infty} \sup_{|u| \leq \epsilon x} P_t(u, B_{\eta x}(0)^C)$$

which converges to 0 as $\epsilon \to 0$. Therefore

$$\lim_{x \to \infty} P^x_t g(u_x) = 0 = g(0) = Q_t h(0)$$

since $Q_t(0, \{0\}) = 1$.
This proves that $P^x_t g$ converges locally uniformly to $Q_t h$, and hence so does f_x. □

Example 5.8. Define a Markov chain $\mathbf{M}$ by

$$M_{j+1} = \big(M_j + Z_{j+1}\big) \mathbb{1}_{\{V_{j+1} > p\}}$$

for a sequence of i.i.d. uniform random variables $(V_j)_{j\in\mathbb{N}}$, a sequence of i.i.d. positive random variables $(Z_j)_{j\in\mathbb{N}}$ and a constant $p \in (0,1)$. Define the intensity function $\lambda(x) = x \vee 1$ and a sequence of random variables τ_j with distribution

$$\tau_j \stackrel{d}{=} \frac{E_j}{\lambda(M_{j-1})}$$

where $(E_j)_{j\in\mathbb{N}}$ is a sequence of i.i.d. Exp(1) random variables. Define the Poisson process $N_t = \max\{n \in \mathbb{N} : \sum_{i=1}^{n} \tau_i \leq t\}$ and the Markov process $X_t = M_{N_t}$.
Suppose a stationary and regularly varying distribution of $\mathbf{M}$ (and therefore of $\mathbf{X}$) exists. The transition kernel of $\mathbf{M}$ is given by

$$\Pi(x, A) = p\mathbb{1}_A(0) + (1-p)\mathbb{P}[x + Z_1 \in A]$$

which satisfies the assumptions of Theorem 5.5. Therefore $\mathbf{M}$ is regularly varying and $\mathbf{X}$ is finite-dimensional regularly varying. However, $\mathbf{X}$ is not regularly varying in D since as we condition on $|X_0| > u$ the parameters $\lambda(X_0)$ become larger and therefore the waiting time gets shorter, eventually going to 0. Now with probability p, the process $\mathbf{X}$ drops down to 0 and thus any process satisfying $\mathcal{L}(\mathbf{X}/u \big| |X_0| > u) \to \mathcal{L}(\mathbf{Y})$ is with probability p not càdlàg in $t = 0$. Therefore $\mathbf{X}$ cannot be regularly varying in D.

The backwards tail process

So far we have only shown that the forward tail process $(Y_t)_{t\geq 0}$ is a Markov process with transition function Q_t. It remains to determine what the backwards tail process $(Y_{-t})_{t\geq 0}$ looks like. Using the time-change formula (Theorem 4.9) we can simply derive the distribution of the backwards tail process only knowing the transition function Q_t. Notably, the backwards tail process will again be a Markov process which is independent of the forward tail process, once conditioned on Y_0.

Proposition 5.9. *Let $\boldsymbol{X}$ be a Markov process satisfying the conditions of Proposition 5.6. Let Q_t be the transition function of the forward tail process of $\boldsymbol{X}$. Then the backwards tail process of $\boldsymbol{X}$ is a Markov process with transition function $\widehat{Q_t}$ defined by*

$$\widehat{Q_t}(u, \cdot) = \mathbb{P}[\Theta_{-t} = 0 \big| \Theta_0 = u]\delta_0 + \widetilde{Q}_t(u, \cdot)$$

where $\widetilde{Q}_t$ is defined as

$$\Lambda_0 \otimes \widetilde{Q_t}(H) = \int_{\mathbb{S}^{d-1}} \int_{\mathbb{R}^d \setminus \{0\}} H\left(\frac{u_1}{|u_1|}, \frac{u_0}{|u_1|}\right) |u_1|^{\alpha} Q_t(u_0, du_1)\Lambda_0(du_0).$$

Proof. The distribution of (Θ_0, Θ_{-t}) can be computed by applying the Time-change formula to (Θ_t, Θ_0) as follows: Let H be a bounded measurable function on $\mathbb{S}^{d-1} \times \mathbb{R}^d$ and let Λ_0 be the distribution of Θ_0. Applying the Time-change formula gives

$$\begin{aligned}\mathbb{E}[H(\Theta_0, \Theta_{-t})\mathbb{1}\{\Theta_{-t} \neq 0\}] &= \mathbb{E}\left[H\left(\frac{\Theta_t}{|\Theta_t|}, \frac{\Theta_0}{|\Theta_t|}\right)|\Theta_t|^\alpha \mathbb{1}\{\Theta_t \neq 0\}\right] \\ &= \int_{\mathbb{S}^{d-1}} \int_{\mathbb{R}^d \setminus \{0\}} H\left(\frac{v}{|v|}, \frac{u}{|v|}\right)|v|^\alpha Q_t(u, dv)\Lambda_0(du)\end{aligned}$$

Thus we define a kernel $\widetilde{Q}_t$ on $\mathbb{S}^{d-1} \times \mathbb{B}(\mathbb{S}^{d-1} \times \mathbb{R}^d \setminus \{0\})$ by

$$\widetilde{Q}_t H(u) = \int_{\mathbb{R}^d} H\left(\frac{v}{|v|}, \frac{u}{|v|}\right)|v|^\alpha Q_t(u, dv), \qquad u \in \mathbb{S}^{d-1}$$

and therefore by definition of $\widetilde{Q}_t$ we have

$$\int_{\mathbb{S}^{d-1}} \widetilde{Q}_t H(u)\Lambda_0(du) = \int_{\mathbb{S}^{d-1}} \int_{\mathbb{R}^d} H\left(\frac{v}{|v|}, \frac{u}{|v|}\right)|v|^\alpha Q_t(u, dv)\Lambda_0(du).$$

Defining the function $\widetilde{H}(u, v) = H(v/|v|, u/|v|) \cdot |v|^\alpha$ for $(u, v) \in \mathbb{S}^{d-1} \times \mathbb{R}^d \setminus \{0\}$ we can rewrite the above expression as $\Lambda_0 \otimes \widetilde{Q}_t(H) = \Lambda_0 \otimes Q_t(\widetilde{H})$.
Combining the equations above gives

$$\mathbb{E}[H(\Theta_0, \Theta_{-t})\mathbb{1}\{\Theta_{-t} \neq 0\}] = \int_{\mathbb{S}^{d-1}} \int_{\mathbb{R}^d} H(u, v)\widetilde{Q}_t(u, dv)$$

The kernel $\widetilde{Q}_t$ then needs to be extended to $\mathbb{R}^d \setminus \{0\} \times \mathbb{B}(\mathbb{S}^{d-1} \times \mathbb{R}^d \setminus \{0\})$. For any function $H : \mathbb{R}^d \setminus \{0\} \times \mathbb{R}^d \setminus \{0\}$ we define

$$\widetilde{Q}_t H(u) := \int_{\mathbb{R}^d} H\left(\frac{v}{|v|}, \frac{u}{|v|}\right)\widetilde{Q}_t\left(\frac{u}{|u|}, \frac{dv}{|u|}\right)$$

from which the homogeneity property $\widetilde{Q}_t(su, sdv) = \widetilde{Q}_t(u, dv)$ follows.

Now for any $H : \mathbb{S}^{d-1} \times (\mathbb{R}^d \setminus \{0\})^n$ define

$$\widetilde{H}(u_0, \ldots, u_n) := H\left(\frac{u_n}{|u_n|}, \ldots, \frac{u_0}{|u_n|}\right)|u_n|^\alpha$$

then we have for $0 < t_1 \leq \cdots \leq t_k$ (see [KS20] Lemma 14.2.9):

$$\Lambda_0 \otimes \widetilde{Q}_{t_1} \otimes \cdots \otimes \widetilde{Q}_{t_k - t_{k-1}}(H) = \Lambda_0 \otimes Q_{t_1} \otimes \cdots \otimes Q_{t_k - t_{k-1}}(\widetilde{H}).$$

Specifically, by the time change formula we obtain

$$\mathbb{E}[H(\Theta_0, \Theta_{-t_1}, \ldots, \Theta_{-t_k})\mathbb{1}\{\Theta_{-t_k} \neq 0\}] = \Lambda_0 \otimes \widetilde{Q}_{t_1} \otimes \cdots \otimes \widetilde{Q}_{t_k - t_{k-1}}(H).$$

Now if $\mathbb{E}|\Theta_t|^\alpha = \mathbb{P}[\Theta_{-t} \neq 0] = 1$ the $(\widetilde{Q}_t)_{t\geq 0}$ define valid transition kernels and we are done.
If not, meaning $0 < \mathbb{P}[\Theta_{-t} = 0] < 1$ we define the kernels

$$\widehat{Q_t}(u,\cdot) = \mathbb{P}[\Theta_{-t} = 0|\Theta_0 = u]\delta_0 + \widetilde{Q}_t(u,\cdot),$$

which are then the transition kernels of $(\Theta_{-t})_{t\geq 0}$. □

Example 4.7 continued
Suppose we only knew that the forward tail process of Example 4.6 was given by $Y_t = e^{-\theta t}Y_0$. Using Proposition 5.9 we can compute the backward transition kernel. Note that $\mathbb{P}[\Theta_{-t} = 0] = 1 - \mathbb{E}|\Theta_t|^\alpha = 1 - \exp(-\alpha\theta t)$. We first compute $\widetilde{Q}_t$ by the relation

$$\Lambda_0 \otimes Q_t(\widetilde{H}) = \Lambda_0 \otimes \widetilde{Q}_t(H)$$

which gives

$$\begin{aligned}
\Lambda_0 \otimes Q_t(\widetilde{H}) &= \int\int H\left(\frac{u_1}{|u_1|}, \frac{u_0}{|u_1|}\right) Q_t(u_0, du_1)\Lambda_0(du_0) \\
&= \int H\left(\frac{\exp(-\theta t)u_0}{\exp(-\theta t)}, \frac{u_0}{\exp(-\theta t)}\right) e^{-\alpha\theta t}\Lambda_0(du_0) \\
&= \int H\big(u_0, e^{\theta t}u_0\big)e^{-\alpha\theta t}\Lambda_0(du_0).
\end{aligned}$$

This implies that the backward tail process has transition kernel

$$\begin{aligned}
\widehat{Q_t}(u, \{e^{\theta t}u\}) &= e^{-\alpha\theta t} \\
\widehat{Q_t}(u, \{0\}) &= 1 - e^{-\alpha\theta t}
\end{aligned}$$

which shows that the backward tail process is given by

$$Y_{-t} = Y_0 e^{\theta t}\mathbb{1}\{t \geq -U\}$$

with $U \sim \text{Exp}(\alpha\theta)$ independent of Y_0.

5.3 Feller Processes

The techniques used in Section 5.2 did not suffice to prove regular variation in D. Proposition 5.5 has shown finite-dimensional weak convergence to a tail process $\mathbf{Y}$. The main benefit that Feller processes have to offer, as opposed to the more general class of Markov processes, is that weak convergence of Feller processes in D can be connected to convergence of their semigroups or generators.
Consider the setup of Proposition 5.6 where our Markov process has transition function P_t satisfying $P_t(ux, uA) \to Q_t(x, A)$ locally uniformly in x. This means

that for every x the measure $P_t^u(x,\cdot) := P_t(ux, u\cdot)$ converges weakly to the measure $Q_t(x,\cdot)$ and therefore for $f^u(z) = f(z/u)$ we get

$$T_t f^u(ux) = \mathbb{E}[f(X_t/u)|X_0 = ux] = \int_{\mathbb{R}^d} f(y) P_t^u(x, dy)$$
$$\to \int_{\mathbb{R}^d} f(y) Q_t(x, dy) =: S_t f(x)$$

Thus for each f the functions $T_t f^u(ux)$ converge locally uniformly in $x \in \mathbb{R}^d \setminus \{0\}$. This suggests that such a convergence, perhaps together with additional assumptions, might be sufficient to prove regular variation of a Feller process $\mathbf{X}$ based on its semigroup $(T_t)_{t\geq 0}$. It would seem that its forward tail process $\mathbf{Y}$ is again a Feller process with semigroup $(S_t)_{t\geq 0}$.

Theorem 5.10. *Let $\boldsymbol{X} = (X_t)_{t\geq 0}$ be a Feller process with semigroup $(T_t)_{t\geq 0}$. Let its transition kernels $(P_t)_{t\geq 0}$ satisfy the assumptions of Proposition 5.6 with $Q_t(x, \{0\}) = 0$ for non-zero x.*
Then $\mathcal{L}\big((X_t/u)_{t\geq 0}\big||X_0| > u\big)$ converges weakly to $\mathcal{L}\big((Y_t)_{t\geq 0}\big)$ in $D([0,\infty), \mathbb{R}^d)$, where $\boldsymbol{Y}$ is a Markov process with initial distribution

$$\mathbb{P}[Y_0 \in A] = \lim_{u\to\infty} \mathbb{P}[X_0 \in A\big||X_0| > u]$$

and transition kernels $(Q_t)_{t\geq 0}$ and semigroup $(S_t)_{t\geq 0}$.

Proof. Let $(u_n)_{n\in\mathbb{N}}$ be an arbitrary non-negative sequence going to infinity. Define a sequence of random variables $\mathbf{X}^{(n)}$ with law $\mathcal{L}\big(\mathbf{X}/u_n\big||X_0| > u_n\big)$. Since X_0 is regularly varying, $X_0^{(n)}$ converges weakly to Y_0.
In Proposition 5.6 we have already shown that the finite-dimensional projections of $\mathbf{X}^{(n)}$ converge weakly to the finite-dimensional projections of $\mathbf{Y}$.
To further show weak convergence in $D([0,\infty), \mathbb{R}^d)$, by Aldous' tightness criterion (see [Kal21], Theorem 23.9 and Theorem 23.11) it suffices to show that for any bounded sequence of optional stopping times $(\tau_n)_{n\in\mathbb{N}}$ and any sequence $0 \leq h_n \to 0$ we have

$$d\big(X_{\tau_n}^{(n)}, X_{\tau_n+h_n}^{(n)}\big) \to 0, \qquad n \to \infty$$

in probability. Since $\mathbf{X}$ satisfies the strong Markov property and is time-homogeneous, this is equivalent to

$$d\big(X_0^{(n)}, X_{h_n}^{(n)}\big) \to 0.$$

Now pick any pair of bounded continuous functions f, g, then since $T_{h_m}^n \to \mathrm{id}$ strongly as $m \to \infty$ and $T_{h_m}^n g \to S_{h_m} g$ locally uniformly as $n \to \infty$, we have

that $T^n_{h_n} g \to g$ locally uniformly as $n \to \infty$. Then by Lemma 5.7 we obtain

$$\begin{aligned}\mathbb{E}\big[f\big(X_0^{(n)}\big) \cdot g\big(X_{h_n}^{(n)}\big)\big] &= \mathbb{E}\big[f\big(X_0^{(n)}\big) \cdot T^n_{h_n} g\big(X_0^{(n)}\big)\big] \\ &= \int_{\mathbb{R}^d} f(x_0) \cdot T^n_{h_n} g(x_0) \mathbb{P}^{X_0^{(n)}}(dx_0) \\ &\xrightarrow[n\to\infty]{} \int_{\mathbb{R}^d} f(x_0) \cdot g(x_0) \mathbb{P}^{Y_0}(dx_0) = \mathbb{E}\big[f(Y_0) \cdot g(Y_0)\big]\end{aligned}$$

which implies $(X_0^{(n)}, X_{h_n}^{(n)}) \Rightarrow (Y_0, Y_0)$ and by the continuous mapping theorem $d(X_0^{(n)}, X_{h_n}^{(n)}) \Rightarrow d(Y_0, Y_0) = 0$. Now since $\mathbf{X}$ is finite-dimensional regularly varying and hence all finite-dimensional projections of $\mathbf{X}^{(n)}$ converge weakly to the finite-dimensional projections of $\mathbf{Y}$, it follows that $\mathbf{X}^{(n)} \Rightarrow \mathbf{Y}$ in D. □

We have proven that $\mathcal{L}((X_t)_{t\geq 0}/u \big| |X_0| > u)$ converges weakly in $D([0,\infty), \mathbb{R}^d)$, which certainly is a necessary condition for $\mathbf{X}$ to be regularly varying in D. Using Proposition 5.9 we can construct a backwards process $(Y_{-t})_{t\geq 0}$ and thus extend $\mathbf{Y}$ to a process in $D(\mathbb{R}, \mathbb{R}^d)$. It would then remain to be proven that $\mathbf{X}/u \big| |X_0| > u$ converges weakly to $\mathbf{Y}$. If $\mathbf{X}$ is stationary and regularly varying in D, then this process $\mathbf{Y}$ must be equal to the tail process. However, weak convergence to a tail process alone does not suffice to verify regular variation in D. Hence Theorem 5.10 only allows us to compute the tail process $\mathbf{Y}$ *assuming* $\mathbf{X}$ is regularly varying, but it does not say anything about whether $\mathbf{X}$ is regularly varying.

Example 4.7 continued
The process $\mathbf{X}$ defined as the solution of

$$dX_t = -\theta X_t dt + \sigma dL_t$$

for a symmetric α-stable Lévy process L has a solution which can be represented by

$$X_t = X_0 e^{-\theta t} + \sigma \int_0^t e^{-\theta(t-s)} dL_s,$$

see [RAS19] Proposition 2.9. It is a Feller process, see [Küh18] Example 4.3, and its transition kernel is given by

$$P_t(x, A) = \mathbb{P}[e^{-\theta t} x + \epsilon_t \in A]$$

for a random variable ϵ_t. This transition function satisfies the assumptions of Proposition 5.6 with $Q_t(x, \{e^{-\theta t} x\}) = 1$. Furthermore, since $\mathbf{X}$ is a Feller process, Theorem 5.10 now implies that the forward tail process of $\mathbf{X}$ is given by $Y_t = Y_0 e^{-\theta t}$.

The Generator

Theorem 5.10 was formulated using the transition kernels P_t or their corresponding semigroups T_t. These can be hard to determine and often times it is easier to describe a process by its generator $\mathcal{A}$. If we assume $\mathcal{A}f^u(ux) \to \mathcal{G}f(x)$ locally uniformly for a generator $\mathcal{G}$, can we derive a Theorem similar to Theorem 5.10 just using the generator?

Theorem 5.11. *Let $\boldsymbol{X}$ be a Feller process with bounded generator $\mathcal{A}$, meaning there exists a $C > 0$ such that for all $f \in C^0$ we have $\|\mathcal{A}f\|_\infty \leq C \cdot \|f\|_\infty$. Suppose that there exists a generator $\mathcal{G}$ such that $\mathcal{A}f^u(ux) \to \mathcal{G}f(x)$ locally uniformly in $\mathbb{R}^d \setminus \{0\}$ and $\mathcal{G}\mathbb{1}_{\{0\}}(x) = 0$.*
Then $(X_t)_{t\geq 0}/u$ conditioned on $|X_0| > u$ converges weakly to $(Y_t)_{t\geq 0}$ where $\boldsymbol{Y}$ has generator $\mathcal{G}$.

Proof. We will prove this Theorem by showing that if the assumptions of Theorem 5.11 are satisfied, the assumptions of 5.10 are also satisfied.
If a generator $\mathcal{A}$ is bounded, we can retrieve the corresponding semigroup T_t by

$$T_t f(x) := \exp(t\mathcal{A})f(x) := \sum_{k=0}^{\infty} \mathcal{A}^k f(x)\frac{t^k}{k!}$$

(see [Kal21] page 367). Therefore we let $T_t = \exp(t\mathcal{A})$ as well as $S_t = \exp(t\mathcal{G})$.
If $\mathcal{A}f^u(u\cdot) \to \mathcal{G}f(\cdot)$ locally uniformly, meaning for all sequence $x_n \to x$ we have $\mathcal{A}f^n(nx_n) \to \mathcal{G}f(x)$, the semigroups also converge locally uniformly:

$$T_t f^n(nx_n) = \exp(t\mathcal{A})f^n(nx_n) = \sum_{k=0}^{\infty} \mathcal{A}^k f^n(nx_n)\frac{t^k}{k!} \to \sum_{k=0}^{\infty} \mathcal{G}^k f(x)\frac{t^k}{k!} = S_t f(x)$$

by dominated convergence, since $\mathcal{A}$ and $\mathcal{G}$ are bounded.
If furthermore $\mathcal{G}\mathbb{1}_{\{0\}}(x) = 0$ for all non-zero x it follows that $\mathcal{G}^k\mathbb{1}_{\{0\}} = 0$ for all k and we get

$$S_t\mathbb{1}_{\{0\}}(x) = \sum_{k=0}^{\infty} \mathcal{G}^k\mathbb{1}_{\{0\}}(x)\frac{t^k}{k!} = 0$$

and therefore T_t and S_t satisfy the assumptions of Theorem 5.10. □

The assumption $\mathcal{G}\mathbb{1}_{\{0\}}(x) = 0$ requires that we can extend the domain of $\mathcal{G}$ such that $\mathcal{G}\mathbb{1}_{\{0\}}$ is well-defined. This is quite a restrictive assumption, since many generators of Feller processes involve differential operators which obviously cannot be applied to $\mathbb{1}_{\{0\}}$.
As an example of a generator for which the assumptions of Theorem 5.11 are satisfied, take the following example of a Markov jump process:

Example 5.12. Suppose $\mathbf{M}$ is a Markov chain satisfying the assumptions of Theorem 5.5 with $K(x,\{0\}) = 0$ for $x \neq 0$. Suppose $(N_t)_{t\geq 0}$ is a Poisson process, dependent on $\mathbf{M}$ with intensity function $\lambda(x)$. Suppose λ is bounded and $\lambda(x) \to \lambda^*$ locally uniformly as $|x| \to \infty$. Then $\mathbf{X}$, defined by $X_t = M_{N_t}$, is regularly varying with forward tail process $Y_t = \hat{Y}_{N_t^*}$ where $\hat{\mathbf{Y}}$ is a Markov chain with kernel K and $(N_t^*)_{t\geq 0}$ is a homogeneous Poisson process with constant intensity λ^*.
The generator of $\mathbf{X}$ is given by

$$\mathcal{A}f(x) = \lambda(x) \int f(y) - f(x)\Pi(x, dy)$$

for all bounded measurable functions and hence

$$\mathcal{A}f^u(ux) = \lambda(ux) \int f(y) - f(x)\Pi^u(x, dy) \to \lambda^* \int f(y) - f(x)K(x, dy) =: \mathcal{G}f(x)$$

This is the generator of $\mathbf{Y} = \hat{\mathbf{Y}} \circ N^*$, as defined above. Note that for both generators their domains can be taken as the space of bounded measurable functions. Since $K(x,\{0\}) = 0$ we have $\mathcal{G}^n \mathbb{1}_{\{0\}}(x) = 0$ for all $n \in \mathbb{N}$ and therefore $Q_t(x,\{0\}) = 0$. It follows that $\mathbf{X}$ has forward tail process $\mathbf{Y}$.

This example can be seen as an extension of Corollary 4.11, as the processes $\mathbf{M}$ and N can be dependent in this example, while Corollary 4.11 assumed independence.
As it turns out, Markov jump processes (as the one in Example 5.12) are the only processes with bounded generator, see [Kal21] Exercise 17.14. Therefore Theorem 5.12 leaves a lot to be desired.

The following Theorem is inspired by Aldous' tightness criterion, see [Bil99] Theorem 16.10. It provides a sufficient condition for regular variation of a process, which in the case of strong Markov processes is easier to verify than the conditions of Theorem 4.11.

Theorem 5.13. *Let $\boldsymbol{X}$ be a stochastically continuous stationary càdlàg process. Suppose $\boldsymbol{X}$ is finite-dimensional regularly varying and for all sequences $(\tau_n)_{n\in\mathbb{N}}$ of bounded optional stopping times and all non-negative sequences $(\delta_n)_{n\in\mathbb{N}}$ converging to 0 we have*

$$\limsup_{n\to\infty} \frac{\mathbb{P}[|X_{\tau_n} - X_{\tau_n+\delta_n}| \geq n\epsilon]}{\mathbb{P}[|X_0| > n]} = 0. \qquad (A)$$

Then $\boldsymbol{X}$ is regularly varying in D.

Proof. The proof is analogous to the one given in [Bil99] Theorem 16.10.
Step 1: First, we prove that assumption (A) implies

$$\lim_{\delta\to 0} \limsup_{n\to\infty} \frac{\mathbb{P}[|X_{\tau_1} - X_{\tau_2}| \geq n\epsilon, \tau_2 - \tau_1 \leq \delta]}{\mathbb{P}[|X_0| > n]} = 0 \qquad (B)$$

for all $\epsilon > 0$ and bounded optional stopping times τ_1, τ_2.
Let $\tau \leq T < \infty$ and choose δ_0, n_0 such that

$$\frac{\mathbb{P}[|X_\tau - X_{\tau+\delta}| \geq n\epsilon]}{\mathbb{P}[|X_0| > n]} \leq \eta, \qquad n \geq n_0, \delta \leq \delta_0.$$

Let μ be the uniform distribution on $[0, \delta]$.
For a given $x \in D$ and $t_1 \leq t_2$ with $t_2 - t_1 \leq \delta$, define $J = [0, 2\delta], I = [0, \delta]$ and

$$M_i = \{s \in J : |x_{t_i+s} - x_{t_i}| < n\epsilon\}.$$

Suppose $\mu(M_i) > 3/4$. Then

$$\mu(M_2 \cap I) = 1 - \mu(M_2^C \cup I^C) \geq 1 - \mu(M_2^C) - \mu(I^C) \geq 1/4$$

and therefore since $t_2 - t_1 \leq \delta$

$$\mu((M_2 + (t_2 - t_1)) \cap J) \geq \mu((M_2 \cap I) + (t_2 - t_1)) = \mu(M_2 \cap I) > 1/4.$$

This means that

$$\mu(M_1) + \mu((M_2 + (t_2 - t_1)) \cap J) > 1$$

and hence $M_1 \cap (M_2 + (t_2 - t_1)) \neq \emptyset$. This means, there exists $s \in M_1$ such that $s - (t_2 - t_1) \in M$ and therefore

$$|x_{t_1} - x_{t_2}| \leq |x_{t_1} - x_{t_1+s}| + |x_{t_2+s-(t_2-t_1)} - x_{t_2}| \leq 2n\epsilon.$$

Equivalently, if $t_2 - t_1 \leq \delta$ but $|x_{t_1} - x_{t_2}| > 2n\epsilon$ then either $\mu(M_1^C) \geq 1/4$ or $\mu(M_2^C) \geq 1/4$. Now we get for $\Theta \sim \mu$ independent of X:

$$\begin{aligned}
&\frac{\mathbb{P}[|X_{\tau_2} - X_{\tau_1}| \geq 2n\epsilon, \tau_2 - \tau_1 \leq \delta]}{\mathbb{P}[|X_0| > n]} \\
&\leq \sum_{i=1}^{2} \frac{\mathbb{P}\Big[\mathbb{P}\big[|X_{\tau_i+\Theta} - X_{\tau_1}| \geq n\epsilon \big| X\big] \geq 1/4\Big]}{\mathbb{P}[|X_0| > n]} \\
\text{(Markov's inequality)} &\leq 4 \sum_{i=1}^{2} \frac{\mathbb{P}[|X_{\tau_i+\Theta} - X_{\tau_i}| \geq n\epsilon]}{\mathbb{P}[|X_0| > n]}
\end{aligned}$$

which converges to 0 since Θ is independent of X. Hence (A) implies (B).

Step 2: Now we verify condition 2 of Theorem 4.11 using condition (B). Suppose (A), and equivalently (B), holds. Define

$$\Delta_K = \left\{ \frac{j}{2^k} : j \in \mathbb{Z} \right\}$$

and the stopping times $\tau_0^n = 0$,

$$\tau_i^n = \min\{t \in \Delta_k : \tau_{i-1}^n < t \leq T, |X_t - X_{\tau_{i-1}^n}| \geq n\epsilon\}$$

where $\tau_i^n = T$ if no such t exists. For given $\epsilon, \eta > 0$ choose n_0, δ_0 such that

$$\frac{\mathbb{P}[|X_{\tau_i^n} - X_{\tau_{i-1}^n}| \geq n\epsilon, \tau_i^n - \tau_{i-1}^n \leq \delta_0]}{\mathbb{P}[|X_0| > n]} \leq \eta, \qquad n \geq n_0.$$

Note that $\tau_i^n < T$ implies that $|X_{\tau_i^n} - X_{\tau_{i-1}^n}| \geq n\epsilon$ and therefore

$$\frac{\mathbb{P}[\tau_i^n < T, \tau_i^n - \tau_{i-1}^n \leq \delta_0]}{\mathbb{P}[|X_0| > n]} \leq \eta, \qquad n \geq n_0.$$

Choose an integer q such that $q \cdot \delta_0 \geq 2T$. For large enough n_0 there is a $\delta > 0$ such that

$$\frac{\mathbb{P}[\tau_i^n < T, \tau_i^n - \tau_{i-1}^n \leq \delta]}{\mathbb{P}[|X_0| > n]} \leq \eta/q, \qquad n \geq n_0.$$

It follows that

$$\begin{aligned}
&\mathbb{P}[|X_0| > n]^{-1} \mathbb{P}\bigg[\bigcup_{i=1}^{q} \{\tau_i^n < T, \tau_i^n - \tau_{i-1}^n \leq \delta\}\bigg] \\
&\leq \mathbb{P}[|X_0| > n]^{-1} \sum_{i=1}^{q} \mathbb{P}[\tau_i^n < T, \tau_i^n - \tau_{i-1}^n \leq \delta] \leq \eta.
\end{aligned}$$

By Markov's inequality we obtain

$$\mathbb{P}[\tau_i^n - \tau_{i-1}^n \geq \delta | \tau_q^n < T] \leq \delta^{-1} \mathbb{E}[\tau_i^n - \tau_{i-1}^n | \tau_q^n < T]$$

which implies

$$\begin{aligned}
\mathbb{E}[\tau_i^n - \tau_{i-1}^n | \tau_q < T] &\geq \delta \mathbb{P}[\tau_i^n - \tau_{i-1}^n \geq \delta | \tau_q^n < T] = \delta(1 - \mathbb{P}[\tau_i^n - \tau_{i-1}^n < \delta | \tau_q^n < T]) \\
&= \delta\bigg(1 - \frac{\mathbb{P}[\tau_i^n - \tau_{i-1}^n < \delta, \tau_q^n < T]}{\mathbb{P}[\tau_q^n < T]}\bigg) \\
&\geq \delta\bigg(1 - \frac{\eta \mathbb{P}[|X_0| > n]}{\mathbb{P}[\tau_q^n < T]}\bigg)
\end{aligned}$$

and hence

$$T \geq \mathbb{E}[\tau_q^n | \tau_q^n < T] = \sum_{i=1}^{q} \mathbb{E}[\tau_i^n - \tau_{i-1}^n | \tau_q^n < T] \geq \delta q \bigg(1 - \frac{\eta \mathbb{P}[|X_0| > n]}{\mathbb{P}[\tau_q^n < T]}\bigg)$$

By choice of q we have $q\delta > 2T$ and therefore

$$\frac{\mathbb{P}[\tau_q^n < T]}{\mathbb{P}[|X_0| > n]} < 2\eta.$$

In total, we get

$$\begin{aligned}
&\frac{\mathbb{P}\bigg[\{\tau_q^n < T\} \cup \bigcup_{i=1}^{q} \{\tau_i^n < T, \tau_i^n - \tau_{i-1}^n \leq \delta\}\bigg]}{\mathbb{P}[|X_0| > n]} \\
&\leq \frac{\mathbb{P}[\tau_q^n < T]}{\mathbb{P}[|X_0| > n]} + \sum_{i=1}^{q} \frac{\mathbb{P}[\tau_i^n < T, \tau_i^n - \tau_{i-1}^n \leq \delta]}{\mathbb{P}[|X_0| > n]} \leq 3\eta.
\end{aligned}$$

Now define the set

$$A_{n,k} = \left(\{\tau_q < T\} \cup \bigcup_{i=1}^{q} \{\tau_i^n < T, \tau_i^n - \tau_{i-1}^n \leq \delta\} \right)^C$$

and the random integer $v = \min\{i : \tau_i^n = T\}$. By definition, on $A_{n,k}$ there is a partition $0 = t_0^k < \cdots < t_v^k = T$ such that $t_i^k - t_{i-1}^k > \delta$ for $1 \leq i < v$ and for $s, t \in \Delta_k \cap [t_{i-1}^k, t_i^k)$ we have $|X_s - X_t| \leq n\epsilon$. Letting $k \to \infty$ and defining $A_n = \limsup_{k\to\infty} A_{n,k}$ we know that since the t_i^k are bounded by T there exists a subsequence $(k_i)_{i\in\mathbb{N}}$ such that the time points $t_j^{k_i}$ converge to some t_j. Then by right-continuity of X we get that $|X_s - X_t| \leq n\epsilon$ for $s, t \in [t_{j-1}, t_j)$. Finally, note that if $w'(X(\omega), 0, T, \delta) \geq n\epsilon$ there must exist s, t with $|t - s| < \delta$ and $|X_t - X_s| \geq n\epsilon$ and hence $\omega \in A_{n,k}^C$ which implies

$$\frac{\mathbb{P}[w'(X, 0, T, \delta) \geq n\epsilon]}{\mathbb{P}[|X_0| > n]} \leq \limsup_{k\to\infty} \frac{\mathbb{P}[A_{n,k}^C]}{\mathbb{P}[|X_0| > n]} = 3\eta, \qquad n \geq n_0.$$

Since η was arbitrary, $\mathbf{X}$ satisfies condition 2 of Theorem 4.11 and hence is regularly varying in D. □

Example 4.7 continued

Using Theorem 5.13 we can now prove that the Ornstein-Uhlenbeck process is regularly varying in D.

Since $\mathbf{X}$ is a Feller process, the tightness criterion of Theorem 5.13 is satisfied if

$$\frac{\mathbb{P}[|X_0 - X_{\delta_n}| \geq n\epsilon]}{\mathbb{P}[|X_0| > n]} \to 0$$

for a sequence $\delta_n \to 0$. Note that

$$\begin{aligned}\mathbb{P}[|X_0 - X_{\delta_n}| \geq n\epsilon] &= \mathbb{P}\left[\left|X_0(1 - e^{-\theta\delta_n}) + \sigma \int_0^{\delta_n} e^{-\theta(\delta_n - s)} dL_s\right| \geq n\epsilon\right] \\ &\leq \mathbb{P}[|X_0(1 - e^{-\theta\delta_n})| \geq n\epsilon/2] + \mathbb{P}\left[\left|\int_0^{\delta_n} e^{-\theta(\delta_n - s)} dL_s\right| \geq n\epsilon/2\right].\end{aligned}$$

Clearly we have

$$\frac{\mathbb{P}[|X_0(1 - e^{-\theta\delta_n})| \geq n\epsilon]}{\mathbb{P}[|X_0| > n]} \to 0.$$

Now for the integral part, we use the estimate given in [RW86] Theorem 2.1 which states

$$\sup_{\lambda>0} \lambda^\alpha \mathbb{P}\left[\sup_{0\leq t\leq\delta_n} \left|\int_0^t e^{-\theta(t-s)} dL_s\right| > \lambda\right] \leq c \cdot \int_0^{\delta_n} e^{-\alpha\theta(\delta_n - s)} ds =: C_{\delta_n}$$

for some constant $c > 0$. This implies that

$$\mathbb{P}\left[\sigma\left|\int_0^{\delta_n} e^{-\theta(\delta_n - s)} dL_s\right| \geq n\epsilon\right] \leq (n\epsilon)^{-\alpha} C_{\delta_n}$$

and together with the asymptotic tail estimate of the symmetric α-stable distribution

$$\mathbb{P}[|X_0| > t] \sim t^{-\alpha}\hat{c}$$

for some constant $\hat{c}$ we get

$$\frac{\mathbb{P}\left[\sigma\left|\int_0^{\delta_n} e^{-\theta(\delta_n - s)} dL_s\right| \geq n\epsilon\right]}{\mathbb{P}[|X_0| > n]} \lesssim \frac{(n\epsilon)^{-\alpha} C_{\delta_n}}{n^{-\alpha}\hat{c}} = \frac{\epsilon^{-\alpha}}{\hat{c}} C_{\delta_n} \to 0.$$

Since X is finite-dimensionally regularly varying and satisfies the tightness criterion of Theorem 5.13, X is regularly varying in D.

6 Summary

In summary, the tail measure on D can be used to study the extremal behaviour of a regularly varying stochastic processes. While this measure may be difficult to determine purely by the expression given in Definition 4.1, we have found how processes like the tail process $\mathbf{Y}$ (Definition 4.5), the spectral process $\mathbf{Z}$ (Theorem 4.13) or the process $\mathbf{Q}$ (Theorem 4.19) can be used to compute the tail measure more efficiently. In order to verify the regular variation of a continuous-time process, more than just the finite-dimensional regular variation (detailed in Section 2) was necessary. Theorem 4.11 gave the additional necessary and sufficient condition to bridge the gap between finite-dimensional regular variation and regular variation in D, which paralleled the convergence theorem given in Section 3.

Our contributions were given in Section 5, where we extended some known theorems about regularly varying Markov chains to continuous-time Markov processes. In our estimation, Theorem 5.10 and Theorem 5.13, which allowed us to more easily verify the regular variation of Markov processes as well as compute their tail processes, are the most noteworthy results.

Theorem 5.13, together with the tail bounds for stochastic integrals given in [RW86] Theorem 2.1, can likely be used to verify regular variation of a much more general class of Lévy-driven stochastic integrals than the one given in Example 4.7.

References

[BHS22] Martin Bladt, Enkelejd Hashorva, and Georgiy Shevchenko. Tail measures and regular variation. *Electronic Journal of Probability*, 27:1–43, 2022.

[Bil99] Patrick Billingsley. *Convergence of probability measures*, volume 2. John Wiley & Sons, 1999.

[BS09] Bojan Basrak and Johan Segers. Regularly varying multivariate time series. *Stochastic processes and their applications*, 119(4):1055–1080, 2009.

[DHS18] Clément Dombry, Enkelejd Hashorva, and Philippe Soulier. Tail measure and spectral tail process of regularly varying time series. *Annals of Applied Probability*, 28(6):3884–3921, 2018.

[EK09] Stewart N Ethier and Thomas G Kurtz. *Markov processes: characterization and convergence*. John Wiley & Sons, 2009.

[HF06] Laurens Haan and Ana Ferreira. *Extreme value theory: an introduction*, volume 3. Springer, 2006.

[HL05] Henrik Hult and Filip Lindskog. Extremal behavior of regularly varying stochastic processes. *Stochastic Processes and their applications*, 115(2):249–274, 2005.

[JS14] Anja Janssen and Johan Segers. Markov tail chains. *Journal of Applied Probability*, 51(4):1133–1153, 2014.

[Kal21] Olav Kallenberg. *Foundations of modern probability*, volume 3. Springer, 2021.

[KS20] Rafal Kulik and Philippe Soulier. *Heavy-tailed time series*. Springer, 2020.

[Küh18] Franziska Kühn. Solutions of lévy-driven sdes with unbounded coefficients as feller processes. *Proceedings of the American Mathematical Society*, 146(8):3591–3604, 2018.

[Lea82] M Ross Leadbetter. Extremes and local dependence in stationary sequences. Technical report, NORTH CAROLINA UNIV AT CHAPEL HILL DEPT OF STATISTICS, 1982.

[Lin04] Filip Lindskog. *Multivariate extremes and regular variation for stochastic processes*. PhD thesis, ETH Zurich, 2004.

[PS18] Hrvoje Planinić and Philippe Soulier. The tail process revisited. *Extremes*, 21:551–579, 2018.

[RAS19] Alfonso Rocha-Arteaga and Ken-iti Sato. Selfsimilar additive processes and stationary ornstein–uhlenbeck type processes. *Topics in Infinitely Divisible Distributions and Lévy Processes, Revised Edition*, pages 61–75, 2019.

[RW86] Jan Rosinski and WA Woyczynski. On itô stochastic integration with respect to p-stable motion: inner clock, integrability of sample paths, double and multiple integrals. *The Annals of Probability*, pages 271–286, 1986.

[RZ13] Sidney I Resnick and David Zeber. Asymptotics of markov kernels and the tail chain. *Advances in Applied Probability*, 45(1):186–213, 2013.

[SO12] Gennady Samorodnitsky and Takashi Owada. Tail measures of stochastic processes or random fields with regularly varying tails. *preprint*, 2012.

[Sou22] Philippe Soulier. The tail process and tail measure of continuous time regularly varying stochastic processes. *Extremes*, 25(1):107–173, 2022.

Hiermit erkläre ich, dass ich die vorliegende Arbeit selbstständig und ohne Benutzung anderer als der angegebenen Quellen und Hilfsmittel angefertigt habe.

Magdeburg, den 20. Februar 2023