KUNGL
TEKNISKA
HÖGSKOLAN

Topics on fractional Brownian motion and regular variation for stochastic processes

Henrik Hult

Stockholm 2003

Doctoral Dissertation
Royal Institute of Technology
Department of Mathematics

Akademisk avhandling som med tillstånd av Kungl Tekniska Högskolan framläggs till offentlig granskning för avläggande av teknologie doktorsexamen fredagen den 3 oktober 2003 kl 10.00 i Sal Q1, Osquldas väg 6, Kungl. Tekniska Högskolan, Stockholm.

ISBN 91-7283-573-7

TRITA-MAT2003-MS01
ISSN 1401-2278

© Henrik Hult, Augusti 2003

Universitetsservice US–AB, Stockholm 2003

Abstract

The first part of this thesis studies tail probabilities for elliptical distributions and probabilities of extreme events for multivariate stochastic processes. It is assumed that the tails of the probability distributions satisfy a regular variation condition. This means, roughly speaking, that there is a non-negligible probability for very large or extreme outcomes to occur. Such models are useful in applications including insurance, finance and telecommunications networks. It is shown how regular variation of the marginals, or the increments, of a stochastic process implies regular variation of functionals of the process. Moreover, the associated tail behavior in terms of a limit measure is derived.

The second part of the thesis studies problems related to parameter estimation in stochastic models with long memory. Emphasis is on the estimation of the drift parameter in some stochastic differential equations driven by the fractional Brownian motion or more generally Volterra-type processes. Observing the process continuously, the maximum likelihood estimator is derived using a Girsanov transformation. In the case of discrete observations the study is carried out for the particular case of the fractional Ornstein-Uhlenbeck process. For this model Whittle's approach is applied to derive an estimator for all unknown parameters.

ISBN 91-7283-573-7 • TRITA-MAT2003-MS01 • ISSN 1401-2278

Contents

1 Introduction **1**
1.1 Regular variation and extremal events 1
1.2 Fractional Brownian motion and parameter estimation 8
References 13

I Regular variation and stochastic processes **17**

2 Multivariate extremes in elliptical distributions **19**
2.1 Introduction 20
2.2 Preliminaries 21
2.3 Elliptical distributions 24
2.4 Main Results 26
2.5 Multivariate extremes for elliptical distributions 28
2.6 Proofs 33
References 43

3 Multivariate regular variation for additive processes **45**
3.1 Introduction 46
3.1.1 Notation 48
3.2 Multivariate regular variation 48
3.3 Additive processes 51
3.4 Sums of regularly varying random vectors 53
3.5 Regular variation for additive processes and functionals 57
References 72

4 On regular variation for stochastic processes **75**
4.1 Introduction 76
4.2 Regular variation on D 79
4.2.1 Proofs 82
4.3 Markov processes with asymptotically independent increments . . 85
4.3.1 Proofs 90

4.4 Filtered Markov processes 95
References 98
Appendix 99

II Fractional Brownian motion and parameter estimation 103

5 Approximating some Volterra type stochastic integrals 105
5.1 Introduction 106
5.2 Preliminaries and definitions 108
5.2.1 Gaussian processes of Volterra type 109
5.2.2 The reproducing kernel Hilbert space 115
5.3 Representation of Volterra type stochastic integrals 117
5.3.1 A wavelet representation of fractional Brownian motion . . 119
5.4 Stochastic integrals with respect to Volterra type processes 122
5.4.1 Malliavin calculus 123
5.4.2 A stochastic integral with respect to a Volterra type process 125
5.5 Applications to parameter estimation 130
5.5.1 Deterministic drift 134
5.5.2 The fractional Ornstein-Uhlenbeck type process. 135
References 137

6 Estimation for the fractional Ornstein-Uhlenbeck process 139
6.1 Introduction 140
6.2 The fractional Ornstein-Uhlenbeck process 141
6.3 Parameter estimation based on discrete observations 144
6.4 Numerical illustrations 148
References 149

Acknowledgments

This thesis have been written during my time as a Ph.D. student at the division of Mathematical Statistics. First of all I would like to thank my supervisor, Prof. Boualem Djehiche, for sharing with me his knowledge on probability theory and for introducing me to the topics studied in the thesis. He has also during the whole time been enthusiastic to discuss problems with me, of all kinds. Moreover, he has always had confidence in me and given me support and continuous encouragement. I am also grateful for the support coming from Prof. Lars Holst who has also introduced me to many fields within the theory of probability. In particular, the theory of weak convergence of measures. I am also grateful for having so many nice colleagues at the Mathematical Statistics division at KTH. Working with them has always been very stimulating.

I wish to send a special thank to Prof. Paul Embrechts who has, on several occasions, welcomed me to stay at ETH in Zürich. There I have had the opportunity to collaborate with my friend and colleague Filip Lindskog and I have also benefited from the academic environment there.

Finally, I want to thank Prof. Tobias Rydén for welcoming me to stay at the division of Mathematical Statistics at Lund Institute of Technology for the last months, where I finished writing the thesis.

Stockholm, 2003-08-20

Chapter 1

Introduction

This thesis is divided into two parts. The first part consists of three papers included in Chapters 2–4 and studies tail probabilities for elliptical distributions and extreme events for stochastic processes. The second part of the thesis consists of two papers included in Chapters 5 and 6 and studies parameter estimation problems related to the fractional Brownian motion. We will first give an introduction to the topics covered in the thesis and highlight the main results.

1.1 Regular variation and extremal events

In this section we give an introduction to regular variation and the study of extremal events. We start by discussing random variables with heavy tails. Suppose X is a random variable with distribution function F. The outcome of X may be thought of as the measurement of a sea-level, the daily loss from investing in a stock, the total amount of claims faced by an insurance company in one year etc. In such applications it is relevant to compute the probability of a very large (extreme) outcome, for instance the probability that the sea-level exceeds a high barrier, the probability that we make a large loss from an investment in the stock, or the probability that the total amount of claims faced by the insurance company in one year exceeds a high threshold. This means, we would have to compute $1 - F(x)$ where x is large. For this reason it is of course important to know what the distribution function F looks like for large x, e.g. at which rate the function $1 - F(x)$ tends to zero as $x \to \infty$. If the decay is fast then the probability mass is concentrated around the center of the distribution. As an example we may consider the standard normal distribution where $1-F(x) \sim (x\sqrt{2})^{-1} \exp(-x^2/2)$ as $x \to \infty$. In this case we say that the distribution has *light* (right) tail. If, on the other hand, the decay is slow, then there is a significant amount of probability mass far out in the (right) tail of the distribution. The slow decay of the probability distribution as $x \to \infty$ is often referred to as *heavy* (right) tail. As an example we may consider

the Pareto distribution where $1 - F(x) \sim x^{-\alpha}$, as $x \to \infty$ and $\alpha > 0$. Similar considerations can of course be made for the left tail as well. Then we consider the rate at which $F(x)$ tends to zero as $x \to -\infty$. There is no precise definition of how fast or slow the decay must be to say that a probability distribution has light or heavy tails. We will work with a more precise concept called regular variation, which specifies the rate at which $1 - F(x)$ tends to zero.

In heavy-tailed distributions large values may occur in a sample with non-negligible probability. This is often observed in insurance data, for instance in the so-called catastrophe insurances including fire, wind-storm and flooding insurances. The large claims may lead to large fluctuations in the cash-flow process faced by the insurance company, increasing the risk in such portfolios. The situation is similar in finance where extremely large losses sometimes occur, which indicate heavy tails of the return distributions. The probability of extreme stock-movements has to be accounted for when analyzing the risk of a portfolio. Another application is queuing models where extreme service times, modeled by heavy-tailed distributions, result in huge waiting times in the system and large fluctuations in the workload process.

In many applications it is appropriate to use a stochastic process $\{X_t : t \geq 0\}$ to model the evolution of the interesting quantities over time. The notion of heavy tails enters naturally in this context either as an assumption on the marginals X_t or as an assumption on the increments $X_{t+h} - X_t$ of the process. However, it is often the case that the marginals or the increments of the process is not the main concern. Instead some functional of the process is important. A natural example is the supremum of the process during a time interval, $\sup_{0 \leq t \leq T} X_t$, i.e. the largest value reached by the process in the interval $[0, T]$. Another example is the mean of the process during a time interval, $T^{-1} \int_0^T X_t \mathrm{d}t$. We are then typically interested in the probability that the functional exceeds some high level, e.g. –What is the probability that the sea-level exceeds a high barrier sometime during $[0, T]$? It may therefore be important to know how the tail behavior of the marginals X_t (or the increments) is related to the tail behavior of functionals of the process. This problem is studied in a multivariate context in Chapters 3 and 4 below.

The concept of regular variation comes from pure mathematics but has found many applications in probability theory. A function f supported on $(0, \infty)$ is said to be *regularly varying at* ∞ *with index* $\rho \in \mathbb{R}$ if for all $x > 0$,

$$\lim_{u \to \infty} \frac{f(ux)}{f(u)} = x^{\rho}. \tag{1.1}$$

If a nonnegative random variable X is distributed according to F we say that X is regularly varying with index $\alpha > 0$ if $1 - F$ is regularly varying at ∞ with index $-\alpha$. In this case we may write the regular variation condition above as

$$\lim_{u \to \infty} \frac{\mathbb{P}(X > ux)}{\mathbb{P}(X > u)} = x^{-\alpha}. \tag{1.2}$$

If X is regularly varying then we can write $\mathbb{P}(X > x) = x^{-\alpha}L(x)$ where L is a slowly varying function (see e.g. Resnick [44]). Hence, the tail of the distribution decays essentially as $x^{-\alpha}$ for large x.

Let us now consider the multivariate case. Suppose that we are dealing with a d-dimensional random vector $\mathbf{X}$ instead of a univariate random variable. This could be interpreted for instance as the measurements of sea-levels at d different locations, the daily losses of d different stocks or the amount of claims in d different insurances in one year. A notable difference between the multivariate case and the univariate case when analyzing extreme values is the possibility to have dependence between the components of the random vector. Large values can for instance tend to occur simultaneously in the different components. To have a good understanding of the dependence between extreme events in the multivariate case may be of great importance in applications. In some cases the dependence between extreme values is even stronger than the dependence of moderate values. As an example we may consider the daily log-returns of BMW and Siemens AG stocks during the period 020189–020196 (see Figure 1.1). Note that large negative shocks tend to occur

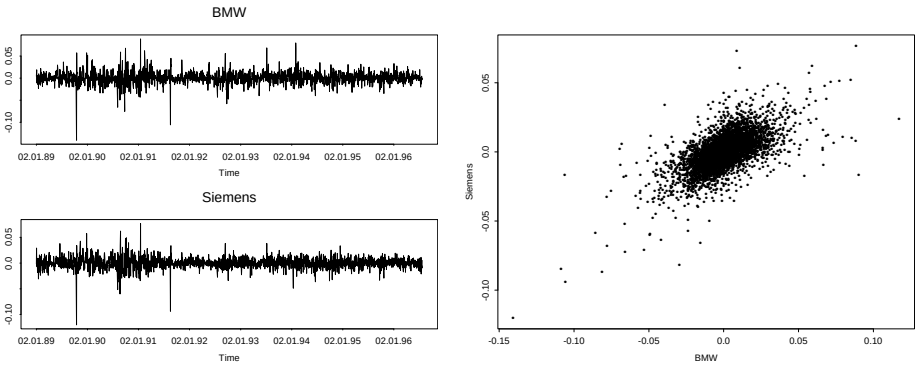

Figure 1.1. The log-returns of BMW and Siemens AG stocks during 020189–020196. Left: The evolution of the log-returns over time. Right: Scatter plot of the log-returns.

simultaneously in both assets. In fact we observe that the dependence seems to be stronger for extreme losses (third quadrant) than for extreme profits (first quadrant) or for moderate values. It is essential to take this fact into account when analyzing the risk of a portfolio with investments in both stocks. For more examples of dependent extreme values in sea-level, storm and wind data see e.g. de Haan and de Ronde [29] and Rootzén and Tajvidi [47]. For examples covering exchange rates in finance see Stărică [52].

In the univariate case we made the assumption that the probability distributions are regularly varying. A similar assumption will be made also in the multivariate setting. This is usually referred to simply as multivariate regular variation. We

denote by $\mathbb{S}^{d-1}$ the unit hypersphere in $\mathbb{R}^d$ with respect to a norm $|\cdot|$, and by $\mathcal{B}(\mathbb{S}^{d-1})$ the Borel σ-algebra on $\mathbb{S}^{d-1}$. A d-dimensional random vector $\mathbf{X}$ is said to be *multivariate regularly varying with index* $\alpha > 0$ if there exists a probability measure σ on $\mathbb{S}^{d-1}$ such that for every $x > 0$, as $u \to \infty$,

$$\frac{\mathbb{P}(|\mathbf{X}| > ux, \mathbf{X}/|\mathbf{X}| \in \cdot\,)}{\mathbb{P}(|\mathbf{X}| > u)} \xrightarrow{w} x^{-\alpha}\sigma(\cdot) \text{ on } \mathcal{B}(\mathbb{S}^{d-1}). \tag{1.3}$$

The probability measure σ is referred to as the *spectral measure* of $\mathbf{X}$ and α is referred to as the *tail index* of $\mathbf{X}$.

The concept of multivariate regular variation is formulated in terms of weak convergence $\xrightarrow{w}$ of measures. Since $\mathbb{S}^{d-1}$ is compact weak convergence coincides with vague convergence (denoted $\xrightarrow{v}$) on this space. Therefore we may alternatively formulate multivariate regular variation in terms of vague convergence (see Chapter 2). For a more complete account on weak and vague convergence we refer to Kallenberg [31] or Daley and Vere-Jones [15]. Note that if we plug in the set $\mathbb{S}^{d-1}$ in the definition then

$$\frac{\mathbb{P}(|\mathbf{X}| > ux)}{\mathbb{P}(|\mathbf{X}| > u)} \to x^{-\alpha},$$

and hence $|\mathbf{X}|$ is regularly varying at ∞ with index α according to (1.2). On the other hand, if we put $x = 1$ then, as $u \to \infty$,

$$\mathbb{P}(\mathbf{X}/|\mathbf{X}| \in \cdot\,|\,|\mathbf{X}| > u) \xrightarrow{w} \sigma(\cdot) \text{ on } \mathcal{B}(\mathbb{S}^{d-1}).$$

The spectral measure gives information on in which directions we are likely to find extreme realizations of the random vector $\mathbf{X}$, whereas α is related to the radial decay of the probability distribution. Note that α does not depend on the direction.

An interesting class of multivariate distributions which is widely applicable is the class of elliptical distributions. This class is an extension of the multivariate normal distributions and may, roughly speaking, be thought of as the class of multivariate distributions whose probability density functions have elliptically shaped level-sets. A thorough introduction to elliptical distributions can be found in Fang, Kotz and Ng [23] and Cambanis, Huang and Simons [9] where many interesting properties are derived. They are also presented in more detail in Chapter 2. The elliptical distributions share many of the tractable properties of the multivariate normal distributions but contrary to the multivariate normal distributions they can also be used in applications where heavy tails are present. This makes the class of elliptical distributions particularly useful, for instance in applications in risk management (see e.g. Embrechts, McNeil and Straumann [22]). As indicated above we sometimes encounter data sets where we have strong dependence of extreme values. Since elliptical distributions have a rather limited flexibility in the dependence structure it is important to understand if their dependence structure is sufficiently rich to accurately model dependent extreme values. In Chapter 2 we study measures of extremal dependence for elliptical distributions.

As previously mentioned it is often appropriate to model the interesting quantities in a dynamical way, as a continuous time stochastic process $\{\mathbf{X}_t : t \in [0,T]\}$. Multivariate regular variation enters naturally as an assumption on the marginals $\mathbf{X}_t$ or as an assumption on the increments $\mathbf{X}_{t+h} - \mathbf{X}_t$ of the process. Similar to the univariate case some functional or vector of functionals of the process may be our primary concern. Natural examples are for instance the componentwise suprema of the process, $(\sup_{0\leq t\leq T} \mathbf{X}_t^{(1)}, \ldots, \sup_{0\leq t\leq T} \mathbf{X}_t^{(d)})$ i.e. the largest value reached by each component of the process in the interval $[0,T]$. Another example is the componentwise mean of the process but other functionals and combinations of functionals may also be of interest. We are then typically interested in the probability that the vector of functionals belongs to some extreme set far away from the origin, e.g. –What is the probability that the sea-level exceeds a high barrier at some (or all) locations sometime during $[0,T]$? To answer this type of questions we need know how the tail behavior of the marginals $\mathbf{X}_t$ is related to the tail behavior of functionals of the process. This is the main problem studied in Chapters 3 and 4.

Let us now give a brief review of the relevant literature. The theory of regularly varying functions is a basic ingredient in the study of weak limits of independent and identically distributed (iid) random variables. The book by Feller [24] is an excellent exposition that makes the connection between regular variation and stable laws. The theory of regularly varying functions is also a fundamental ingredient in the study of extreme values for iid observations, which is explained in de Haan [28]. For an exhaustive treatment of regular variation and its applications in probability theory we refer to Bingham, Goldie and Teugels [7]. For a more recent account on the applications of regularly varying functions and extreme value theory we refer to Embrechts, Klüppelberg and Mikosch [20]. The study of regular variation for stochastic processes is, in the univariate case, often included in more general studies of subexponentiality. Embrechts, Goldie and Veraverbeke [19] proves tail equivalence of a subexponential infinitely divisible random variable and its associated Lévy measure. For results on the tail behavior of functionals of stable processes we refer to Samorodnitsky and Taqqu [50] and references therein. Rosiński and Samorodnitsky [48] derived general results on the tail behavior of subadditive functionals for infinitely divisible stochastic processes. Their paper covers many of the previously known results. Braverman, Mikosch and Samorodnitsky [8] also study the tail behavior of functionals of univariate Lévy processes. Multivariate regular variation was originally used to characterize the domain of attraction of sums of independent identically distributed random vectors that converges in distribution to a multivariate stable distribution (see Rvačeva [49]). The connection between multivariate regular variation and multivariate extreme value theory is explained in Resnick [44]. Kesten [32] used a formulation of multivariate regular variation in terms of linear combinations to conclude that the stationary solution of a multivariate linear stochastic recurrence equation is regularly varying. Several equivalent formulations of multivariate regular variation can be found in the literature. Many of them are documented in Basrak [4]. We refer also to Basrak, Davis and Miksoch

[5] for partial results on the equivalence of the formulation of multivariate regular variation used here (1.3) and the one used by Kesten [32] in terms of linear combinations (see also Chapter 3).

Next, we will give the outline of the first part of the thesis and highlight the main results. Chapter 2 is primarily concerned with the connection between elliptical distributions and multivariate regular variation. Elliptical distributions may be thought of as the class of multivariate distributions whose densities have elliptically shaped level-sets. The main interest in elliptical distributions comes from their usefulness in practice. This class of distributions provides a rich source of multivariate distributions which share many of the tractable properties of the multivariate normal distributions. However, contrary to the multivariate normal distributions, they enable the modeling of multivariate extremes and other forms of non-normal dependences such as tail dependence. It should be noted that, although the elliptical distributions are attractive in many applications, they do not provide a very flexible class of dependence structures. Chapter 2 studies the dependence structure of elliptical distributions and associated dependence measures. Particular emphasis is on the extremal dependence, i.e. dependence of extreme values. The simple dependence structure of elliptical distributions enables explicit computations of interesting dependence measures such as the coefficients of tail dependence and spectral measures associated with regularly varying random vectors.

The main results in Chapter 2 include Theorem 2.22 and a counterexample in Section 2.4. Theorem 2.22 states that for elliptical distributions the existence of tail dependence of the bivariate marginals and the fulfillment of the condition of multivariate regular variation is equivalent to univariate regular variation of the radial random variable in the general representation of elliptical distributions. Moreover, we derive an explicit formula for the coefficient of tail dependence for elliptical distributions. As for the counterexample we show that contrary to Kendall's tau, Spearman's rho is not invariant in the class of elliptical distributions with continuous marginals and a fixed dispersion matrix. We are also able to explicitly compute spectral measures with respect to different norms in Section 2.5.

Chapter 3 is concerned with multivariate stochastic processes with independent increments having marginal distributions which satisfy a multivariate regular variation condition. We find the tail-asymptotics of the distribution of vectors of functionals acting on the process. By tail-asymptotics we mean the limit measure associated with multivariate regular variation. This topic is rather well studied in a univariate context and even though the intuition behind the univariate results to a large extent extends to the multivariate case, proving results in the multivariate case requires other tools such as vague convergence of measures. In the univariate case much of the studies are done for the class of subexponential distributions which includes the class of regularly varying distributions. However, although attempts have been made to formulate subexponentiality in a multivariate framework (see Cline and Resnick [11]), the concept of multivariate subexponential distributions is not well understood. Therefore we have chosen to work with the class of multivariate regularly varying distributions.

The main results in Chapter 3 include Theorem 3.16, Theorem 3.19, Theorem 3.20 and Theorem 3.22. In Theorem 3.16 we prove tail equivalence between an infinitely divisible random vector $\mathbf{X}$ and its associated Lévy measure ν. This can be seen as a multivariate version, in the regularly varying case, of a result in Embrechts, Goldie and Veraverbeke [19] which says that

$$\mathbb{P}(X > x) \sim \nu(\{y \in \mathbb{R} : y > x\}) \quad \text{as } x \to \infty,$$

for X subexponential. In Theorem 3.19 we determine the implications of regular variation of $\mathbf{X}_t$ on the joint tail behavior of vectors of some functionals acting on the process. In particular, we study the tail behavior of the vector of componentwise suprema

$$\mathbf{X}_t^* = \Big(\sup_{0\le s\le t} X_s^{(1)}, \dots, \sup_{0\le s\le t} X_s^{(d)} \Big)$$

and the vector of componentwise suprema of the jumps

$$\mathbf{X}_t^\Delta = \Big(\sup_{0< s\le t} \Delta X_s^{(1)}, \dots, \sup_{0< s\le t} \Delta X_s^{(d)} \Big).$$

In Theorem 3.20 we give a formulation of regular variation for the graph of an additive process and relates it to regular variation of the marginals of the process. Theorem 3.22 determines the joint tail behavior of the vector of componentwise integrals

$$\mathbf{I}_t = \Big(\int_0^t X_s^{(1)} \mathrm{d}s, \dots, \int_0^t X_s^{(d)} \mathrm{d}s \Big).$$

In Chapter 4 we study general stochastic processes with sample paths in the space $D([0,1], \mathbb{R}^d)$ of right-continuous functions on $\mathbb{R}^d$ with left limits. A natural definition of regular variation for a stochastic process would be to say that it is regularly varying if all its finite dimensional distributions are multivariate regularly varying. However, the finite dimensional distributions alone give limited insights to the extremal behavior of the process and the tail behavior of functionals of the process. Instead we will treat the stochastic process as an element in $D([0,1], \mathbb{R}^d)$ and formulate regular variation on $D([0,1], \mathbb{R}^d)$. The definition is similar to the definition of multivariate regular variation. For explanation of the notation and for technical details we refer to Chapter 4. We say that a stochastic process with sample paths in $D([0,1], \mathbb{R}^d)$ is regularly varying if there exist $\alpha > 0$ and a probability measure σ on $D_1([0,1], \mathbb{R}^d) = \{\mathbf{x} \in D([0,1], \mathbb{R}^d) : \sup_{t\in[0,1]} |\mathbf{x}_t| = 1\}$ such that, as $u \to \infty$,

$$\frac{\mathbb{P}(|\mathbf{X}|_\infty > ux, \mathbf{X}/|\mathbf{X}|_\infty \in \cdot\,)}{\mathbb{P}(|\mathbf{X}|_\infty > u)} \xrightarrow{w} x^{-\alpha}\sigma(\cdot) \quad \text{on } \mathcal{B}(D_1([0,1], \mathbb{R}^d))$$

for every $x > 0$. Here $|\mathbf{x}|_\infty = \sup_{t\in[0,1]} |\mathbf{x}_t|$. The spectral measure σ contains all necessary information for understanding the extremal behavior of the process $\mathbf{X} = \{\mathbf{X}_t : t \in [0,1]\}$. One advantage with this formulation is that we can derive

a Continuous Mapping Theorem to obtain the tail behavior of many interesting mappings and functionals of the process. In Chapter 3 we studied processes with independent increments. For these processes extreme values are essentially due to one big jump. This intuition may be formalized in terms of the support of the spectral measure. For additive processes the spectral measure concentrates on

$$\{\mathbf{x} \in D_1([0,1],\mathbb{R}^d) : \mathbf{x}(\cdot) = \mathbf{y}1_{[u,1]}(\cdot), u \in [0,1], \mathbf{y} \in \mathbb{S}^{d-1}\},$$

i.e. functions with (at most) one jump. This is actually true also for a larger class of of regularly varying Markov processes which satisfy a condition of asymptotically independent increments. An application of the Continuous Mapping Theorem enables us to study of some filtered stochastic processes of the form

$$\mathbf{Y}_t = \int_0^t f(t,s)\mathrm{d}\mathbf{X}_s, \quad t \in [0,1], \tag{1.4}$$

where $\mathbf{X}$ is a regularly varying Markov process of finite variation. This includes some stable processes and serves as an example on how the formulation of regular variation on $D([0,1],\mathbb{R}^d)$ and the Continuous Mapping Theorem may be applied to obtain the tail behavior of more complicated processes.

The main findings in Chapter 4 include Theorem 4.5, Theorem 4.6, Theorem 4.8, Theorem 4.12 and Theorem 4.18. In Theorem 4.5 we derive a Continuous Mapping Theorem which enables us to obtain the tail behavior of mappings and functionals of a regularly varying stochastic process in $D([0,1],\mathbb{R}^d)$. In Theorem 4.6 we derive necessary and sufficient conditions for a stochastic processes with sample paths in $D([0,1],\mathbb{R}^d)$ to be regularly varying. Theorem 4.8 gives simplified sufficient conditions for regular variation on $D([0,1],\mathbb{R}^d)$ for a class of Markov processes with asymptotically independent increments. In Theorem 4.12 we derive the support of the spectral measure for this class of Markov processes. From this result we may conclude that the extremal behavior for these processes is due to (at most) one big jump. Finally, in Theorem 4.18 we derive the tail behavior of some filtered Markov processes with asymptotically independent increments and paths of finite variation.

1.2 Fractional Brownian motion and parameter estimation

In this section we provide a short introduction to the fractional Brownian motion and some of its applications. For a more thorough introduction we refer to Samorodnitsky and Taqqu [50]. The fractional Brownian motion was first studied by Kolmogorov [34] and it was given its name by Mandelbrot and Van Ness [39]. The study of the fractional Brownian motion appeared within the theory of self-similar processes. Self-similarity is a scaling property of the finite-dimensional distributions. A stochastic process $\{X_t : t \in \mathbb{R}\}$ is self-similar with parameter H

if $\{X_{ct} : t \in \mathbb{R}\}$ and $\{c^H X_t : t \in \mathbb{R}\}$ have the same finite-dimensional distributions for all $c > 0$. This scaling property is surprisingly often observed in various applications. For instance in telecommunications networks, hydrology and finance. It turns out that the fractional Brownian motion is the only Gaussian self-similar process with stationary increments. This is the main motivation for studying it. There are many ways to define the fractional Brownian motion. The simplest, to our knowledge, being the following. The fractional Brownian motion with index $H \in (0,1)$, denoted $\{B_t^H : t \in \mathbb{R}\}$ is the zero mean Gaussian process with covariance function

$$r(t,s) = \frac{1}{2}(|t|^{2H} + |s|^{2H} - |t-s|^{2H}).$$

When $H = 1/2$ then $r(t,s) = \min(t,s)$ and the fractional Brownian motion coincides with the standard Brownian motion. Note that for $H < 1/2$ the increments are negatively correlated whereas they are positively correlated for $H > 1/2$. For $H = 1/2$ they are independent. Let us now have a closer look at the increments of the fractional Brownian motion. Consider for simplicity increments of length 1. Let

$$Y_j = B_{j+1}^H - B_j^H, \quad j \in \mathbb{Z}.$$

This is a stationary process and its covariance function is

$$r(j) = \frac{1}{2}(|j+1|^{2H} + |j-1|^{2H} - 2|j|^{2H}), \quad j \in \mathbb{Z}.$$

The process $\{Y_j : j \in \mathbb{Z}\}$ is often referred to as fractional Gaussian noise. An interesting property is that

$$r(j) \sim H(2H-1)j^{2H-2}, \quad \text{as } j \to \infty.$$

This means that for $H > 1/2$ the sum of correlations diverges, i.e.

$$\sum_{j=0}^{\infty} r(j) = \infty.$$

This property is often referred to as long memory or long-range dependence. It has also served as a motivation for studying the fractional Brownian motion and the fractional Gaussian noise.

We note at this point that according to (1.1) the covariance function $r(j)$ is regularly varying at ∞ with index $\rho = 2H-2$. This is probably the only connection between the first and second part of the thesis.

The most commonly used representation of the fractional Brownian motion is the so-called moving average representation. Let $\{B_t : t \in \mathbb{R}\}$ be a standard

Brownian motion on the real line. Then the fractional Brownian motion with index $H \in (0,1)$ can be represented as

$$B_t^H = C_1^{-1}(H) \int_{-\infty}^{\infty} \left[(t-s)_+^{H-1/2} - (-s)_+^{H-1/2}\right] \mathrm{d}B_s, \quad t \in \mathbb{R},$$
$$C_1(H) = \Big\{ \int_0^{\infty} \left((1+s)^{H-1/2} - s^{H-1/2}\right)^2 \mathrm{d}s + \frac{1}{2H} \Big\}^{1/2}.$$

The fractional Brownian motion can also be given an integral representation on a compact interval. By self-similarity it is sufficient to give a representation on $[0,1]$. Let $\{B_t : t \in [0,1]\}$ be a standard Brownian motion. Then

$$B_t^H = \int_0^t K_H(t,s) \mathrm{d}B_s, \quad t \in [0,1]$$

is the fractional Brownian motion on $[0,1]$ and

$$K_H(t,s) = \frac{1}{\sqrt{V_H}\Gamma(H+\frac{1}{2})} (t-s)^{H-\frac{1}{2}} {}_1F_2\Big(H-\frac{1}{2}, \frac{1}{2}-H, H+\frac{1}{2}, 1-\frac{t}{s}\Big) 1_{[0,t)}(s),$$

with ${}_1F_2$ the Gauss hypergeometric function and

$$V_H = \frac{\Gamma(2-2H)\cos(\pi H)}{\pi H(1-2H)}. \tag{1.5}$$

This representation is due to Decreusefond and Üstünel [16]. See also Norros, Valkeila and Virtamo [42].

We will now give a brief review of the extensive literature on the fractional Brownian motion. This list is far from being complete but aims at giving relevant references for the current thesis. For general self-similar processes we refer to the extensive list of references in Taqqu [53] and for more recent work to Embrechts and Maejima [21]. For many applications of self-similar processes in economics and natural sciences see Mandelbrot [38]. As the increments of the fractional Brownian motion exhibits long memory for $H > 1/2$ new methods for parameter estimation have been developed. Many of the existing estimation methods for discrete time models related to the fractional Brownian motion and the fractional Gaussian noise are summarized in Beran [6]. In particular, Fox and Taqqu [26] have developed the idea of Whittle estimation for such processes. Results that have been extended by Dahlhaus [14]. For a semi-parametric approach to estimation of long memory see Robinson [45]. Models with long memory such as the fractional Gaussian noise may appear as the limit of aggregated dynamic models. This is shown in Granger [27]. Parameter estimation for continuous time models with long memory has also been studied in Comte [12] and Comte and Renault [13]. For applications of the fractional Brownian motion in hydrology we refer to Liu, Molz and Szulga [37]. For applications of the fractional Brownian motion in telecommunications and network traffic we refer Leland, Taqqu, Willinger and Wilson [35]. See also Mikosch,

Resnick, Rootzén and Stegeman [41] and the references therein. The use of wavelet analysis in network traffic and processes with long memory has been considered for instance in Abry, Flandrin, Taqqu and Veitch [1] and Abry and Veitch [2]. See also Flandrin [25] and Meyer, Sellan and Taqqu [40] for wavelet analysis and synthesis of the fractional Brownian motion. In economics the fractional Brownian motion was suggested as a model by Mandelbrot and Van Ness [39]. For further applications in finance see Shiryaev [51]. Recently, there has been rather extensive criticism against using the fractional Brownian motion to model asset returns since many of the natural models driven by the fractional Brownian motion admits arbitrage opportunities. See for instance Rogers [46] and Cheridito [10]. However, the fractional Brownian motion may also be used to model the volatility of asset prices, see Djehiche and Eddahbi [17]. The wide use of models driven by the fractional Brownian motion in applied fields has led researchers to develop a stochastic calculus for the fractional Brownian motion. For this reason a lot of efforts have been put into defining stochastic integrals with respect to the fractional Brownian motion. These question have been studied by Lin [36], Zähle [55], Decreusefond and Üstünel [16], Duncan, Hu and Pasik-Duncan [18], Pipiras and Taqqu [43], Alós, Mazét and Nualart [3] and many others.

Next, we will give an outline of the second part of the thesis and highlight the main results. In Chapter 5 we study Gaussian processes admitting representation as a Volterra type stochastic integral with respect to the standard Brownian motion. That is, processes $X = \{X_t : t \in T\}$, where T is an index set, admitting a representation of the form

$$X_t = \int_T V(t,r)\mathrm{d}B_r, \quad t \in T, \tag{1.6}$$

where $\{B_t : t \in T\}$ is a standard Brownian motion. Our primary example is the fractional Brownian motion but other interesting processes also fall within this more general definition. Chapter 5 has two main purposes. The first is to study Mercer-type representations of the Volterra process $\{X_t : t \in T\}$, i.e. representations of the form

$$X_t = \sum_{j,k} \Psi_{j,k}(t)\xi_{j,k}, \tag{1.7}$$

where $\{\Psi_{j,k}\}$ is a basis in a function space and the coefficients $\{\xi_{j,k}\}$ is a sequence of Gaussian random variables. Ideally we want the coefficients to be independent standard normal random variables. This is achieved if $\{\Psi_{j,k}\}$ is an orthonormal basis in the reproducing kernel Hilbert space associated with the process X. This representation have been known for a long time for Gaussian processes but to our knowledge it has received very little attention the literature on the fractional Brownian motion. The first four sections of Chapter 5 illustrates how this representation can be used in the context of fractional Brownian motion.

In the last section we study a Girsanov transformation with respect to the fractional Brownian motion and derive a maximum likelihood estimator of the drift parameter θ in models of the type

$$Y(t,\omega) = \theta A(t,\omega) + X(t,\omega),$$

where $A(\cdot)$ is a sufficiently smooth drift function. Here it is assumed that the process Y is observed continuously. Our results extend previous work by Norros, Valkeila and Virtamo [42], who study the case where $A(t) = t$ and X is the fractional Brownian motion, and work by Kleptsyna and Le Breton [33] who study the fractional Ornstein-Uhlenbeck process.

The main results of Chapter 5 include Corollary 5.17, Proposition 5.29, Proposition 5.34, Theorem 5.31, Theorem 5.32 and Theorem 5.33. In Corollary 5.17 we prove that the representation (1.7) holds almost surely uniformly on compacts if X is an almost surely continuous Gaussian process and $\{\Psi_{j,k}\}$ is an orthonormal base in the associated reproducing kernel Hilbert space. This result was suggested by Janson [30]. In Proposition 5.29 we derive a spectral representation of Skorohod integrals with respect to Volterra type processes. In Proposition 5.34 we derive a formula for computing the coefficients $\{\xi_{j,k}\}$ in the representation (1.7). In Theorem 5.31 we derive a Girsanov transformation for Volterra type processes. Finally, in Theorems 5.32 and 5.33 we derive consistency and asymptotic normality of estimators of the drift parameter in models driven by a Volterra type process.

Although the estimation techniques studied in the last section of Chapter 5 is interesting from a theoretical point of view the estimators seem rather difficult to implement in practice. In particular the derived estimators assume that we have continuous observations of the process, whereas in most applications only discrete observations are at hand. In Chapter 6 we have chosen to study a particular model, namely the fractional Ornstein-Uhlenbeck process, where we consider the joint estimation of all parameters in this model based on discrete observations. The fractional Ornstein-Uhlenbeck process is the stationary solution to the equation

$$X_t - X_s = -\theta \int_s^t X_u \mathrm{d}u + \sigma(B_t^H - B_s^H), \quad s < t,$$

where $\sigma > 0$, $\theta > 0$ and $H \in (0,1)$ and $\{B_t^H : t \in \mathbb{R}\}$ is the fractional Brownian motion with Hurst index H. The increments of the fractional Brownian motion has long memory if $H > 1/2$ and this property is transfered to the fractional Ornstein-Uhlenbeck process. We found it natural and simple to estimate the unknown parameters using an idea introduced in Whittle [54] and further developed for Gaussian sequences with long memory in Fox and Taqqu [26] and Dahlhaus [14].

The main findings in Chapter 6 include the computation of the spectral density and the covariance function of the fractional Ornstein-Uhlenbeck process in Proposition 6.2. Moreover, in Theorem 6.3 we prove consistency and asymptotic normality of the Whittle estimator for the joint estimation of all parameters in the fractional Ornstein-Uhlenbeck process.

References

[1] Abry, P., Flandrin, P., Taqqu, M. and Veitch, D. (2000) Wavelets for the analysis, estimation and synthesis of scaling data, in: K. Park and W. Willinger eds., *Self-Similar Network Traffic and Performance Evaluation*, Wiley, New York, 39–88.

[2] Abry, P. and Veitch, D. (1998) Wavelet analysis of long range dependent traffic, *IEEE Trans. Inform. Theory*, Vol. 44 No. 1, 2–15.

[3] Alòs, E., Mazet, O. and Nualart, D. (2000) Stochastic calculus with respect to Gaussian processes, *Ann. Probab.*, Vol. 29 No. 2, 766–801.

[4] Basrak, B. (2000) *The Sample Autocorrelation Function of Non-Linear Time Series*, Ph.D. Thesis, University of Groningen, Department of Mathematics.

[5] Basrak, B., Davis, R.A. and Mikosch, T. (2002) A Characterization of Multivariate Regular Variation, *Ann. Appl. Probab.*, Vol. 12, 908–920.

[6] Beran, J. (1994) *Statistics for long-memory processes*, Chapman & Hall, New York.

[7] Bingham, N.H., Goldie, C.M. and Teugels, J.L. (1987) *Regular variation*, Encyclopedia of mathematics and its applications 27, Cambridge University Press.

[8] Braverman, M., Mikosch, T. and Samorodnitsky, G. (2002) Tail probabilities of subadditive functionals acting on Lévy processes, *Ann. Appl. Probab.*, Vol. 12, 69–100.

[9] Cambanis, S., Huang, S. and Simons, G. (1981) On the theory of elliptically contoured distributions, *J. Multivariate Anal.*, Vol. 11, 368–385.

[10] Cheridito, P. (2002) Regularizing fractional Brownian motion with a view towards stock price modelling, Diss. ETH No. 14051, Ph.D. Thesis, ETH, Zürich.

[11] Cline, D.B.H. and Resnick, S.I. (1992) Multivariate subexponential distributions, *Stochastic Process. Appl.*, Vol. 42, 49–72.

[12] Comte, F. (1996) Simulation and estimation of long memory continuous time models, *J. Time Ser. Anal.*, Vol. 17 No. 1, 19–36.

[13] Comte, F. and Renault, E. (1996) Long memory continuous time models, *J. Econometrics*, Vol. 73, 101–149.

[14] Dahlhaus, R. (1989) Efficient parameter estimation for self-similar processes, *Ann. Statist.*, Vol. 17 No. 4, 1749–1766.

[15] Daley, D.J. and Vere-Jones, D. (1988) *An Introduction to the Theory of Point Processes*, Springer Verlag, New York.

[16] Decreusefond, L. and Üstünel, A.S. (1999) Stochastic Analysis of the Fractional Brownian Motion, *Potential Anal.*, Vol. 10, 177–214.

[17] Djehiche, B. and Eddahbi, M. (2001) Hedging options in market models modulated by fractional Brownian motion, *Stochastic Anal. Appl.*, Vol. 19 No. 5, 753–770.

[18] Duncan, T.E., Hu, Y. and Pasik-Duncan, B. (2000) Stochastic calculus for fractional Brownian motion I Theory, *SIAM J. Control Optim.*, Vol. 38 No. 2, 582–612.

[19] Embrechts, P., Goldie, C.M. and Veraverbeke, N. (1979) Subexponentiality and Infinite Divisibility, *Z. Wahrsch. Verw. Gebiete*, Vol. 49, 335–347.

[20] Embrechts, P., Klüppelberg C. and Mikosch, T. (1997) *Modelling Extremal Events for Insurance and Finance*, Springer, Berlin.

[21] Embrechts, P. and Maejima, M. (2002) *Self-similar processes*, Princeton University Press.

[22] Embrechts, P., McNeil A. and Straumann D. (2002) Correlation and Dependence in Risk Management: Properties and Pitfalls, in: M.A.H. Dempster ed., *Risk Management: Value at Risk and Beyond*, Cambridge University Press, Cambridge, 176–223.

[23] Fang, K.-T., Kotz, S. and Ng, K.-W. (1987) *Symmetric Multivariate and Related Distributions*, Chapman & Hall, London.

[24] Feller, W. (1972) *An introduction to probability theory and its applications*, Vol. 2, 2nd edition, Wiley, New York.

[25] Flandrin, P. (1992) Wavelet analysis and synthesis of fractional Brownian motion, *IEEE Trans. Inform. Theory*, IT-38 No. 2, 910–917.

[26] Fox, R. and Taqqu, M. (1986) Large-sample properties of parameter estimates for strongly dependent stationary Gaussian time series, *Ann. Statist.*, Vol. 14 No. 2, 517–532.

[27] Granger, C.W.J. (1980) Long memory relationships and the aggregation of dynamic models, *J. Econometrics*, Vol. 14, 227–238.

[28] de Haan, L. (1970) *On regular variation and its applications to the weak convergence of sample extremes*, Mathematical Centre Tract 32, Mathematics Centre, Amsterdam.

[29] de Haan, L. and de Ronde, J. (1998) Sea and Wind: Multivariate extremes at work, *Extremes*, Vol. 1 No. 1, 7–45.

[30] Janson, S. (1997) *Gaussian Hilbert Spaces*, Cambridge University Press.

[31] Kallenberg, O. (1983) *Random Measures*, 3rd edition, Akademie Verlag, Berlin.

[32] Kesten, H. (1973) Random difference equations and renewal theory for products of random matrices, *Acta Math.*, Vol. 131, 207–248.

[33] Kleptsyna M.L. and Le Breton, A. (2001) Statistical analysis of the fractional Ornstein-Uhlenbeck type process, *Stat. Inference Stoch. Process.*, Vol. 5 No. 3, 229–248.

[34] Kolmogorov, A.N. (1940) Wienerische spiralen und einige andere interessante Kurven in Hilbertschen raum, *C.R. (Doklady) Acad. Sci. USSR (N.S.)*, Vol. 26, 115–118.

[35] Leland, W.E., Taqqu, M.S., Willinger, W. and Wilson, D.V. (1994) On the self-similar nature of Ethernet traffic (Extended version), *IEEE/ACM Trans. Networking*, Vol. 2, 1–15.

[36] Lin, S.J. (1995) Stochastic analysis of fractional Brownian motions, *Stochastics Stochastics Rep.*, Vol. 55, 121–140.

[37] Liu, H.H., Molz, F.J. and Szulga, J. (1997) Fractional Brownian motion and fractional Gaussian noise in subsurface hydrology: A review, presentation of fundamental properties, and extensions, *Water Resour. Res.* , Vol. 33 No. 10, p. 2273 (97WR01982).

[38] Mandelbrot, B.B. (1982) *The fractal geometry of nature*, San Francisco: W.H. Freeman.

[39] Mandelbrot, B.B. and Van Ness, J.W. (1968) Fractional Brownian motions, fractional noises and applications, *SIAM Review*, Vol. 10, 422–437.

[40] Meyer, Y., Sellan, F. and Taqqu, M. (1999) Wavelets, Generalized White Noise and Fractional Integration: The Synthesis of Fractional Brownian Motion, *J. Fourier Anal. Appl.*, Vol. 5 No. 5, 465–494.

[41] Mikosch, T., Resnick, S., Rootzén, H. and Stegeman, A. (2002) Is network traffic approximated by stable Lévy motion or fractional Brownian motion? *Ann. Appl. Probab.*, Vol. 12 No. 1, 23–60.

[42] Norros, I., Valkeila, E. and Virtamo, J. (1999) An elementary approach to a Girsanov formula and other analytical results on fractional Brownian motions, *Bernoulli*, Vol. 5 No. 4, 571–588.

[43] Pipiras, V. and Taqqu, M. (2000) Integration questions related to fractional Brownian motion, *Probab. Theory Related Fields*, Vol. 118, 251–291.

[44] Resnick, S.I. (1987) *Extreme Values, Regular Variation, and Point Processes*, Springer Verlag, New York.

[45] Robinson, P.M. (1995) Gaussian semiparametric estimation of long-range dependence, *Ann. Statist.*, Vol. 5, 1630–1661.

[46] Rogers, L.C.G. (1997) Arbitrage from fractional Brownian motion, *Math. Finance* , Vol. 7, 95–105.

[47] Rootzén, H. and Tajvidi, N. (1997) Extreme value statistics and wind storm losses: a case study, *Scand. Actuar. J.* , Vol. 97 No. 1, 70–94.

[48] Rosiński, J. and Samorodnitsky, G. (1993) Distributions of subadditive functionals of sample paths of infinitely divisible processes, *Ann. Probab.*, Vol. 21, 996–1014.

[49] Rvačeva, E.L. (1962) On domains of attraction of multi-dimensional distributions, *Select. Transl. Math., Statist. and Probability*, American Mathematical Society, Providence, R.I. Vol. 2, 183–205.

[50] Samorodnitsky, G. and Taqqu, M. (1994) *Stable Non-Gaussian Random Processes*, Chapman and Hall.

[51] Shiryaev, A.N. (1999) *Essentials of stochastic finance. Facts, Models, Theory*, World Scientific, Singapore.

[52] Stărică, C. (1999) Multivariate extremes for models with constant conditional correlations, *J. Empirical Finance*, Vol. 6, 515–553 and *Extremes and Integrated Risk Management,* in P. Embrechts ed., Risk Books, (2000).

[53] Taqqu, M.S. (1986) A bibliographical guide to self-similar processes and long-range dependence, in: Eberlein, E. and Taqqu, M.S. eds., *Dependence in probability and statistics*, Birkhuser, Boston, 137–164.

[54] Whittle, P. (1951) *Hypothesis testing in time series analysis*, Hafner, New York.

[55] Zähle, M. (1998) Integration with respect to fractal functions and stochastic calculus I, *Probab. Theory Related Fields*, Vol. 111, 333–374.

Part I

Regular variation and stochastic processes

Chapter 2

Multivariate extremes, aggregation and dependence in elliptical distributions

Henrik Hult and Filip Lindskog (2002) Multivariate extremes, aggregation and dependence in elliptical distribution, *Advances in Applied Probability*, Vol. 34 No. 3, 587–609.

Abstract. In this paper we clarify dependence properties of elliptical distributions by deriving general but explicit formulas for the coefficients of upper and lower tail dependence and spectral measures with respect to different norms. We show that an elliptically distributed random vector is regularly varying if and only if the bivariate marginal distributions have tail dependence. Furthermore, the tail dependence coefficients are fully determined by the tail index of the random vector (or equivalently of its components) and the linear correlation coefficient. Whereas Kendall's tau is invariant in the class of elliptical distributions with continuous marginals and a fixed dispersion matrix, we show that this is not true for Spearman's rho. We also show that sums of elliptically distributed random vectors with the same dispersion matrix (up to a positive constant factor) remain elliptical if they are dependent only through their radial parts.

2000 *Mathematics Subject Classification.* 60E05 (primary); 62H20 (secondary).
Keywords and phrases. Elliptical distributions; Multivariate extremes; Regular variation; Tail dependence; Kendall's tau; Spearman's rho.
Acknowledgments. The authors want to thank Boualem Djehiche, Paul Embrechts and Uwe Schmock for comments on the manuscript.

2.1 Introduction

The class of elliptical distributions provides a rich source of multivariate distributions which share many of the tractable properties of the multivariate normal distribution and enables modeling of multivariate extremes and other forms of non-normal dependences.

In this paper we aim to clarify dependence properties of elliptical distributions. Dependence between the components of a random vector is, of course, related to the shape of the joint distribution. For elliptically distributed random vectors the shape of the distribution is given by the dispersion matrix Σ and the radial random variable R (see Theorem 2.11). The simple structure of elliptical distributions enables explicit computations of interesting quantities such as the coefficients of tail dependence and spectral measures associated with regularly varying random vectors (see Definition 2.7). From our explicit formula for the coefficient of tail dependence we conclude that it is fully determined by the corresponding linear correlation coefficient (as defined in Definition 2.16) and the tail index of the radial random variable in the general representation (see Theorem 2.11) of elliptically distributed random vectors. For this class of multivariate distributions, regular variation and tail dependence are closely related. Existence of tail dependence of the bivariate marginals and of regular variation is equated to regular variation of the radial random variable in the general representation.

Standard estimators of the linear correlation coefficient for elliptical distributions are based on the assumption of finite second moments. Kendall's tau and Spearman's rho (and their sample versions) do not rely on the existence of certain moments. It has been proved in Lindskog, McNeil and Schmock [8] that Kendall's tau is invariant in the class of elliptical distributions with continuous univariate marginals and a fixed dispersion matrix (up to a positive constant factor). This implies that the robust estimator of Kendall's tau can be used to estimate linear correlation coefficients without any other assumption on the underlying distribution than that of continuity of the univariate margins and joint ellipticality. One might also expect Spearman's rho to be invariant in the class of elliptical distributions with continuous marginals and a fixed dispersion matrix. We give a counterexample showing that this is not true.

It is known that sums of independent elliptically distributed random vectors with the same dispersion matrix are elliptical. In Theorem 2.18 we prove that sums of elliptically distributed random vectors with the same dispersion matrix are elliptical if they are dependent only through their radial parts. This result has applications to multivariate time series. It should be noted that the dispersion matrices are allowed to differ by a positive constant factor, see Remark 2.12(b) for details.

In this paper we use the spectral measure to answer questions about dependence of extremes for regularly varying elliptically distributed random vectors. In doing so it is crucial to consider a spectral measure with respect to a norm which corresponds to the question one is trying to answer. We discuss and exemplify this in Section 2.5.

In particular, for a bivariate elliptically distributed random vector, we compute the spectral measure with respect to the Euclidean 2-norm and with respect to the max-norm.

The paper is organized as follows. In Section 2.2 we recall the definitions of various dependence concepts. Section 2.3 introduces elliptical distributions, in particular we give the general stochastic representation of elliptically distributed random vectors. This representation is fundamental for the subsequent analysis. Section 2.4 contains the main results and in Section 2.5 we discuss the interpretation of the spectral measure with respect to different norms. All proofs are given in Section 2.6.

2.2 Preliminaries

To begin with we recall the definitions of the concordance measures Kendall's tau and Spearman's rho.

Definition 2.1. *Kendall's tau* for the random vector $(X_1, X_2)^{\mathrm{T}}$ is defined as

$$\tau(X_1, X_2) \triangleq \mathbb{P}((X_1 - X_1')(X_2 - X_2') > 0) - \mathbb{P}((X_1 - X_1')(X_2 - X_2') < 0),$$

where $(X_1', X_2')^{\mathrm{T}}$ is an independent copy of $(X_1, X_2)^{\mathrm{T}}$.

Definition 2.2. *Spearman's rho* for the random vector $(X_1, X_2)^{\mathrm{T}}$ is defined as

$$\varrho_S(X_1, X_2) \triangleq 3\left(\mathbb{P}((X_1 - X_1')(X_2 - X_2'') > 0) - \mathbb{P}((X_1 - X_1')(X_2 - X_2'') < 0)\right),$$

where $(X_1', X_2')^{\mathrm{T}}$ and $(X_1'', X_2'')^{\mathrm{T}}$ are independent copies of $(X_1, X_2)^{\mathrm{T}}$.

An important property of Kendall's tau and Spearman's rho is that they are invariant under strictly increasing transformations of the underlying random variables. If $(X_1, X_2)^{\mathrm{T}}$ is a random vector with continuous univariate marginal distributions and T_1 and T_2 are strictly increasing transformations on the range of X_1 and X_2 respectively, then $\tau(T_1(X_1), T_2(X_2)) = \tau(X_1, X_2)$. The same property holds for Spearman's rho. Note that this implies that Kendall's tau and Spearman's rho do not depend on the (marginal) distributions of X_1 and X_2.

Next we introduce two measures of dependence of multivariate extremes. Perhaps the most commonly encountered measure of dependence of bivariate extremes is the coefficient of upper (lower) tail dependence.

Let F be a univariate distribution function. We define the generalized inverse of F as $F^{-1}(u) \triangleq \inf\{x \in \mathbb{R} \mid F(x) \geq u\}$ for all u in $(0, 1)$.

Definition 2.3. Let $(X_1, X_2)^{\mathrm{T}}$ be a random vector with marginal distribution functions F_1 and F_2. The *coefficient of upper tail dependence* of $(X_1, X_2)^{\mathrm{T}}$ is defined as

$$\lambda_U(X_1, X_2) \triangleq \lim_{u \nearrow 1} \mathbb{P}(X_2 > F_2^{-1}(u) \mid X_1 > F_1^{-1}(u)),$$

provided that the limit $\lambda_U \in [0, 1]$ exists. The *coefficient of lower tail dependence* is defined as

$$\lambda_L(X_1, X_2) \triangleq \lim_{u \searrow 0} \mathbb{P}(X_2 \leq F_2^{-1}(u) \mid X_1 \leq F_1^{-1}(u)),$$

provided that the limit $\lambda_L \in [0, 1]$ exists. If $\lambda_U > 0$ ($\lambda_L > 0$), then we say that $(X_1, X_2)^{\mathrm{T}}$ has upper (lower) tail dependence.

For a pair of random variables, upper (lower) tail dependence is a measure of *joint* extremes. That is, it measures the probability that one component is extremely large (small) given that the other one is extremely large (small), relative to the marginal distributions.

The second measure of dependence in multivariate extremes that we discuss in this paper is the spectral measure associated with a regularly varying random vector (see Definition 2.7 below). Let us first recall the definition of regular variation for a (univariate) random variable.

Definition 2.4. The non-negative random variable R is said to be *regularly varying at ∞ with index $\alpha > 0$* if for all $x > 0$,

$$\lim_{u \to \infty} \frac{\mathbb{P}(R > ux)}{\mathbb{P}(R > u)} = x^{-\alpha}.$$

Throughout the paper we use the shorter "regularly varying with index $\alpha > 0$" for "regularly varying at ∞ with index $\alpha > 0$".

To prepare for the definition of regular variation for random vectors, we recall the concept of vague convergence. Let $\mathcal{X}$ be a Polish space, i.e. there exists a metric ρ on $\mathcal{X}$ which makes $(\mathcal{X}, \rho)$ a separable and complete metric space. A set $B \subset \mathcal{X}$ is said to be relatively compact if its closure $\overline{B}$ is compact. Let $\mathcal{B}(\mathcal{X})$ be the Borel σ-algebra on $\mathcal{X}$, generated by ρ. A measure μ on $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$ is called a Radon measure if $\mu(B) < \infty$ for all relatively compact sets $B \in \mathcal{B}(\mathcal{X})$.

Definition 2.5. Let $\mu, \mu_1, \mu_2, \ldots$ be Radon measures on $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$. We say that $\{\mu_n\}_{n \in \mathbb{N}}$ *converges to μ vaguely*, written $\mu_n \xrightarrow{v} \mu$, if

$$\lim_{n \to \infty} \int_{\mathcal{X}} f(s) \mu_n(\mathrm{d}s) = \int_{\mathcal{X}} f(s) \mu(\mathrm{d}s)$$

for all continuous functions $f : \mathcal{X} \to \mathbb{R}_+$ with compact support.

A useful equivalent formulation of vague convergence is given in the following theorem.

Theorem 2.6. *Let $\mu, \mu_1, \mu_2, \ldots$ be Radon measures on $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$. Then the following statements are equivalent.*

(1) $\mu_n \xrightarrow{v} \mu$ *as* $n \to \infty$.

(2) $\lim_{n\to\infty} \mu_n(B) = \mu(B)$ *for all relatively compact* $B \in \mathcal{B}(\mathcal{X})$ *with* $\mu(\partial B) = 0$.

For a proof, see Kallenberg [7] p. 169. For further details about vague convergence we refer to Kallenberg [7] and Resnick [10].

We denote by $\mathbb{S}^{d-1}$ the unit hypersphere in $\mathbb{R}^d$ with respect to a norm $|\cdot|$, and by $\mathcal{B}(\mathbb{S}^{d-1})$ the Borel σ-algebra on $\mathbb{S}^{d-1}$. We write $\mathbb{S}_2^{d-1} = \{\mathbf{z} \in \mathbb{R}^d \,|\, \mathbf{z}^{\mathrm{T}}\mathbf{z} = 1\}$ for the unit hypersphere in $\mathbb{R}^d$ with respect to the Euclidean 2-norm $|\cdot|_2$.

Definition 2.7. The d-dimensional random vector $\mathbf{X}$ is said to be *regularly varying with index* $\alpha > 0$ if there exists a probability measure σ on $\mathbb{S}^{d-1}$ such that for all $x > 0$, as $u \to \infty$,

$$\frac{\mathbb{P}(|\mathbf{X}| > ux, \mathbf{X}/|\mathbf{X}| \in \cdot\,)}{\mathbb{P}(|\mathbf{X}| > u)} \xrightarrow{v} x^{-\alpha}\sigma(\cdot) \text{ on } \mathcal{B}(\mathbb{S}^{d-1}).$$

The probability measure σ is referred to as the *spectral measure* of $\mathbf{X}$ and α is referred to as the *tail index* of $\mathbf{X}$.

For more on multivariate regular variation we refer to Resnick [10] and Mikosch [9]. In the definition we did not specify the choice of norm $|\cdot|$. The reason for this is that whether a random vector is regularly varying or not does not depend on the choice of norm in Definition 2.7. This is stated in the following lemma which proof is given in Section 2.6.

Lemma 2.8. *Let $|\cdot|_A$ and $|\cdot|_B$ be two norms on $\mathbb{R}^d$ and let $\mathbf{X}$ be a d-dimensional random vector. Then $\mathbf{X}$ is regularly varying with index $\alpha > 0$ with respect to the norm $|\cdot|_A$ if and only if $\mathbf{X}$ is regularly varying with index $\alpha > 0$ with respect to the norm $|\cdot|_B$.*

It is clear that the corresponding spectral measures do not coincide for different norms, see Section 2.5 for explicit examples. When we want to emphasize the choice of norm, we say that σ is the spectral measure of $\mathbf{X}$ with respect to the norm $|\cdot|$.

The following result on the effect of adding a constant vector to a regularly varying random vector turns out to be useful in the study of regular variation properties of elliptical distributions. The proof is given in Section 2.6.

Lemma 2.9. *Let $\mathbf{X}$ be a d-dimensional regularly varying random vector with tail index $\alpha > 0$ and spectral measure σ with respect to the norm $|\cdot|$, and let $b \in \mathbb{R}^d$ be a constant vector. Then $\mathbf{X} + \mathbf{b}$ is regularly varying with the same tail index and the same spectral measure with respect to the norm $|\cdot|$.*

2.3 Elliptical distributions

The main topic of this paper is to understand various measures of dependence through elliptical distributions. In this section we introduce the class of elliptically distributed random vectors and give some of their properties. For further details about elliptical distributions we refer to Fang, Kotz and Ng [6] and Cambanis, Huang and Simons [3].

Definition 2.10. If $\mathbf{X}$ is a d-dimensional random vector and, for some vector $\mu \in \mathbb{R}^d$, some $d \times d$ non-negative definite symmetric matrix Σ and some function $\phi : [0, \infty) \to \mathbb{R}$, the characteristic function $\varphi_{\mathbf{X}-\mu}$ of $\mathbf{X} - \mu$ is of the form $\varphi_{\mathbf{X}-\mu}(\mathbf{t}) = \phi(\mathbf{t}^{\mathrm{T}}\Sigma\mathbf{t})$, we say that $\mathbf{X}$ has an *elliptical distribution* with parameters μ, Σ and ϕ, and we write $\mathbf{X} \sim E_d(\mu, \Sigma, \phi)$.

The function ϕ is referred to as the *characteristic generator* of $\mathbf{X}$. When $d = 1$, the class of elliptical distributions coincides with the class of one-dimensional symmetric distributions. For elliptically distributed random vectors, we have the following general representation theorem.

Theorem 2.11. $\mathbf{X} \sim E_d(\mu, \Sigma, \phi)$ *with* $\operatorname{rank}(\Sigma) = k$ *if and only if there exist a non-negative random variable* R *independent of* $\mathbf{U}$*, a* k*-dimensional random vector uniformly distributed on the unit hypersphere* $\mathbb{S}_2^{k-1} = \{\mathbf{z} \in \mathbb{R}^k \,|\, \mathbf{z}^{\mathrm{T}}\mathbf{z} = 1\}$*, and a* $d \times k$ *matrix* A *with* $AA^{\mathrm{T}} = \Sigma$*, such that*

$$\mathbf{X} \stackrel{\mathrm{d}}{=} \mu + RA\mathbf{U}. \tag{2.1}$$

For the proof of Theorem 2.11 and details about the relation between R and ϕ, see Fang, Kotz and Ng [6] or Cambanis, Huang and Simons [3].

Remark 2.12. (a) Note that the representation (2.1) is not unique: if $\mathcal{O}$ is an orthogonal $k \times k$ matrix, then (2.1) also holds with $A' \triangleq A\mathcal{O}$ and $\mathbf{U}' \triangleq \mathcal{O}^{\mathrm{T}}\mathbf{U}$.
(b) Note that elliptical distributions with different parameters can be equal: if $\mathbf{X} \sim E_d(\mu, \Sigma, \phi)$, then $\mathbf{X} \sim E_d(\mu, c\Sigma, \phi_c)$ for every $c > 0$, where $\phi_c(s) \triangleq \phi(s/c)$ for all $s \geq 0$.

Example 2.13. Classical examples of elliptical distributions are the multivariate normal and the multivariate t-distributions. Let $\mathbf{X} \stackrel{\mathrm{d}}{=} \mu + RA\mathbf{U} \sim E_d(\mu, \Sigma, \phi)$, where $\operatorname{rank}(\Sigma) = d$. Then $\mathbf{X}$ is normally distributed if and only if $R^2 \sim \chi^2_d$, and $\mathbf{X}$ is t-distributed with ν degrees of freedom if and only if $R^2/d \sim F(d, \nu)$, where $F(d, \nu)$ denotes an F-distribution with d and ν degrees of freedom.

If the elliptically distributed random vector $\mathbf{X}$ has finite second moments, then we can always find a representation such that $\operatorname{Cov}(\mathbf{X}) = \Sigma$. To see this we use Theorem 2.11 to obtain

$$\operatorname{Cov}(\mathbf{X}) = \operatorname{Cov}(\mu + RA\mathbf{U}) = A\,\mathbb{E}(R^2)\operatorname{Cov}(\mathbf{U})\,A^{\mathrm{T}},$$

i.e. $\operatorname{Cov}(\mathbf{X})$ exists if and only if $\mathbb{E}(R^2) < \infty$. To compute $\operatorname{Cov}(\mathbf{U})$ let $\mathbf{Y} \sim N_d(\mathbf{0}, I_d)$. Then $\mathbf{Y} \stackrel{\mathrm{d}}{=} |\mathbf{Y}|_2\mathbf{U}$, where $|\mathbf{Y}|_2$ and $\mathbf{U}$ are independent. Furthermore $|\mathbf{Y}|_2^2 \sim \chi^2_d$, so

$\mathbb{E}(|\mathbf{Y}|_2^2) = d$. Since $\mathrm{Cov}(\mathbf{Y}) = I_d$ we see that if $\mathbf{U}$ is uniformly distributed on the unit hypersphere in $\mathbb{R}^d$, then $\mathrm{Cov}(\mathbf{U}) = I_d/d$. Thus $\mathrm{Cov}(\mathbf{X}) = AA^{\mathrm{T}}\mathbb{E}(R^2)/d$. By choosing the characteristic generator $\phi^*(s) = \phi(s/c)$, where $c = \mathbb{E}(R^2)/d$, we get $\mathrm{Cov}(\mathbf{X}) = \Sigma$.

The following result provides the basis of most applications of elliptical distributions.

Lemma 2.14. *Let $\mathbf{X} \sim E_d(\mu, \Sigma, \phi)$, let B be a $q \times d$ matrix and let $b \in \mathbb{R}^q$. Then*

$$b + B\mathbf{X} \sim E_q(b + B\mu, B\Sigma B^{\mathrm{T}}, \phi).$$

Proof. By Theorem 2.11, $b + B\mathbf{X}$ has a stochastic representation

$$b + B\mathbf{X} \stackrel{\mathrm{d}}{=} b + B\mu + RBA\mathbf{U}$$

and the conclusion follows from Definition 2.10. □

If we partition $\mathbf{X}$, μ and Σ into

$$\mathbf{X} = \begin{pmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{pmatrix}, \quad \mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \quad \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix},$$

where $\mathbf{X}_1$ and μ_1 are $r \times 1$ vectors and Σ_{11} is a $r \times r$ matrix, then we have the following consequence of Lemma 2.14.

Corollary 2.15. *Let $\mathbf{X} \sim E_d(\mu, \Sigma, \phi)$. Then*

$$\mathbf{X}_1 \sim E_r(\mu_1, \Sigma_{11}, \phi), \quad \mathbf{X}_2 \sim E_{d-r}(\mu_2, \Sigma_{22}, \phi).$$

Hence, marginal distributions of elliptical distributions are elliptical and of the same type (with the same characteristic generator). Next we introduce the linear correlation coefficient for a pair of random variables with a joint elliptical distribution.

Definition 2.16. Let $\mathbf{X} \sim E_d(\mu, \Sigma, \phi)$. For $i, j \in \{1, \ldots, d\}$, if $\Sigma_{ii} > 0$ and $\Sigma_{jj} > 0$, then we call

$$\varrho_{ij} \triangleq \Sigma_{ij}/\sqrt{\Sigma_{ii}\Sigma_{jj}} \tag{2.2}$$

the *linear correlation coefficient* for X_i and X_j.

If $\mathrm{Var}(X_i), \mathrm{Var}(X_j) \in (0, \infty)$, then $\varrho_{ij} = \mathrm{Cov}(X_i, X_j)/\sqrt{\mathrm{Var}(X_i)\,\mathrm{Var}(X_j)}$. We want to emphasize that the linear correlation coefficient as defined by (2.2) is an extension of the usual definition in terms of variances and covariances. We want to interpret the linear correlation coefficient as a scalar measure of dependence and, as such, it should not rely on finiteness of certain moments. Clearly (2.2) only makes sense for elliptical distributions. On the other hand, linear correlation is not always a meaningful measure of dependence for non-elliptical distributions, whereas

Kendall's tau and Spearman's rho remain meaningful; see for example Embrechts, McNeil and Straumann [5] p. 25.

In this paper we primarily consider elliptically distributed random vectors having components with continuous distributions. It is therefore relevant to present necessary and sufficient conditions for the components of an elliptically distributed random vector to be continuous random variables. Throughout the paper we say that a random variable is continuous whenever it has a continuous distribution function. The proof of Lemma 2.17 is given in Section 2.6.

Lemma 2.17. *Let* $\mathbf{X} \sim E_d(\mu, \Sigma, \phi)$*, with* $\mathbb{P}(X_i = \mu_i) < 1$ *for all* $i \in \{1, \dots, d\}$ *and with representation* $\mathbf{X} \stackrel{\mathrm{d}}{=} \mu + RA\mathbf{U}$ *according to Theorem 2.11. If* $\operatorname{rank}(\Sigma) = 1$*, then* $X_1, \dots, X_d$ *are continuous random variables if and only if* R *is continuous. If* $\operatorname{rank}(\Sigma) \geq 2$*, then* $X_1, \dots, X_d$ *are continuous random variables if and only if* $\mathbb{P}(X_i = \mu_i) = 0$ *for all* i*, or equivalently, if and only if* $\mathbb{P}(R = 0) = 0$*.*

2.4 Main Results

The sum of two independent elliptical random vectors with the same dispersion matrix is elliptical. The next theorem shows that the sum of two dependent elliptical random vectors with the same dispersion matrix, which are dependent only through their radial parts, is also elliptical.

Theorem 2.18. *Let* R *and* $\widetilde{R}$ *be non-negative random variables and let* $\mathbf{X} \triangleq \mu + R\mathbf{Z} \sim E_d(\mu, \Sigma, \phi)$ *and* $\widetilde{\mathbf{X}} \triangleq \widetilde{\mu} + \widetilde{R}\widetilde{\mathbf{Z}} \sim E_d(\widetilde{\mu}, \Sigma, \widetilde{\phi})$*, where* $(R, \widetilde{R}), \mathbf{Z}, \widetilde{\mathbf{Z}}$ *are independent. Then* $\mathbf{X} + \widetilde{\mathbf{X}} \sim E_d(\mu + \widetilde{\mu}, \Sigma, \phi^*)$*. Moreover, if* R *and* $\widetilde{R}$ *are independent, then* $\phi^*(u) = \phi(u)\widetilde{\phi}(u)$*.*

For the expression of the characteristic generator, ϕ^*, we refer to the proof in Section 2.6. A natural application of Theorem 2.18 is in the context of multivariate time series.

Example 2.19. Let $\mathbf{X}_t = \sigma_t \mathbf{Z}_t$, $t \in \mathbb{Z}$, where the random vectors $\mathbf{Z}_t \sim E_d(\mathbf{0}, \Sigma, \phi_t)$ are mutually independent and independent of the non-negative (univariate) random variables σ_t for all t. The σ_t's are allowed to be dependent. Then for every $t \in \mathbb{Z}$, $\mathbf{X}_t$ is elliptically distributed with dispersion matrix Σ, and so are all partial sums $\mathbf{S}_n = \sum_{t=1}^{n} \mathbf{X}_t$.

The relations (given below) between Kendall's tau, Spearman's rho and the linear correlation coefficient are well known for bivariate normally distributed random vectors. As stated in the next theorem the relation between Kendall's tau and the linear correlation coefficient holds more generally for all elliptically distributed random vectors with continuous univariate marginals.

Theorem 2.20. *Let* $\mathbf{X} \sim E_d(\mu, \Sigma, \phi)$*, where for* $i, j \in \{1, \dots, d\}$*,* X_i *and* X_j *are continuous. Then,*

$$\tau(X_i, X_j) = \frac{2}{\pi} \arcsin \varrho_{ij}. \tag{2.3}$$

For a proof of an extended version, see Lindskog, McNeil and Schmock [8]. As a consequence we have the following well-known result for Spearman's rho, for which we give an easy proof in Section 2.6.

Corollary 2.21. *Let* $\mathbf{X} \sim N_d(\mu, \Sigma)$*, where for* $i, j \in \{1, \ldots, d\}$*,* $\Sigma_{ii}, \Sigma_{jj} > 0$*. Then*

$$\varrho_S(X_i, X_j) = \frac{6}{\pi} \arcsin(\varrho_{ij}/2). \tag{2.4}$$

In the light of Theorem 2.20 one might expect Spearman's rho to be invariant in the class of elliptical distributions with continuous univariate marginals and a fixed dispersion matrix. However, the counterexample below shows that this is *not* true.

Counterexample. Let $\mathbf{X} \sim N_2(\mu, \Sigma)$, where $\Sigma_{11}, \Sigma_{22} > 0$. According to Theorem 2.11, $\mathbf{X}$ has stochastic representation $\mathbf{X} \stackrel{\mathrm{d}}{=} \mu + RA\mathbf{U}$, where $R \sim \chi^2_2$. We construct a counterexample by deriving a relation between Spearman's rho and the linear correlation coefficient for the bivariate elliptically distributed random vector $\mathbf{W} \triangleq A\mathbf{U}$. The relation is given by

$$\varrho_S(W_1, W_2) = 3\left(\frac{\arcsin \varrho}{\pi}\right) - 4\left(\frac{\arcsin \varrho}{\pi}\right)^3,$$

where $\varrho = \Sigma_{12}/\sqrt{\Sigma_{11}\Sigma_{22}}$. This relation differs from the relation (2.4) between Spearman's rho and the linear correlation coefficient for a bivariate normal distribution. The difference $\varrho_S(X_1, X_2) - \varrho_S(W_1, W_2)$ as a function of the linear correlation coefficient ϱ is plotted in Figure 2.1. We see that the difference is small but clearly not zero. For more details on this counterexample we refer to Section 2.6. It should be noted that there are other choices of R (other than $R^2 \sim \chi^2_2$) for which the difference $\varrho_S(X_1, X_2) - \varrho_S(W_1, W_2)$ becomes much bigger.

In Section 2.2 we introduced two concepts for measuring dependence of multivariate extremes of random vectors, the coefficient of tail dependence and the spectral measure associated with a regularly varying random vector. In the next theorem we clarify the connection between these two concepts. We also derive an explicit expression for the coefficient of tail dependence for two random variables with a joint elliptical distribution.

Theorem 2.22. *Let* $\mathbf{X} \stackrel{\mathrm{d}}{=} \mu + RA\mathbf{U} \sim E_d(\mu, \Sigma, \phi)$*, with* $\Sigma_{ii} > 0$ *for* $i = 1, \ldots, d$*,* $|\varrho_{ij}| < 1$ *for all* $i \neq j$*, and where* μ*,* R*,* A *and* $\mathbf{U}$ *are as in Theorem 2.11. Then the following statements are equivalent.*

(1) *R is regularly varying with index $\alpha > 0$.*
(2) *$\mathbf{X}$ is regularly varying with index $\alpha > 0$.*
(3) *For all $i \neq j$, $(X_i, X_j)^{\mathrm{T}}$ has tail dependence.*

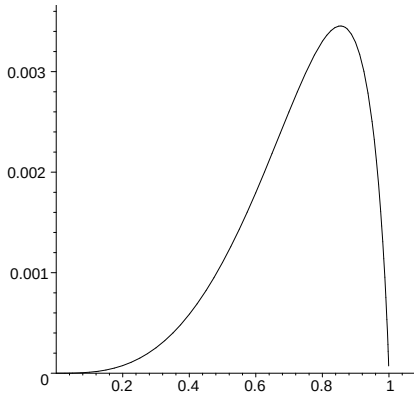

Figure 2.1. The difference between $(6/\pi)\arcsin(\varrho/2)$ and $(3/\pi)\arcsin\varrho - (4/\pi^3)(\arcsin\varrho)^3$ as a function of ϱ for $\varrho \in [0,1]$ (see the counterexample in Section 2.4).

Moreover, if R is regularly varying with index $\alpha > 0$, then for all $i \neq j$,

$$\lambda_U(X_i, X_j) = \lambda_L(X_i, X_j) = \frac{\int_{(\pi/2-\arcsin\varrho_{ij})/2}^{\pi/2} \cos^\alpha t dt}{\int_0^{\pi/2} \cos^\alpha t dt}.$$

Remark 2.23. Note that (1) and (2) are equivalent even if the condition $|\varrho_{ij}| < 1$ for all $i \neq j$ is not satisfied.

From the theorem above we can conclude that whether the bivariate marginals of an elliptically distributed vector $\mathbf{X}$ have tail dependence or not depends only on whether the radial random variable R in the representation $\mathbf{X} \stackrel{\mathrm{d}}{=} \mu + RA\mathbf{U}$ is regularly varying or not. The linear correlation coefficient ϱ_{ij} only effects the magnitude of the coefficient of tail dependence. An interesting consequence is that if $\mathbf{X} \sim E_d(\mu, \Sigma, \phi)$, then $(X_i, X_j)^{\mathrm{T}}$ can have a coefficient of tail dependence significantly larger than zero even if the linear correlation coefficient for $(X_i, X_j)^{\mathrm{T}}$ is zero or negative. In Figure 2.2 we have plotted the coefficient of tail dependence for an elliptically distributed bivariate random vector with uncorrelated components as a function of the tail index α.

2.5 Multivariate extremes for elliptical distributions

In this section we discuss how to interpret the spectral measure with respect to different norms. The discussion is general but in the case of elliptical distributions we can explicitly compute the spectral measure with respect to different norms and

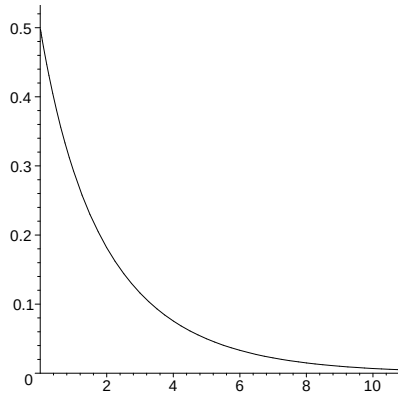

Figure 2.2. The coefficients of upper and lower tail dependence for regularly varying bivariate elliptical distributions with $(\Sigma_{11}, \Sigma_{22}, \varrho_{12}) = (1, 1, 0)$, as a function of the tail index α (see Remark 2.23).

compare different choices. Real data, e.g. financial asset price log returns, often indicate that the underlying distribution is elliptical or at least close to elliptical, and many statistical models are based on the assumption of ellipticality. Hence the following discussion should be relevant for many applications, especially in risk management.

By Lemma 2.8 we know that if a random vector $\mathbf{X}$ is regularly varying with respect to some norm on $\mathbb{R}^d$, then it is regularly varying with respect to every norm on $\mathbb{R}^d$. For every choice of the norm the spectral measure is a measure of dependence between extreme values. However, the choice of norm becomes essential when interpreting the spectral measure. The choice of norm must be related to the question we are trying to answer. A natural question would be: *What is the dependence between the components of a random vector given that at least one of its components is extreme?* In the literature (see e.g. Stărică [11]) most authors consider the Euclidean 2-norm, $|\cdot|_2$. However, if we want a measure of dependence between the components - given that at least one of the components is extreme - then we should use the max-norm $|\mathbf{X}|_\infty \triangleq \max\{|X_1|, \ldots, |X_d|\}$. Clearly, if we take $x = 1$ in Definition 2.7, we have that

$$\begin{aligned} \sigma_\infty(\cdot) &= \lim_{u\to\infty} \mathbb{P}(\mathbf{X}/|\mathbf{X}|_\infty \in \cdot \mid |\mathbf{X}|_\infty > u) \\ &= \lim_{u\to\infty} \mathbb{P}(\mathbf{X}/|\mathbf{X}|_\infty \in \cdot \mid |X_1| > u \cup \cdots \cup |X_d| > u), \end{aligned}$$

from which it is seen that the max-norm corresponds to the question posed. However, if the components are not identically distributed, then it might be more natural to condition on the events $\{|X_1| > G_1^{-1}(u) \cup \cdots \cup |X_d| > G_d^{-1}(u)\}$, where G_i is the distribution function of $|X_i|$ and $u \nearrow 1$. For $\mathbf{X} \sim E_d(\mathbf{0}, \Sigma, \phi)$ this is achieved by

considering the weighted max-norm $|\mathbf{X}|_{\infty,\Sigma} \triangleq \max\{|X_1|/\sqrt{\Sigma_{11}}, \ldots, |X_d|/\sqrt{\Sigma_{dd}}\}$, since in this case,

$$\begin{aligned}\sigma_{\infty,\Sigma}(\cdot) &= \lim_{z\to\infty} \mathbb{P}(\mathbf{X}/|\mathbf{X}|_{\infty,\Sigma} \in \cdot \mid |\mathbf{X}|_{\infty,\Sigma} > z) \\ &= \lim_{u\nearrow 1} \mathbb{P}(\mathbf{X}/|\mathbf{X}|_{\infty,\Sigma} \in \cdot \mid |X_1| > G_1^{-1}(u) \cup \cdots \cup |X_d| > G_d^{-1}(u)),\end{aligned}$$

is the spectral measure of $\mathbf{X}$ with respect to the norm $|\cdot|_{\infty,\Sigma}$. The corresponding question in this case would be: *What is the dependence between the components of a random vector given that at least one of its components is extreme relative to its marginal distribution?*

In the following two examples we compute the spectral measure with respect to the Euclidean 2-norm and the max-norm for bivariate regularly varying elliptical distributions. This can also be done for elliptical distributions of higher dimension, but the corresponding computations in spherical coordinates become quite tedious.

Example 2.24. Let $\mathbf{X} \sim E_2(\mathbf{0}, \Sigma, \phi)$, with $\Sigma_{11}, \Sigma_{22} > 0$, be regularly varying with index $\alpha > 0$, and let $\mathbf{X} \stackrel{\mathrm{d}}{=} RA\mathbf{U}$ be a stochastic representation according to Theorem 2.11. Without loss of generality we can choose A and $\mathbf{U}$ such that

$$\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \stackrel{\mathrm{d}}{=} R \begin{pmatrix} \sqrt{\Sigma_{11}} & 0 \\ \sqrt{\Sigma_{22}}\varrho_{12} & \sqrt{\Sigma_{22}}\sqrt{1-\varrho_{12}^2} \end{pmatrix} \begin{pmatrix} \cos\varphi \\ \sin\varphi \end{pmatrix},$$

where $\varphi \sim U(-\pi/2, 3\pi/2)$, i.e.

$$\begin{aligned}(X_1, X_2)^{\mathrm{T}} &\stackrel{\mathrm{d}}{=} (\sqrt{\Sigma_{11}} R\cos\varphi, \sqrt{\Sigma_{22}}\varrho_{12} R\cos\varphi + \sqrt{\Sigma_{22}}\sqrt{1-\varrho_{12}^2} R\sin\varphi)^{\mathrm{T}} \\ &= (\sqrt{\Sigma_{11}} R\cos\varphi, \sqrt{\Sigma_{22}} R\sin(\arcsin\varrho_{12} + \varphi))^{\mathrm{T}},\end{aligned}$$

where $\varphi \sim U(-\pi/2, 3\pi/2)$. Let

$$\begin{aligned}f(t) &\triangleq \left(\Sigma_{11}\cos^2 t + \Sigma_{22}\sin^2(\arcsin\varrho_{ij} + t)\right)^{1/2}, \\ g(t) &\triangleq \begin{cases} -\pi/2 & \text{for } t = -\pi/2, \\ \arctan\left(\sqrt{\frac{\Sigma_{22}}{\Sigma_{11}}}(\varrho_{12} + \sqrt{1-\varrho_{12}^2}\tan t)\right) & \text{for } t \in (-\pi/2, \pi/2), \\ g(t-\pi) + \pi & \text{for } t \in [\pi/2, 3\pi/2). \end{cases}\end{aligned}$$

Then,

$$RA\mathbf{U} = R|A\mathbf{U}|_2 \frac{A\mathbf{U}}{|A\mathbf{U}|_2} \stackrel{\mathrm{d}}{=} Rf(\varphi)\begin{pmatrix} \cos g(\varphi) \\ \sin g(\varphi) \end{pmatrix}.$$

Since $\mathbf{X}$ is regularly varying and $\mathbf{X}/|\mathbf{X}|_2$ has continuous distribution on $\mathbb{S}_2^1$, there exists a probability measure σ such that for every $x > 0$ and every $S \in \mathcal{B}(\mathbb{S}_2^1)$,

$$\lim_{z\to\infty} \frac{\mathbb{P}(R|A\mathbf{U}|_2 > zx, A\mathbf{U}/|A\mathbf{U}|_2 \in S)}{\mathbb{P}(R|A\mathbf{U}|_2 > z)} = x^{-\alpha}\sigma(S).$$

Moreover, by Theorem 2.20, R is regularly varying which implies that there exists a slowly varying function L, i.e., a positive, Lebesgue measurable function on $(0, \infty)$

satisfying $\lim_{t\to\infty} L(tx)/L(t) = 1$, for $x > 0$, such that $\mathbb{P}(R > x) = x^{-\alpha}L(x)$. Let $S_{\theta_1,\theta_2} = \{(\cos t, \sin t)^{\mathrm{T}} \mid \theta_1 < t < \theta_2\}$, where by symmetry we can assume that $-\pi/2 < \theta_1 < \theta_2 < \pi/2$. The case $|\varrho_{12}| = 1$ is trivial, so we consider only the case $|\varrho_{12}| < 1$. Then,

$$g^{-1}(t) = \arctan\left(\frac{1}{\sqrt{1-\varrho_{12}^2}}\left(\frac{\sqrt{\Sigma_{11}}}{\sqrt{\Sigma_{22}}}\tan t - \varrho_{12}\right)\right) \quad \text{for } -\pi/2 < t < \pi/2,$$

and

$$\begin{aligned}
\lim_{z\to\infty} \frac{\mathbb{P}(R|A\mathbf{U}|_2 > zx, A\mathbf{U}/|A\mathbf{U}|_2 \in S)}{\mathbb{P}(R|A\mathbf{U}|_2 > z)} &= \lim_{z\to\infty} \frac{\int_{g^{-1}(\theta_1)}^{g^{-1}(\theta_2)} z^{-\alpha}x^{-\alpha}f(t)^\alpha L(zx/f(t))\mathrm{d}t}{\int_0^{2\pi} z^{-\alpha}f(t)^\alpha L(z/f(t))\mathrm{d}t} \\
&= x^{-\alpha}\lim_{z\to\infty} \frac{\int_{g^{-1}(\theta_1)}^{g^{-1}(\theta_2)} f(t)^\alpha L(zx/f(t))/L(z)\mathrm{d}t}{\int_0^{2\pi} f(t)^\alpha L(z/f(t))/L(z)\mathrm{d}t} = x^{-\alpha}\frac{\int_{g^{-1}(\theta_1)}^{g^{-1}(\theta_2)} f(t)^\alpha \mathrm{d}t}{\int_0^{2\pi} f(t)^\alpha \mathrm{d}t} \\
&= x^{-\alpha}\frac{\int_{g^{-1}(\theta_1)}^{g^{-1}(\theta_2)} \left(\Sigma_{11}\cos^2 t + \Sigma_{22}\sin^2(\arcsin\varrho_{12}+t)\right)^{\alpha/2}\mathrm{d}t}{\int_0^{2\pi}\left(\Sigma_{11}\cos^2 t + \Sigma_{22}\sin^2(\arcsin\varrho_{12}+t)\right)^{\alpha/2}\mathrm{d}t}.
\end{aligned}$$

The third equality follows from the fact that $L(tx)/L(t) \to 1$ uniformly on intervals $[a, b]$, $0 < a \le b < \infty$ (see Theorem A3.2 in Embrechts, Klüppelberg and Mikosch [4] p. 566) and from the fact that there are constants $0 < c_1 < c_2 < \infty$ such that $c_1 < 1/f(t) < c_2$ for all $t \in [-\pi/2, \pi/2]$. Now we can identify the spectral measure as

$$\sigma(S_{\theta_1,\theta_2}) = \frac{\int_{g^{-1}(\theta_1)}^{g^{-1}(\theta_2)} \left(\Sigma_{11}\cos^2 t + \Sigma_{22}\sin^2(\arcsin\varrho_{12}+t)\right)^{\alpha/2}\mathrm{d}t}{\int_0^{2\pi}\left(\Sigma_{11}\cos^2 t + \Sigma_{22}\sin^2(\arcsin\varrho_{12}+t)\right)^{\alpha/2}\mathrm{d}t}.$$

Note that the spectral measure depends on the tail index α. Furthermore, note that $\lim_{\alpha\to 0}\sigma(S) = \mathbb{P}(A\mathbf{U}/|A\mathbf{U}|_2 \in S)$ for all $S \in \mathcal{B}(\mathbb{S}_2^{d-1})$. We see that the spectral measure is absolutely continuous and hence it has a density. The density is plotted in Figure 2.3 for bivariate regularly varying elliptical distributions with $(\Sigma_{11}, \Sigma_{22}, \varrho_{12}) = (1, 1, 0.5)$ and with tail indices $\alpha = 0, 2, 4, 8, 16$ (we write $\alpha = 0$ for the limit measure $\lim_{\alpha\to 0}\sigma(\cdot)$). From this figure it can be seen that as α increases - that is as the tails become lighter - the probability mass becomes *more concentrated* in the main directions of the ellipse (in this case $\pi/4$ and $5\pi/4$).

Example 2.25. Let us now compute the spectral measure with respect to the norm $|\cdot|_\infty$. Proceeding analogously to the previous example but replacing the function f by

$$\widetilde{f}(t) \triangleq \max\{\sqrt{\Sigma_{11}}\,|\cos t|, \sqrt{\Sigma_{22}}\,|\sin(\arcsin\varrho_{12}+t)|\,\},$$

we find that

$$RA\mathbf{U} = R|A\mathbf{U}|_\infty \frac{A\mathbf{U}}{|A\mathbf{U}|_\infty}$$

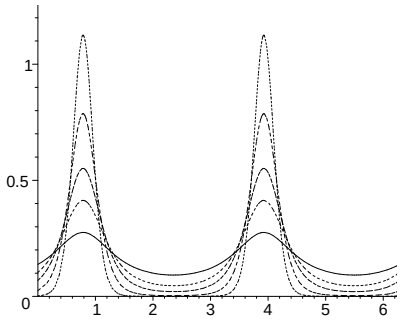

Figure 2.3. Densities of the spectral measure of $\mathbf{X} \sim E_2(\mu, \Sigma, \phi)$ with respect to the Euclidean 2-norm, where $(\Sigma_{11}, \Sigma_{22}, \varrho_{12}) = (1, 1, 0.5)$, and tail index $\alpha = 0, 2, 4, 8, 16$. Larger tail indices correspond to higher peaks (see Example 2.24).

where $A\mathbf{U}/|A\mathbf{U}|_\infty \stackrel{\mathrm{d}}{=} \widetilde{f}(\varphi)$ with $\varphi \sim U(-\pi/2, 3\pi/2)$. Following the computations in the previous example we obtain the spectral measure with respect to the max-norm as

$$\sigma_\infty(\widetilde{S}_{\theta_1,\theta_2}) = \frac{\int_{g^{-1}(\theta_1)}^{g^{-1}(\theta_2)} \left(\max\{\sqrt{\Sigma_{11}}\,|\cos t|, \sqrt{\Sigma_{22}}\,|\sin(\arcsin\varrho_{12} + t)|\,\}\right)^\alpha \mathrm{d}t}{\int_0^{2\pi} \left(\max\{\sqrt{\Sigma_{11}}\,|\cos t|, \sqrt{\Sigma_{22}}\,|\sin(\arcsin\varrho_{12} + t)|\,\}\right)^\alpha \mathrm{d}t},$$

where $\widetilde{S}_{\theta_1,\theta_2}$ is the radial projection of S_{θ_1,θ_2} (see Example 2.24) on $\mathbb{S}_\infty^{d-1}$, the the unit circle with respect to the max-norm. The density of the spectral measure is plotted in Figure 2.4 for bivariate regularly varying elliptical distributions with $(\Sigma_{11}, \Sigma_{22}, \varrho_{12}) = (1, 1, 0.5)$ and with tail indices $\alpha = 0, 2, 4, 8, 16$. From this figure it can be seen that as α increases - that is, as the tails become lighter - the probability mass becomes *less concentrated* in the main directions of the ellipse (in this case $\pi/4$ and $5\pi/4$). This is quite intuitive, for (bivariate) regularly varying elliptical distributions with lighter tails, the probability of joint extremes (that both components are extreme) becomes very small compared to the probability that one component is extreme. This can be seen from the fact that the coefficient of tail dependence tends to zero as the tail index increases (see Remark 2.23 and Figure 2.2).

Note the striking difference between the spectral measure with respect to the Euclidean norm and the spectral measure with respect to the max-norm. By choosing a norm which does not correspond to the question one is trying to answer, one might draw completely wrong conclusions about dependences between extremes. The best illustration of this is the comparison of Figure 2.3 with Figure 2.4.

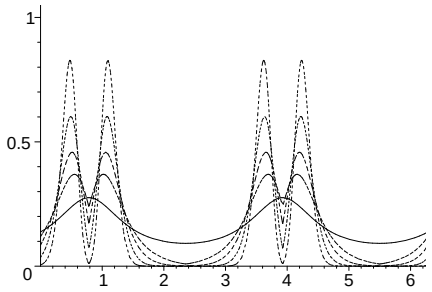

Figure 2.4. Densities of the spectral measure of $\mathbf{X} \sim E_2(\mu, \Sigma, \phi)$ with respect to the max-norm, where $(\Sigma_{11}, \Sigma_{22}, \varrho_{12}) = (1, 1, 0.5)$, and tail index $\alpha = 0, 2, 4, 8, 16$. Larger tail indices correspond to higher peaks (see Example 2.25).

2.6 Proofs

There exist several definitions of (multivariate) regular variation equivalent to Definition 2.7 (see e.g. Basrak [1]), one of which will be useful in the proof of Lemma 2.8 and Theorem 2.22. A statement similar to the following has been proved in [1], but we include a proof for completeness.

The following lemma gives an equivalent definition of (multivariate) regular variation in terms of vague convergence of measures on $\overline{\mathbb{R}}^d \backslash \{\mathbf{0}\}$, where $\overline{\mathbb{R}} \triangleq \mathbb{R} \cup \{-\infty, \infty\}$. There is a metric ρ on this space which makes it separable and complete and the σ-algebra $\mathcal{B}(\overline{\mathbb{R}}^d \backslash \{\mathbf{0}\})$, generated by ρ, coincides on $\mathbb{R}^d \backslash \{\mathbf{0}\}$ with the standard Borel σ-algebra $\mathcal{R}^d$ on $\mathbb{R}^d$, i.e. $\mathcal{B}(\overline{\mathbb{R}}^d \backslash \{\mathbf{0}\}) \cap (\mathbb{R}^d \backslash \{\mathbf{0}\}) = \mathcal{R}^d \cap (\mathbb{R}^d \backslash \{\mathbf{0}\})$. Finally, for $x > 0$ and $S \in \mathcal{B}(\mathbb{S}^{d-1})$, let $V_{x,S} \triangleq \{\mathbf{y} \in \mathbb{R}^d \backslash \{\mathbf{0}\} : |\mathbf{y}| > x, \mathbf{y}/|\mathbf{y}| \in S\}$.

Lemma 2.26. *Let $\mathbf{X}$ be a d-dimensional random vector. Then the following statements are equivalent.*

(1) $\mathbf{X}$ is regularly varying with index $\alpha > 0$ in the sense of Definition 2.7.

(2) There exists a non-zero Radon measure μ on $\mathcal{B}(\overline{\mathbb{R}}^d \backslash \{\mathbf{0}\})$ with $\mu(\overline{\mathbb{R}}^d \backslash \mathbb{R}^d) = 0$ and a relatively compact set $E \in \mathcal{B}(\overline{\mathbb{R}}^d \backslash \{\mathbf{0}\})$ with $\mu(\partial xE) = 0$ for every $x > 0$ such that, as $u \to \infty$,

$$\frac{\mathbb{P}(\mathbf{X} \in u\,\cdot\,)}{\mathbb{P}(\mathbf{X} \in uE)} \xrightarrow{v} \mu(\cdot) \text{ on } \mathcal{B}(\overline{\mathbb{R}}^d \backslash \{\mathbf{0}\}). \tag{2.5}$$

Moreover, if (2.5) holds, then there exists $\alpha > 0$ such that $\mu(xD) = x^{-\alpha}\mu(D)$ for every $x > 0$ and relatively compact $D \in \mathcal{B}(\overline{\mathbb{R}}^d \backslash \{0\})$ with $\mu(\partial D) = 0$.

Remark 2.27. If (2) holds, then

$$\widetilde{\mu}_u(\cdot) \triangleq \mathbb{P}(\mathbf{X} \in u\,\cdot\,)/\,\mathbb{P}(\mathbf{X} \in u\widetilde{E}) \xrightarrow{v} \widetilde{\mu}(\cdot) \text{ as } u \to \infty$$

holds with $\widetilde{\mu}(\cdot) = \mu(\cdot)/\mu(\widetilde{E})$ for any relatively compact set $\widetilde{E} \in \mathcal{B}(\overline{\mathbb{R}}^d \backslash \{\mathbf{0}\})$ such that $\mu(\widetilde{E}) > 0$ and $\mu(\partial\widetilde{E}) = 0$. This follows directly from the fact that

$$\frac{\mathbb{P}(X \in u\,\cdot\,)}{\mathbb{P}(X \in uE)} \frac{\mathbb{P}(X \in uE)}{\mathbb{P}(X \in u\widetilde{E})} \xrightarrow{v} \widetilde{\mu}(\cdot) \text{ as } u \to \infty$$

if the set $\widetilde{E}$ satisfies the above conditions.

Remark 2.28. If μ is a Radon measure on $\mathcal{B}(\overline{\mathbb{R}}^d \backslash \{\mathbf{0}\})$ such that for some $\alpha > 0$, $\mu(xD) = x^{-\alpha}\mu(D)$ for every $x > 0$ and relatively compact $D \in \mathcal{B}(\overline{\mathbb{R}}^d \backslash \{\mathbf{0}\})$ with $\mu(\partial D) = 0$, then it follows that μ assigns no mass to spheres (with respect to any norm on $\mathbb{R}^d$). To see this note that since μ is a Radon measure there exists a $y > 0$ such that $\mu(y\mathbb{S}^{d-1}) = 0$. Hence, by the scaling property we have $\mu(x\mathbb{S}^{d-1}) = (x/y)^{-\alpha}\mu(y\mathbb{S}^{d-1}) = 0$. Under the condition that $\mu(\overline{\mathbb{R}}^d \backslash \mathbb{R}^d) = 0$ we have $\mu(\partial V_{x,\mathbb{S}^{d-1}}) = 0$ for every $x > 0$. This implies in particular that $\mu(\partial V_{x,S}) = \mu(V_{x,\partial S})$ for every $x > 0$ and $S \in \mathcal{B}(\mathbb{S}^{d-1})$.

Proof. (2) $\Rightarrow$ (1) By Remark 2.28, $\mu(\partial V_{1,\mathbb{S}^{d-1}}) = 0$. Since $\mu(\overline{\mathbb{R}}^d \backslash \mathbb{R}^d) = 0$ we can without loss of generality take $E \subset \mathbb{R}^d$, and since E is relatively compact there exists a $x > 0$ such that $E \subset V_{x,\mathbb{S}^{d-1}}$, and $\mu(V_{1,\mathbb{S}^{d-1}}) = x^\alpha \mu(V_{x,\mathbb{S}^{d-1}}) > 0$ since $\mu(E) > 0$. Furthermore, $V_{1,\mathbb{S}^{d-1}}$ is relatively compact. Hence, by Remark 2.27, we may put $E = V_{1,\mathbb{S}^{d-1}}$. For any $x > 0$ and $S \in \mathcal{B}(\mathbb{S}^{d-1})$ with $\mu(\partial V_{x,S}) = 0$,

$$\lim_{u\to\infty} \mathbb{P}(|\mathbf{X}| > ux, \mathbf{X}/|\mathbf{X}| \in S)/\,\mathbb{P}(|\mathbf{X}| > u) = \lim_{u\to\infty} \mu_u(V_{x,S}) = \mu(V_{x,S}) = x^{-\alpha}\mu(V_{1,S}).$$

Since $\mu(V_{1,\cdot})$ is a probability measure on $\mathbb{S}^{d-1}$ we may put $\sigma(\cdot) = \mu(V_{1,\cdot})$. Moreover, by Remark 2.28, $\mu(\partial V_{x,S}) = \mu(V_{x,\partial S})$. Hence $\mathbf{X}$ is regularly varying with index $\alpha > 0$.

(1) $\Rightarrow$ (2) We will prove the implication with $E = V_{1,\mathbb{S}^{d-1}}$. For $u, x > 0$ define the measures

$$\begin{aligned} \mu_u(V_{x,\cdot}) &\triangleq \mathbb{P}(|\mathbf{X}| > ux, \mathbf{X}/|\mathbf{X}| \in \cdot)/\,\mathbb{P}(|\mathbf{X}| > u), \\ \mu(V_{x,\cdot}) &\triangleq x^{-\alpha}\sigma(\cdot) \end{aligned}$$

on $\mathcal{B}(\mathbb{S}^{d-1})$. This also defines set functions μ_u and μ on the semi-ring $\mathcal{P} \triangleq \{V_{x,S} : x > 0, S \in \mathcal{B}(\mathbb{S}^{d-1})\}$. By Theorem 11.3, p. 166 in Billingsley [2], μ_u and μ can be extended to measures on $\sigma(\mathcal{P}) = \mathcal{B}(\overline{\mathbb{R}}^d \backslash \{\mathbf{0}\}) \cap \mathbb{R}^d$. The extensions are unique since $\mathbb{R}^d \backslash \{\mathbf{0}\} = \cup_{k=1}^\infty V_{1/k,\mathbb{S}^{d-1}}$, where $\mu_u(V_{1/k,\mathbb{S}^{d-1}}) < \infty$ and $\mu(V_{1/k,\mathbb{S}^{d-1}}) < \infty$ for each $k \geq 1$. By definition of μ, $\mu(V_{x,S}) = x^{-\alpha}\mu(V_{1,S})$ for $x > 0$ and $S \in \mathcal{B}(\mathbb{S}^{d-1})$. Since

$\mathcal{P}$ is a π-system, this scaling property holds for arbitrary sets in $\sigma(\mathcal{P})$. We extend μ_u and μ to $\mathcal{B}(\overline{\mathbb{R}}^d\backslash\{\mathbf{0}\})$ by requiring that $\mu_u(\overline{\mathbb{R}}^d\backslash\mathbb{R}^d) = \mu(\overline{\mathbb{R}}^d\backslash\mathbb{R}^d) = 0$. Let

$$\mathcal{A} \triangleq \{V_{x,S}\backslash V_{y,S} : 0 < x \le y,\, S \in \mathcal{B}(\mathbb{S}^{d-1})\}.$$

Now $\mu_u(V_{x,\cdot}) \xrightarrow{v} \mu(V_{x,\cdot})$ is equivalent to that for each $S \in \mathcal{B}(\mathbb{S}^{d-1})$ with $\mu(V_{x,\partial S}) = 0$, $\mu_u(V_{x,S}) \to \mu(V_{x,S})$. By Remark 2.28, $\mu(\partial V_{x,S}) = \mu(V_{x,\partial S})$ for $x > 0$ and $S \in \mathcal{B}(\mathbb{S}^{d-1})$. Hence, for each set $V_{x,S}\backslash V_{y,S} \in \mathcal{A}$ with $\mu(\partial V_{x,S}) = 0$, as $u \to \infty$,

$$\mu_u(V_{x,S}\backslash V_{y,S}) = \mu_u(V_{x,S}) - \mu_u(V_{y,S}) \to \mu(V_{x,S}) - \mu(V_{y,S}) = \mu(V_{x,S}\backslash V_{y,S}).$$

Since $\mu(\overline{\mathbb{R}}^d\backslash\mathbb{R}^d) = 0$, convergence on $\mathcal{A}$ implies vague convergence on $\mathcal{B}(\overline{\mathbb{R}}^d\backslash\{0\})$. Hence $\mu_u \xrightarrow{v} \mu$ on $\mathcal{B}(\overline{\mathbb{R}}^d\backslash\{0\})$.
Let us also prove that if (2.5) holds, then there exists $\alpha > 0$ such that $\mu(xD) = x^{-\alpha}\mu(D)$ for every $x > 0$ and relatively compact $D \in \mathcal{B}(\overline{\mathbb{R}}^d\backslash\{0\})$ with $\mu(\partial D) = 0$. Indeed, for $x, y > 0$,

$$\begin{aligned}\mu(xyE) &= \lim_{u\to\infty} \frac{\mathbb{P}(X \in uxyE)}{\mathbb{P}(X \in uE)} \\ &= \lim_{u\to\infty} \frac{\mathbb{P}(X \in uxyE)}{\mathbb{P}(X \in uxE)} \frac{\mathbb{P}(X \in uxE)}{\mathbb{P}(X \in uE)} = \mu(yE)\mu(xE),\end{aligned}$$

i.e. $x \mapsto \mu(xE)$ satisfies the Hamel equation, and hence $\mu(xE) = x^{-\alpha}$ for some $\alpha \in \mathbb{R}$. Since μ is a Radon measure (μ assigns finite measure to relatively compact sets) we have $\alpha \ge 0$, and since μ does not charge $\overline{\mathbb{R}}^d\backslash\mathbb{R}^d$ we have $\alpha \ne 0$. For $x > 0$ and a relatively compact D with $\mu(\partial D) = 0$,

$$\mu(xD) = \lim_{u\to\infty} \frac{\mathbb{P}(X \in uxD)}{\mathbb{P}(X \in uE)} = \lim_{u\to\infty} \frac{\mathbb{P}(X \in uxD)}{\mathbb{P}(X \in uxE)} \frac{\mathbb{P}(X \in uxE)}{\mathbb{P}(X \in uE)} = \mu(D)x^{-\alpha}.$$

□

Proof of Lemma 2.8. Assume that $\mathbf{X}$ is regularly varying with index $\alpha > 0$ with respect to the norm $|\cdot|_A$. Then statement (2) of Lemma 2.26 holds for some μ. Proceeding as in the first part of the proof of Lemma 2.26 with the norm $|\cdot|_B$ proves the claim. □

Proof of Lemma 2.9. Let $x > 0$ be arbitrary but fixed and let $S \in \mathcal{B}(\mathbb{S}^{d-1})$ be arbitrary but fixed with $\sigma(\partial S) = 0$. For $u > 0$ let

$$\begin{aligned}A_u &\triangleq \frac{\mathbb{P}(|\mathbf{X}+\mathbf{b}| > ux, (\mathbf{X}+\mathbf{b})/|\mathbf{X}+\mathbf{b}| \in S)}{\mathbb{P}(|\mathbf{X}+\mathbf{b}| > u)}, \\ L_u &\triangleq \frac{\mathbb{P}(|\mathbf{X}| > ux + |\mathbf{b}|, (\mathbf{X}+\mathbf{b})/|\mathbf{X}+\mathbf{b}| \in S)}{\mathbb{P}(|\mathbf{X}| > u - |\mathbf{b}|)}, \\ U_u &\triangleq \frac{\mathbb{P}(|\mathbf{X}| > ux - |\mathbf{b}|, (\mathbf{X}+\mathbf{b})/|\mathbf{X}+\mathbf{b}| \in S)}{\mathbb{P}(|\mathbf{X}| > u + |\mathbf{b}|)}.\end{aligned}$$

Then

$$\begin{aligned} L_u &= \frac{\mathbb{P}(|\mathbf{X}| > ux + |\mathbf{b}|, (\mathbf{X}+\mathbf{b})/|\mathbf{X}+\mathbf{b}| \in S)}{\mathbb{P}(|\mathbf{X}| > u + |\mathbf{b}|/x)} \frac{\mathbb{P}(|\mathbf{X}| > u + |\mathbf{b}|/x)}{\mathbb{P}(|\mathbf{X}| > u - |\mathbf{b}|)} \triangleq L_u^{(1)} L_u^{(2)}, \\ U_u &= \frac{\mathbb{P}(|\mathbf{X}| > ux - |\mathbf{b}|, (\mathbf{X}+\mathbf{b})/|\mathbf{X}+\mathbf{b}| \in S)}{\mathbb{P}(|\mathbf{X}| > u - |\mathbf{b}|/x)} \frac{\mathbb{P}(|\mathbf{X}| > u - |\mathbf{b}|/x)}{\mathbb{P}(|\mathbf{X}| > u + |\mathbf{b}|)} \triangleq U_u^{(1)} U_u^{(2)}. \end{aligned}$$

Since $L_u \leq A_u \leq U_u$ for all $u > 0$, $L_u^{(2)}, U_u^{(2)} \to 1$ as $u \to \infty$ and $L_u^{(1)} = U_s^{(1)}$ with $s = u + 2|\mathbf{b}|/x$,

$$\begin{aligned} \lim_{u\to\infty} A_u &= \lim_{u\to\infty} L_u = \lim_{u\to\infty} U_u = \lim_{u\to\infty} \frac{\mathbb{P}(|\mathbf{X}| > ux, (\mathbf{X}+\mathbf{b})/|\mathbf{X}+\mathbf{b}| \in S)}{\mathbb{P}(|\mathbf{X}| > u)} \\ &= \lim_{u\to\infty} \frac{\mathbb{P}(|\mathbf{X}| > ux, \mathbf{X}/|\mathbf{X}| \in S)}{\mathbb{P}(|\mathbf{X}| > u)} \\ &\quad + \lim_{u\to\infty} \frac{\mathbb{P}(|\mathbf{X}| > ux, (\mathbf{X}+\mathbf{b})/|\mathbf{X}+\mathbf{b}| \in S, \mathbf{X}/|\mathbf{X}| \notin S)}{\mathbb{P}(|\mathbf{X}| > u)} \\ &\quad - \lim_{u\to\infty} \frac{\mathbb{P}(|\mathbf{X}| > ux, (\mathbf{X}+\mathbf{b})/|\mathbf{X}+\mathbf{b}| \notin S, \mathbf{X}/|\mathbf{X}| \in S)}{\mathbb{P}(|\mathbf{X}| > u)}. \end{aligned}$$

The second to last term can be written as

$$\begin{aligned} &\lim_{u\to\infty} \frac{\mathbb{P}(|\mathbf{X}| > ux, (\mathbf{X}+\mathbf{b})/|\mathbf{X}+\mathbf{b}| \in S, \mathbf{X}/|\mathbf{X}| \notin S)}{\mathbb{P}(|\mathbf{X}| > u)} \\ &= \lim_{u\to\infty} \frac{\mathbb{P}(|\mathbf{X}| > ux)}{\mathbb{P}(|\mathbf{X}| > u)} \mathbb{P}((\mathbf{X}+\mathbf{b})/|\mathbf{X}+\mathbf{b}| \in S, \mathbf{X}/|\mathbf{X}| \notin S \mid |\mathbf{X}| > ux) \\ &\leq \lim_{u\to\infty} \frac{\mathbb{P}(|\mathbf{X}| > ux)}{\mathbb{P}(|\mathbf{X}| > u)} \mathbb{P}(\mathbf{X}/|\mathbf{X}| \in \partial S \mid |\mathbf{X}| > ux) = x^{-\alpha}\sigma(\partial S) = 0, \end{aligned}$$

and similarly for the last term, from which the conclusion follows. □

Proof of Lemma 2.17. Let $\mathbf{X} \stackrel{\mathrm{d}}{=} \mu + RA\mathbf{U}$ be a stochastic representation according to Theorem 2.11. Suppose $\operatorname{rank}(\Sigma) = 1$, then A is a $d \times 1$ matrix and U is symmetric $\{1, -1\}$-valued. Furthermore, $\mathbb{P}(X_i = \mu_i) < 1$ implies $A_{i1} \neq 0$. Hence, if $\operatorname{rank}(\Sigma) = 1$, then $X_1, \ldots, X_d$ are continuous random variables if and only if R is continuous. Now suppose $\operatorname{rank}(\Sigma) = k \geq 2$. Define $\mathbf{A}_i \triangleq (A_{i1}, \ldots, A_{ik})$ and $a \triangleq \mathbf{A}_i \mathbf{A}_i^{\mathrm{T}}$. Since $\mathbb{P}(X_i = \mu_i) < 1$, the case $a = 0$ is excluded. By choosing an orthogonal $k \times k$ matrix $\mathcal{O}$ whose first column is $\mathbf{A}_i^{\mathrm{T}}/a$ and using Remark 2.12(a) if necessary, we may assume that $\mathbf{A}_i = (a, 0, \ldots, 0)$, hence $X_i \stackrel{\mathrm{d}}{=} \mu_i + aRU_1$. Note that U_1 is a continuous random variable because $k \geq 2$. Hence $\mathbb{P}(aRU_1 = x) = 0$ for all $x \in \mathbb{R}\backslash\{0\}$. Hence, if $\operatorname{rank}(\Sigma) \geq 2$, then $X_1, \ldots, X_d$ are continuous random variables if and only if $\mathbb{P}(X_i = \mu_i) = 0$ for $i = 1, \ldots, d$, or equivalently, if and only if $\mathbb{P}(R = 0) = 0$. □

Proof of Theorem 2.18. Let $\widetilde{\phi}^{(r)}$ be the characteristic generator of $(\widetilde{R} \mid R = r)\widetilde{\mathbf{Z}}$, let ϕ' be the characteristic generator of $\mathbf{Z}$, and let F_R be the distribution function of R. Then for all $t \in \mathbb{R}^d$,

$$\begin{aligned}\varphi_{R\mathbf{Z}+\widetilde{R}\widetilde{\mathbf{Z}}}(t) &= \int_0^\infty \varphi_{r\mathbf{Z}}(t)\varphi_{(\widetilde{R}|R=r)\widetilde{\mathbf{Z}}}(t)\mathrm{d}F_R(r) \\ &= \int_0^\infty \phi'(r^2 t^{\mathrm{T}}\Sigma t)\widetilde{\phi}^{(r)}(t^{\mathrm{T}}\Sigma t)\mathrm{d}F_R(r),\end{aligned}$$

from which it follows that $\mathbf{X} + \widetilde{\mathbf{X}} \sim E_d(\mu + \widetilde{\mu}, \Sigma, \phi^*)$, with

$$\phi^*(u) = \int_0^\infty \phi'(r^2 u)\widetilde{\phi}^{(r)}(u)\mathrm{d}F_R(r).$$

Moreover, if R and $\widetilde{R}$ are independent, then $\widetilde{\phi}^{(r)}(u) = \widetilde{\phi}(u)$ and

$$\phi^*(u) = \int_0^\infty \phi'(r^2 u)\widetilde{\phi}^{(r)}(u)\mathrm{d}F_R(r) = \widetilde{\phi}(u)\int_0^\infty \phi'(r^2 u)\mathrm{d}F_R(r) = \phi(u)\widetilde{\phi}(u).$$

□

Proof of Corollary 2.21. Recall that $\varrho_{ij} \triangleq \Sigma_{ij}/\sqrt{\Sigma_{ii}\Sigma_{jj}}$. Let $\widetilde{X}_i \overset{\mathrm{d}}{=} X_i$ for $i = 1, \ldots, d$ be mutually independent, and independent of $\mathbf{X}$. Then $\widetilde{\mathbf{X}} \sim N_d(\mu, \widetilde{\Sigma})$, where $\widetilde{\Sigma} = \mathrm{diag}(\Sigma_{11}, \ldots, \Sigma_{dd})$. Hence, $\mathbf{X}^* \triangleq \mathbf{X} - \widetilde{\mathbf{X}} \sim N_d(\mathbf{0}, \Sigma^*)$, where $\Sigma^* = \Sigma + \widetilde{\Sigma}$. Let $\varrho^*_{ij} \triangleq \Sigma^*_{ij}/\sqrt{\Sigma^*_{ii}\Sigma^*_{jj}}$. Then,

$$\varrho_S(X_i, X_j) = 3\tau(X_i^*, X_j^*) = 3\left(\frac{2}{\pi}\arcsin\varrho^*_{ij}\right) = \frac{6}{\pi}\arcsin(\varrho_{ij}/2),$$

where the second equality follows from Theorem 2.20 and the fact that the dispersion matrix of a sum of two independent identically distributed elliptical random vectors differs from those of the terms by at most a positive constant factor. □

Proof of Theorem 2.22. By Lemma 2.9 we can without loss of generality assume that $\mu = 0$. (1) $\iff$ (2) If $\mathrm{rank}(\Sigma) = k < d$, denote by $\Sigma^{(-1)} \triangleq (A^{(-1)})^{\mathrm{T}}A^{(-1)}$ the generalized inverse of Σ, where $A^{(-1)} \triangleq (A^{\mathrm{T}}A)^{-1}A^{\mathrm{T}}$, i.e. $A^{(-1)}$ solves $A^{(-1)}A = I_k$, where I_k denotes the $k \times k$ identity matrix. Note that $\Sigma^{(-1)} = \Sigma^{-1}$ if $\mathrm{rank}(\Sigma) = d$. By choosing the norm $|\mathbf{x}|_\Sigma \triangleq (\mathbf{x}^{\mathrm{T}}\Sigma^{(-1)}\mathbf{x})^{1/2}$ in the definition of regular variation (by Lemma 2.8 we are allowed to choose any norm), we obtain

$$\frac{\mathbb{P}(|\mathbf{X}|_\Sigma > ux, \mathbf{X}/|\mathbf{X}|_\Sigma \in \cdot)}{\mathbb{P}(|\mathbf{X}|_\Sigma > u)} = \frac{\mathbb{P}(R > ux, A\mathbf{U} \in \cdot)}{\mathbb{P}(R > u)} = \frac{\mathbb{P}(R > ux)\,\mathbb{P}(A\mathbf{U} \in \cdot)}{\mathbb{P}(R > u)}.$$

If R is regularly varying, then $\lim_{u\to\infty}\mathbb{P}(R > ux)/\mathbb{P}(R > u) = x^{-\alpha}$, and hence $\mathbb{P}(|\mathbf{X}|_\Sigma > ux, \mathbf{X}/|\mathbf{X}|_\Sigma \in \cdot)/\mathbb{P}(|\mathbf{X}|_\Sigma > u) \xrightarrow{v} x^{-\alpha}\,\mathbb{P}(A\mathbf{U} \in \cdot)$ as $u \to \infty$. Conversely,

if $\mathbb{P}(|\mathbf{X}|_\Sigma > ux, \mathbf{X}/|\mathbf{X}|_\Sigma \in \cdot)/\,\mathbb{P}(|\mathbf{X}|_\Sigma > u) \xrightarrow{v} x^{-\alpha}\sigma(\cdot)$ as $u \to \infty$, then we must have $\sigma(\cdot) = \mathbb{P}(A\mathbf{U} \in \cdot)$ and $\lim_{u\to\infty} \mathbb{P}(R > ux)/\,\mathbb{P}(R > u) = x^{-\alpha}$.

(1) $\iff$ (3) First note that if $\mathbf{X} \sim E_d(\mathbf{0}, \Sigma, \phi)$, with $X_i \sim F_i$ and $X_j \sim F_j$, then $F_i^{-1}(u) = \sqrt{\Sigma_{ii}/\Sigma_{jj}}F_j^{-1}(u)$ for $u \in (0,1)$. Secondly, if $\lim_{u\nearrow 1} F_i^{-1}(u) < \infty$, i.e. if X_i is a bounded random variable, then there exists a $u_0 \in (0,1)$ such that the events $\{X_i > F_i^{-1}(u)\}$ and $\{X_j > F_j^{-1}(u)\}$ are disjoint for $u > u_0$, and hence

$$\lim_{u\nearrow 1} \frac{\mathbb{P}(X_i > F_i^{-1}(u), X_j > F_j^{-1}(u))}{\mathbb{P}(X_i > F_i^{-1}(u))} = 0.$$

Therefore, without loss of generality, we only consider random vectors whose marginal distributions are such that $\lim_{u\nearrow 1} F_i^{-1}(u) = \infty$, i.e. random variables with unbounded support. Then,

$$\lambda_U(X_i, X_j) = \lim_{z\to\infty} \frac{\mathbb{P}(X_i > \sqrt{\Sigma_{ii}}z, X_j > \sqrt{\Sigma_{jj}}z)}{\mathbb{P}(X_i > \sqrt{\Sigma_{ii}}z)}.$$

Since $X \stackrel{\mathrm{d}}{=} RA\mathbf{U}$,

$$\begin{pmatrix} X_i \\ X_j \end{pmatrix} \stackrel{\mathrm{d}}{=} R \begin{pmatrix} \sqrt{\Sigma_{ii}} & 0 \\ \sqrt{\Sigma_{jj}}\varrho_{ij} & \sqrt{\Sigma_{jj}}\sqrt{1-\varrho_{ij}^2} \end{pmatrix} \begin{pmatrix} \cos\varphi \\ \sin\varphi \end{pmatrix},$$

where $\varphi \sim U(-\pi, \pi)$, i.e.

$$\begin{aligned} (X_i, X_j)^{\mathrm{T}} &\stackrel{\mathrm{d}}{=} (\sqrt{\Sigma_{ii}}R\cos\varphi, \sqrt{\Sigma_{jj}}\varrho_{ij}R\cos\varphi + \sqrt{\Sigma_{jj}}\sqrt{1-\varrho_{ij}^2}R\sin\varphi)^{\mathrm{T}} \\ &= (\sqrt{\Sigma_{ii}}R\cos\varphi, \sqrt{\Sigma_{jj}}R\sin(\arcsin\varrho_{ij} + \varphi))^{\mathrm{T}}, \end{aligned}$$

with $\varphi \sim U(-\pi, \pi)$. Hence

$$\lambda_U(X_i, X_j) = \lim_{z\to\infty} \frac{\mathbb{P}(R\cos\varphi > z, R\sin(\arcsin\varrho_{ij} + \varphi) > z)}{\mathbb{P}(R\cos\varphi > z)}.$$

The numerator can be written as

$$\begin{aligned} &\mathbb{P}(R\cos\varphi > z, R\sin(\arcsin\varrho_{ij} + \varphi) > z) \\ &= \mathbb{P}\left(R > z\max\Big(\frac{1}{\cos\varphi}, \frac{1}{\sin(\arcsin\varrho_{ij} + \varphi)}\Big), \cos\varphi > 0, \sin(\arcsin\varrho_{ij} + \varphi) > 0\right) \\ &= \frac{1}{\pi}\int_{(\pi/2 - \arcsin\varrho_{ij})/2}^{\pi/2} \mathbb{P}(R > z/\cos t)\mathrm{d}t, \end{aligned}$$

and the denominator can be written as

$$\mathbb{P}(R\cos\varphi > z) = \mathbb{P}(R > z/\cos\varphi, \cos\varphi > 0) = \frac{1}{\pi}\int_0^{\pi/2} \mathbb{P}(R > z/\cos t)\mathrm{d}t.$$

Suppose R is regularly varying with index $\alpha > 0$. By Theorem A.3.2 in Embrechts, Klüppelberg and Mikosch [4] p. 566, for $a > 0$,

$$\mathbb{P}(R > zx)/\,\mathbb{P}(R > z) \to x^{-\alpha} \text{ as } z \to \infty,$$

uniformly in x on each $[a, \infty)$. In particular, with $x = 1/\cos t$,

$$\mathbb{P}(R > z/\cos t)/\,\mathbb{P}(R > z) \to \cos^{\alpha} t \text{ as } z \to \infty,$$

uniformly in t on $[0, \pi/2)$. Let

$$f_z(t) \triangleq \begin{cases} \mathbb{P}(R > z/\cos t)/\,\mathbb{P}(R > z) & \text{for } t \in [0, \pi/2), \\ 0 & \text{for } t = \pi/2. \end{cases}$$

Then $f_z(\cdot) \to \cos^{\alpha}(\cdot)$ uniformly on $[0, \pi/2]$, and hence

$$\lambda_U(X_i, X_j) = \lim_{z \to \infty} \frac{\int_{(\pi/2 - \arcsin \varrho_{ij})/2}^{\pi/2} f_z(t)\mathrm{d}t}{\int_0^{\pi/2} f_z(t)\mathrm{d}t} = \frac{\int_{(\pi/2 - \arcsin \varrho_{ij})/2}^{\pi/2} \cos^{\alpha} t\mathrm{d}t}{\int_0^{\pi/2} \cos^{\alpha} t\mathrm{d}t}.$$

Since

$$\lim_{\alpha \to \infty} \frac{\int_{(\pi/2 - \arcsin \varrho_{ij})/2}^{\pi/2} \cos^{\alpha} t\mathrm{d}t}{\int_0^{\pi/2} \cos^{\alpha} t\mathrm{d}t} = 0,$$

we conclude that $\lambda_U(X_i, X_j) > 0$ if and only if R is regularly varying with index $\alpha > 0$ and in that case

$$\lambda_U(X_i, X_j) = \frac{\int_{(\pi/2 - \arcsin \varrho_{ij})/2}^{\pi/2} \cos^{\alpha} t\mathrm{d}t}{\int_0^{\pi/2} \cos^{\alpha} t\mathrm{d}t}.$$

Moreover, because elliptically distributed random vectors are radially symmetric about μ, $\lambda_U(X_i, X_j) = \lambda_L(X_i, X_j)$. □

We close this section with a more detailed version of the counterexample already discussed in Section 2.4.

Counterexample. Let $\mathbf{X} \overset{\mathrm{d}}{=} \mu + RA\mathbf{U} \sim E_2(\mu, \Sigma, \phi)$, where $\Sigma_{11}, \Sigma_{22} > 0$ and μ, R, A and $\mathbf{U}$ are as in Theorem 2.11. To construct a counterexample we derive the relation between Spearman's rho and the linear correlation coefficient $\varrho = \Sigma_{12}/\sqrt{\Sigma_{11}\Sigma_{22}}$ for $\mathbf{W} \triangleq A\mathbf{U}$. We only consider the case with $\operatorname{rank}(\Sigma) = 2$, since the case with $\operatorname{rank}(\Sigma) = 1$ is trivial. From the invariance of Spearman's rho under componentwise strictly increasing transformations of the underlying random vector we can without

loss of generality assume that $\Sigma_{11} = \Sigma_{22} = 1$ and $\Sigma_{12} = \Sigma_{21} = \varrho$. We show that the following relation holds,

$$\varrho_S(W_1, W_2) = 3\left(\frac{\arcsin \varrho}{\pi}\right) - 4\left(\frac{\arcsin \varrho}{\pi}\right)^3. \tag{2.6}$$

In the case of a bivariate normal distribution, i.e. $R \sim \chi^2_2$, we know from Corollary 2.21 that the relation between Spearman's rho and the linear correlation coefficient is

$$\varrho_S(X_1, X_2) = \frac{6}{\pi} \arcsin(\varrho/2). \tag{2.7}$$

Since these two relations differ (the difference is plotted in Figure 2.1) we conclude that, contrary to Kendall's tau, Spearman's rho is not invariant in the class of elliptical distributions with a fixed dispersion matrix. It remains to be shown that (2.6) holds. This can be done following the steps below.

Step 1. Let (W_1, W_2), (W_1', W_2') and (W_1'', W_2'') be independent copies. Then

$$\varrho_S(W_1, W_2) = 12\,\mathbb{P}(W_1' \leq W_1, W_2'' \leq W_2) - 3.$$

Step 2. For (W_1, W_2), W_1', W_2'' as above we have that,

$$\begin{aligned}\mathbb{P}(W_1' \leq W_1, W_2'' \leq W_2) = \frac{1}{2\pi}\int_0^{2\pi} \Bigg(&\frac{1}{2} + \frac{1}{\pi}\arcsin(\sin(\arcsin\varrho + t)) \\ &- \frac{1}{2\pi}\arccos(\cos t) \\ &- \frac{1}{\pi^2}\arccos(\cos t)\arcsin(\sin(\arcsin\varrho + t))\Bigg)\,\mathrm{d}t.\end{aligned}$$

Step 3. The following equalities hold:

(i) $\displaystyle\int_0^{2\pi} \arcsin(\sin(\arcsin\varrho + t))\mathrm{d}t = 0.$

(ii) $\displaystyle\int_0^{2\pi} \arccos(\cos t)\mathrm{d}t = \pi^2.$

(iii) $\displaystyle\int_0^{2\pi} \arccos(\cos t)\arcsin(\sin(\arcsin\varrho + t))\mathrm{d}t = \frac{2}{3}(\arcsin\varrho)^3 - \frac{\pi^2}{2}\arcsin\varrho.$

Combining Steps 1-3 yields (2.6),

$$\begin{aligned}\varrho_S(W_1, W_2) &= 12\,\mathbb{P}(W_1' \leq W_1, W_2'' \leq W_2) - 3 \\ &= \frac{12}{2\pi}\left(\pi - \frac{\pi}{2} - \frac{2}{3\pi^2}(\arcsin\varrho)^3 + \frac{1}{2}\arcsin\varrho\right) - 3 \\ &= 3\left(\frac{\arcsin\varrho}{\pi}\right) - 4\left(\frac{\arcsin\varrho}{\pi}\right)^3.\end{aligned}$$

Proof of Step 1. Straightforward computations of Spearman's rho for continuous random variables yields

$$\begin{aligned}\varrho_S(W_1, W_2) &= 3\,(2\,\mathbb{P}((W_1 - W_1')(W_2 - W_2'') > 0) - 1) \\ &= 3\,(4\,\mathbb{P}(W_1' \leq W_1, W_2'' \leq W_2) - 1) \\ &= 12\,\mathbb{P}(W_1' \leq W_1, W_2'' \leq W_2) - 3.\end{aligned}$$

□

Proof of Step 2. Let $\varphi, \varphi', \varphi'' \sim U(0, 2\pi)$ be independent. Then

$$\begin{aligned}(W_1, W_2) &\stackrel{\mathrm{d}}{=} (\cos\varphi, \sin(\arcsin\varrho + \varphi)), \\ (W_1', W_2') &\stackrel{\mathrm{d}}{=} (\cos\varphi', \sin(\arcsin\varrho + \varphi')), \\ (W_1'', W_2'') &\stackrel{\mathrm{d}}{=} (\cos\varphi'', \sin(\arcsin\varrho + \varphi'')).\end{aligned}$$

Conditioning on φ yields,

$$\begin{aligned}&\mathbb{P}(W_1' \leq W_1, W_2'' \leq W_2) = \mathbb{P}(\cos\varphi' \leq \cos\varphi, \sin(\arcsin\varrho + \varphi'') \leq \sin(\arcsin\varrho + \varphi)) \\ &= \frac{1}{2\pi}\int_0^{2\pi} \mathbb{P}(\cos t - \cos\varphi' \geq 0)\,\mathbb{P}(\sin(\arcsin\varrho + t) - \sin(\arcsin\varrho + \varphi'') \geq 0)\mathrm{d}t.\end{aligned}$$

The factors in the integrand can be written as

$$\begin{aligned}&\mathbb{P}(\cos t - \cos\varphi' \geq 0) = 1 - \frac{1}{2\pi} 2\arccos(\cos t), \\ &\mathbb{P}(\sin(\arcsin\varrho + t) - \sin(\arcsin\varrho + \varphi'') \geq 0) \\ &\quad = 1 - \frac{1}{2\pi}\left(\pi - \arcsin(\sin(\arcsin\varrho + t)) - \arcsin(\sin(\arcsin\varrho + t))\right) \\ &\quad = \frac{1}{2} + \frac{1}{2\pi} 2\arcsin(\sin(\arcsin\varrho + t)).\end{aligned}$$

Combining these expressions yields,

$$\begin{aligned}&\mathbb{P}(\cos t - \cos\varphi' \geq 0)\,\mathbb{P}(\sin(\arcsin\varrho + t) - \sin(\arcsin\varrho + \varphi'') \geq 0) \\ &\qquad = \frac{1}{2} + \frac{1}{\pi}\arcsin(\sin(\arcsin\varrho + t)) - \frac{1}{2\pi}\arccos(\cos t) \\ &\qquad\quad - \frac{1}{\pi^2}\arccos(\cos t)\arcsin(\sin(\arcsin\varrho + t)).\end{aligned}$$

□

Proof of Step 3. (i) and (ii) are elementary. To compute (iii) we first split the integral depending on $\arccos(\cos t)$ and then use a variable transformation to obtain

$$\begin{aligned}
\mathcal{I} \triangleq & \frac{1}{\pi^2}\int_0^{2\pi} \arccos(\cos t)\arcsin(\sin(\arcsin\varrho + t))\mathrm{d}t \\
= & \frac{1}{\pi^2}\left(\int_0^{\pi} t\arcsin(\sin(\arcsin\varrho + t))\mathrm{d}t \right. \\
& \left. + \int_{\pi}^{2\pi}(2\pi - t)\arcsin(\sin(\arcsin\varrho + t))\mathrm{d}t\right) \\
= & \frac{1}{\pi^2}\left(\int_{\arcsin\varrho}^{\pi+\arcsin\varrho}(u - \arcsin\varrho)\arcsin(\sin u)\mathrm{d}u \right. \\
& \left. + \int_{\pi+\arcsin\varrho}^{2\pi+\arcsin\varrho}(2\pi - u + \arcsin\varrho)\arcsin(\sin u)\mathrm{d}u\right) \\
= & \frac{1}{\pi^2}\left(\underbrace{\int_0^{\pi}(u - \arcsin\varrho)\arcsin(\sin u)\mathrm{d}u}_{I} - \underbrace{\int_0^{\arcsin\varrho}(u - \arcsin\varrho)u\mathrm{d}u}_{II}\right. \\
& + \underbrace{\int_{\pi}^{\pi+\arcsin\varrho}(u - \arcsin\varrho)(\pi - u)\mathrm{d}u}_{III} \\
& + \underbrace{\int_{\pi}^{2\pi}(2\pi - u + \arcsin\varrho)\arcsin(sinu)\mathrm{d}u}_{IV} \\
& - \underbrace{\int_{\pi}^{\pi+\arcsin\varrho}(2\pi - u + \arcsin\varrho)(\pi - u)\mathrm{d}u}_{V} \\
& \left. + \underbrace{\int_{2\pi}^{2\pi+\arcsin\varrho}(2\pi - u + \arcsin\varrho)(u - 2\pi)\mathrm{d}u}_{VI}\right)
\end{aligned}$$

The different parts can now be computed separately.

$$\begin{aligned}
I = & \int_0^{\pi}(u - \arcsin\varrho)\arcsin(\sin u)\mathrm{d}u = \int_0^{\pi/2}(u - \arcsin\varrho)u\mathrm{d}u \\
& + \int_{\pi/2}^{\pi}(u - \arcsin\varrho)(\pi - u)\mathrm{d}u = \pi^3/8 - \pi^2(\arcsin\varrho)/4
\end{aligned}$$

$$
\begin{aligned}
II &= \int_0^{\arcsin\varrho} (u - \arcsin\varrho) u \mathrm{d}u = -(\arcsin\varrho)^3/6 \\
III &= \int_\pi^{\pi+\arcsin\varrho} (u - \arcsin\varrho)(\pi - u)\mathrm{d}u = -\pi(\arcsin\varrho)^2/2 + (\arcsin\varrho)^3/6 \\
IV &= \int_\pi^{2\pi} (2\pi - u + \arcsin\varrho)\arcsin(\sin u)\mathrm{d}u \\
&= \int_\pi^{3\pi/2} (2\pi - u + \arcsin\varrho)(\pi - u)\mathrm{d}u + \int_{3\pi/2}^{2\pi} (2\pi - u + \arcsin\varrho)(u - 2\pi)\mathrm{d}u \\
&= -\pi^3/8 - \pi^2(\arcsin\varrho)/4 \\
V &= \int_\pi^{\pi+\arcsin\varrho} (2\pi - u + \arcsin\varrho)(\pi - u)\mathrm{d}u = -\pi(\arcsin\varrho)^2/2 - (\arcsin\varrho)^3/6 \\
VI &= \int_{2\pi}^{2\pi+\arcsin\varrho} (2\pi - u + \arcsin\varrho)(u - 2\pi)\mathrm{d}u = (\arcsin\varrho)^3/6
\end{aligned}
$$

Putting everything together yields

$$
\begin{aligned}
\mathcal{I} &= \frac{1}{\pi^2}(I - II + III + IV - V + VI) = \frac{1}{\pi^2}\left(\frac{2}{3}(\arcsin\varrho)^3 - \frac{\pi^2}{2}\arcsin\varrho\right) \\
&= \frac{2}{3\pi^2}(\arcsin\varrho)^3 - \frac{1}{2}\arcsin\varrho.
\end{aligned}
$$

□

References

[1] Basrak, B. (2000) *The Sample Autocorrelation Function of Non-Linear Time Series*, Ph.D. Thesis, University of Groningen, Department of Mathematics.

[2] Billingsley, P. (1995) *Probability and Measure*, 3rd edition, Wiley, New York.

[3] Cambanis, S., Huang, S. and Simons, G. (1981) On the theory of elliptically contoured distributions, *J. Multivariate Anal.*, Vol. 11, 368–385.

[4] Embrechts, P., Klüppelberg C. and Mikosch, T. (1997) *Modelling Extremal Events for Insurance and Finance*, Springer, Berlin.

[5] Embrechts, P., McNeil A. and Straumann D. (2002) Correlation and Dependence in Risk Management: Properties and Pitfalls, in: M.A.H. Dempster ed., *Risk Management: Value at Risk and Beyond*, Cambridge University Press, Cambridge, 176–223.

[6] Fang, K.-T., Kotz, S. and Ng, K.-W. (1987) *Symmetric Multivariate and Related Distributions*, Chapman & Hall, London.

[7] Kallenberg, O. (1983) *Random Measures*, 3rd edition, Akademie Verlag, Berlin.

[8] Lindskog, F., McNeil, A.J. and Schmock, U. (2001) A note on Kendall's tau for Elliptical distributions, *Preprint, available at www.risklab.ch/∼Papers*

[9] Mikosch, T. (2001) Modelling dependence and tails of financial time series, In *SemStat: Seminaire Europeen de Statistique, Extreme Value Theory and Applications*, Chapman & Hall, London, 75 pages.

[10] Resnick, S.I. (1987) *Extreme Values, Regular Variation, and Point Processes*, Springer Verlag, New York.

[11] Stărică, C. (1999) Multivariate extremes for models with constant conditional correlations, *J. Empirical Finance*, Vol. 6, 515–553.

Chapter 3

Multivariate regular variation for additive processes

Henrik Hult and Filip Lindskog (2002) Multivariate regular variation for additive processes, *submitted*.

Abstract. We study the joint tail behavior of heavy-tailed $\mathbb{R}^d$-valued additive processes, i.e. stochastically continuous processes with independent increments, and the joint tail behavior of vectors of functionals acting on each component of such processes. More precisely we consider additive processes which at a fixed time $t > 0$ satisfy a multivariate regular variation condition. Tail equivalence between the process at time t and its Lévy measure, in the sense of having the same multivariate regular variation limit measure, is established. We also derive regular variation limit measures for vectors consisting of the componentwise suprema, the componentwise suprema of the jumps and the componentwise integrals over the time interval $[0, t]$. In order to derive the limit measure for the vector of componentwise integrals we study convergence of the appropriately scaled probability that the process reaches sets in the product space $[0, t] \times \mathbb{R}^d$, and we establish several equivalent formulations of this convergence.

2000 *Mathematics Subject Classification.* 60G51 (primary); 60G70, 60E07 (secondary).
Keywords and phrases. Multivariate regular variation; Lévy processes; Vague convergence; Extreme value theory.
Acknowledgments. The authors want to thank Boualem Djehiche, Paul Embrechts and Thomas Mikosch for comments on the manuscript.

3.1 Introduction

Stochastic processes with heavy-tailed marginal distributions have become increasingly important in many applications such as communication networks, hydrology, insurance mathematics and mathematical finance. A lot of effort has been put into studying such processes and into finding the tail asymptotics of the distribution of functionals of the processes (see e.g. Embrechts, Goldie and Veraverbeke [7], Rosiński and Samorodnitsky [14] and Braverman, Mikosch and Samorodnitsky [5]). However, whereas problems concerning the tail behavior for univariate stochastic processes have been studied successfully for quite some time, multivariate processes have received far less attention. Even though the intuition behind the univariate results to a large extent extends to the multivariate case, proving results in the multivariate case requires other tools such as vague convergence of measures.

The notion of heavy tails used in this paper is that of multivariate regular variation. It says that an $\mathbb{R}^d$-valued random vector $\mathbf{X}$ with unbounded support is *regularly varying* if there exists a sequence $\{a_n\}$, $0 < a_n \uparrow \infty$, and a nonzero Radon measure μ defined on $\mathcal{B}(\overline{\mathbb{R}}^d \backslash \{\mathbf{0}\})$ with $\mu(\overline{\mathbb{R}}^d \backslash \mathbb{R}^d) = 0$ such that, as $n \to \infty$,

$$n\,\mathbb{P}(a_n^{-1}\mathbf{X} \in \cdot\,) \xrightarrow{v} \mu(\cdot) \quad \text{on } \mathcal{B}(\overline{\mathbb{R}}^d \backslash \{\mathbf{0}\}).$$

Here $\xrightarrow{v}$ denotes vague convergence on $\mathcal{B}(\overline{\mathbb{R}}^d \backslash \{\mathbf{0}\})$. We refer to Kallenberg [9] for details on vague convergence. The reason for working on the space $\overline{\mathbb{R}}^d \backslash \{\mathbf{0}\}$ instead of $\mathbb{R}^d$ has to do with vague convergence; we want sets in $\mathbb{R}^d$ which are bounded away from the origin to be relatively compact. We focus on multivariate additive processes, i.e. stochastically continuous processes with independent increments, which at some fixed time $t > 0$ satisfy the above regular variation condition, and we study the tail behavior of vectors of functionals acting on such processes. By tail behavior we mean a limit measure of the above type. The basic intuition underlying all the results is the following: the process reaches a set far away from the origin by making one big jump at some time $\tau < t$, and in comparison to the size of the jump the process does not move much before τ nor between τ and t.

We prove tail equivalence between the distribution of $\mathbf{X}_t$ and its associated Lévy measure ν_t in the sense that if μ_t is a nonzero Radon measure on $\mathcal{B}(\overline{\mathbb{R}}^d \backslash \{\mathbf{0}\})$ with $\mu_t(\overline{\mathbb{R}}^d \backslash \mathbb{R}^d) = 0$, then

$$n\,\mathbb{P}(a_n^{-1}\mathbf{X}_t \in \cdot\,) \xrightarrow{v} \mu_t \quad \text{on } \mathcal{B}(\overline{\mathbb{R}}^d \backslash \{\mathbf{0}\}) \tag{3.1}$$

if and only if

$$n\nu_t(a_n \cdot\,) \xrightarrow{v} \mu_t \quad \text{on } \mathcal{B}(\overline{\mathbb{R}}^d \backslash \{\mathbf{0}\}). \tag{3.2}$$

This is a multivariate version, in the regularly varying case, of a result in Embrechts, Goldie and Veraverbeke [7] which says that

$$\mathbb{P}(X_t > x) \sim \nu_t(\{y \in \mathbb{R} : y > x\}) \quad \text{as } x \to \infty,$$

for X_t subexponential. Moreover, we determine the implications of regular variation of $\mathbf{X}_t$ on the joint tail behavior of vectors of functionals acting on the underlying process $\{\mathbf{X}_s : s \geq 0\}$. For example, we study the componentwise suprema

$$\mathbf{X}_t^* = \Big(\sup_{0\leq s\leq t} X_s^{(1)}, \ldots, \sup_{0\leq s\leq t} X_s^{(d)} \Big)$$

and the componentwise suprema of the jumps

$$\mathbf{X}_t^\Delta = \Big(\sup_{0< s\leq t} \Delta X_s^{(1)}, \ldots, \sup_{0< s\leq t} \Delta X_s^{(d)} \Big),$$

and we show that if the additive process $\{\mathbf{X}_s : s \geq 0\}$ satisfies the regular variation condition (3.1), then

$$n\,\mathbb{P}(a_n^{-1}\mathbf{X}_t^* \in \cdot\,) \xrightarrow{v} \mu_t(\cdot) \quad \text{and} \quad n\,\mathbb{P}(a_n^{-1}\mathbf{X}_t^\Delta \in \cdot\,) \xrightarrow{v} \mu_t(\cdot) \quad \text{on } \mathcal{B}(\overline{\mathbb{R}}_+^d\backslash\{\mathbf{0}\}),$$

i.e. we have convergence to the same limit measure. In this case, knowing the joint tail behavior of $\mathbf{X}_t$ is enough to determine the joint tail behavior of $\mathbf{X}_t^*$ and $\mathbf{X}_t^\Delta$ respectively. However, there are many other interesting choices of functionals for which the joint tail behavior cannot be determined by that of $\mathbf{X}_t$ alone. Typically, we would need information of the type: the probability that $\mathbf{X}_s$ for some $s \in [0,t]$ reaches a set $a_n B_s$, where B_s is allowed to vary with s. This can be formulated in terms of a vague limit on $\mathcal{B}([0,t] \times (\overline{\mathbb{R}}^d\backslash\{\mathbf{0}\}))$ for the graph $\{(s, \mathbf{X}_s) : 0 \leq s \leq t\}$ of the process. We prove tail equivalence between the graph of the process and the associated measure $\widetilde{\nu}$, where $\widetilde{\nu}([0,s] \times B) = \nu_s(B)$ for $s \in [0,t]$ and $B \in \mathcal{B}(\overline{\mathbb{R}}^d\backslash\{\mathbf{0}\})$, in the sense of having the same vague limit on $\mathcal{B}([0,t] \times (\overline{\mathbb{R}}^d\backslash\{\mathbf{0}\}))$. Since the precise formulation of the result (Theorem 3.20) requires some additional notation and a few technicalities, we refrain from stating it at this point. Using this result we can determine the joint tail behavior of the componentwise integrals

$$\mathbf{I}_t = \Big(\int_0^t X_s^{(1)} \mathrm{d}s, \ldots, \int_0^t X_s^{(d)} \mathrm{d}s \Big)$$

in the sense of the vague limit of $n\,\mathbb{P}(a_n^{-1}\mathbf{I}_t \in \cdot\,)$ on $\mathcal{B}(\overline{\mathbb{R}}^d\backslash\{\mathbf{0}\})$. As a special case, if $\{\mathbf{X}_s : s \geq 0\}$ is a Lévy process, then it follows that

$$n\,\mathbb{P}(a_n^{-1}\mathbf{I}_t \in \cdot\,) \xrightarrow{v} \frac{t^{\alpha+1}}{\alpha+1}\mu(\cdot) \quad \text{on } \mathcal{B}(\overline{\mathbb{R}}^d\backslash\{\mathbf{0}\}),$$

where μ and $\alpha > 0$ are such that $\mu_t(\cdot) = t\mu(\cdot)$ and $\mu(x\,\cdot) = x^{-\alpha}\mu(\cdot)$ for $x > 0$.

The paper is organized as follows. We begin in Section 3.2 by reviewing the notion of multivariate regular variation and we introduce the additive processes in Section 3.3 and recall some of their properties. In Section 3.4 we consider sums of independent regularly varying random vectors and prove some new results. In Section 3.5 these results are used to prove the main results on tail equivalence and joint tail behavior of functionals acting on the components of the processes.

3.1.1 Notation

We denote by $\overline{\mathbb{R}} = \mathbb{R} \cup \{-\infty, \infty\}$ the extended real line and $\overline{\mathbb{R}}_+ = [0, \infty]$. For sets $B \subset A \subset \overline{\mathbb{R}}^d$,

$$A \backslash B = \{\mathbf{x} \in \overline{\mathbb{R}}^d : \mathbf{x} \in A \cap B^c\}.$$

We denote by $\mathcal{R}^d$ the standard Borel σ-algebra on $\mathbb{R}^d$. There is a metric ρ on $\overline{\mathbb{R}}^d \backslash \{\mathbf{0}\}$ which makes $(\overline{\mathbb{R}}^d \backslash \{\mathbf{0}\}, \rho)$ a separable and complete metric space, we denote by $\mathcal{B}(\overline{\mathbb{R}}^d \backslash \{\mathbf{0}\})$ the Borel σ-algebra generated by the open sets with respect to ρ. Furthermore, on $\mathbb{R}^d \backslash \{\mathbf{0}\}$, $\mathcal{B}(\overline{\mathbb{R}}^d \backslash \{\mathbf{0}\})$ and $\mathcal{R}^d$ coincide. For a set $E \subset \overline{\mathbb{R}}^d \backslash \{\mathbf{0}\}$ we denote by $\mathcal{B}(E)$ the σ-algebra generated by the subspace topology. For a set $A \subset \overline{\mathbb{R}}^d \backslash \{\mathbf{0}\}$ ($\overline{\mathbb{R}}^d_+ \backslash \{\mathbf{0}\}$) we denote by ∂A its boundary (∂A is always measurable since it is closed). If μ is a measure on $\mathcal{B}(\overline{\mathbb{R}}^d \backslash \{\mathbf{0}\})$ ($\mathcal{B}(\overline{\mathbb{R}}^d_+ \backslash \{\mathbf{0}\})$), then A is called a μ-continuity set if $\mu(\partial A) = 0$.

Operations between $\mathbb{R}^d$-valued ($\overline{\mathbb{R}}^d$-valued) vectors are always interpreted as componentwise operations, e.g. $\mathbf{a} < \mathbf{b}$ means $a^{(i)} < b^{(i)}$ for $i = 1, \ldots, d$. Using this notation we have $[\mathbf{a}, \mathbf{b}) = \{\mathbf{x} \in \overline{\mathbb{R}}^d : \mathbf{a} \leq \mathbf{x} < \mathbf{b}\}$. We denote by $|\cdot|$ an arbitrary norm on $\mathbb{R}^d$ and by $\mathbb{S}^{d-1} = \{\mathbf{z} \in \mathbb{R}^d : |\mathbf{z}| = 1\}$ and $B_{\mathbf{x},\epsilon} = \{\mathbf{z} \in \mathbb{R}^d : |\mathbf{z}| < \epsilon\}$ the unit sphere and the open ball with radius ϵ centered at $\mathbf{x}$, respectively, with respect to this norm.

The symbol $\sim$ denotes both asymptotic equivalence, i.e. $f(x) \sim g(x)$ as $x \to \infty$ if $\lim_{x\to\infty} f(x)/g(x) = 1$, and that a random vector has a certain probability distribution, i.e. $\mathbf{X} \sim F$ means that $\mathbf{X}$ has the distribution F. The dual use of $\sim$ should not cause any confusion.

3.2 Multivariate regular variation

The notion of multivariate regularly varying random vectors has appeared in several apparently different applications such as the study of stochastic recurrence equations (see Kesten [10]), multivariate extreme value theory (see e.g. Resnick [13]) and the description of weak limits of point processes constructed from stationary sequences of random vectors (see Davis and Hsing [6]). In recent years some effort has been made to establish equivalence of the different notions of multivariate regular variation (see Basrak [1] and Basrak, Davis and Mikosch [3]) and as a result several equivalent definitions are at hand. The following definition is perhaps the most intuitive one.

Definition 3.1. An $\mathbb{R}^d$-valued random vector $\mathbf{X}$ with unbounded support is said to be *regularly varying with index* $\alpha > 0$ if there exists a probability measure σ on $\mathbb{S}^{d-1}$ such that for all $x > 0$, as $u \to \infty$,

$$\frac{\mathbb{P}(|\mathbf{X}| > ux, \mathbf{X}/|\mathbf{X}| \in \cdot)}{\mathbb{P}(|\mathbf{X}| > u)} \xrightarrow{v} x^{-\alpha}\sigma(\cdot) \quad \text{on } \mathcal{B}(\mathbb{S}^{d-1}). \tag{3.3}$$

The probability measure σ is referred to as the *spectral measure* of $\mathbf{X}$ and α is referred to as the *tail index* of $\mathbf{X}$.

An equivalent formulation is the following (see Basrak [1]).

Theorem 3.2. *Let $\mathbf{X}$ be an $\mathbb{R}^d$-valued random vector with unbounded support. Then $\mathbf{X}$ is regularly varying in the sense of Definition 3.1 if and only if there exists a sequence $\{a_n\}$, $0 < a_n \uparrow \infty$, and a nonzero Radon measure μ on $\mathcal{B}(\overline{\mathbb{R}}^d \backslash \{\mathbf{0}\})$ with $\mu(\overline{\mathbb{R}}^d \backslash \mathbb{R}^d) = 0$ such that, as $n \to \infty$,*

$$n\,\mathbb{P}(a_n^{-1}\mathbf{X} \in \cdot\,) \xrightarrow{v} \mu(\cdot) \quad \text{on } \mathcal{B}(\overline{\mathbb{R}}^d \backslash \{\mathbf{0}\}). \tag{3.4}$$

Moreover, if (3.4) holds, then there exists $\alpha > 0$ such that $\mu(xB) = x^{-\alpha}\mu(B)$ for every $x > 0$ and $B \in \mathcal{B}(\overline{\mathbb{R}}^d \backslash \{0\})$.

Remark 3.3. (i) The scaling property of the limit measure μ implies that μ assigns no mass to spheres centered at the origin, i.e. sets of the form $r\mathbb{S}^{d-1}$ for $r > 0$, with respect to any norm on $\mathbb{R}^d$. In particular μ has no atoms (see Basrak [1]).
(ii) A standard regular variation argument shows that if (3.4) holds, then the sequence $\{a_n\}$ is regularly varying with index $1/\alpha$, i.e. for every $\lambda > 0$, $a_{[\lambda n]}/a_n \to \lambda^{1/\alpha}$ as $n \to \infty$.

Let us now consider yet another possible formulation of multivariate regular variation; there exists an $\alpha > 0$ and a slowly varying function $L(u)$ such that for every $\mathbf{x} \in \mathbb{R}^d \backslash \{\mathbf{0}\}$

$$\lim_{u\to\infty} \frac{\mathbb{P}(\langle \mathbf{x}, \mathbf{X}\rangle > u)}{u^{-\alpha}L(u)} = w(\mathbf{x}) \quad \text{exists and is finite} \tag{3.5}$$

and $w(\mathbf{x}) = 0$ is possible for some but not all $\mathbf{x} \in \mathbb{R}^d \backslash \{\mathbf{0}\}$. It follows immediately that the function w is homogeneous, $w(r\mathbf{x}) = r^\alpha w(\mathbf{x})$ for every $r > 0$ and $\mathbf{x} \in \mathbb{R}^d \backslash \{\mathbf{0}\}$. It is easy to show that (3.3) implies (3.5). In Basrak, Davis and Mikosch [3] the following theorem proves equivalence of (3.3) and (3.5) under some additional assumptions.

Theorem 3.4. *Let $\mathbf{X}$ be an $\mathbb{R}^d$-valued random vector with unbounded support. Then (3.3) and (3.5) are equivalent if either (i) α is a positive noninteger or (ii) $\mathbf{X}$ has nonnegative components and α is an odd integer.*

Let $\mathbf{X}$ be a random vector satisfying (3.3), i.e. $\mathbf{X}$ is regularly varying. Then $\mathbf{X}$ satisfies (3.5) and the limit function w is uniquely determined by α and σ. On the other hand, the limit function w determines α but not necessarily the spectral measure σ if α is a positive integer. Consider the following example.

Example 3.5. Fix an integer $\alpha \geq 1$. We will construct two regularly varying random vectors $\mathbf{X}_1$ and $\mathbf{X}_2$ with tail index α and different spectral measures $\sigma_1(\cdot) = \mathbb{P}(\Theta_1 \in \cdot\,)$ and $\sigma_2(\cdot) = \mathbb{P}(\Theta_2 \in \cdot\,)$, such that the limit functions w_1 and w_2 in (3.5) coincide.

Let Θ_1 be a $[0, 2\pi)$-valued random variable with density function $f_1(\theta) > \epsilon$ for all $\theta \in [0, 2\pi)$ and some $\epsilon > 0$. Take $\delta \in (0, \epsilon)$ and let Θ_2 have density function f_2 given by

$$f_2(\theta) = f_1(\theta) + \delta \sin((\alpha + 2)\theta), \quad \theta \in [0, 2\pi).$$

Let $R \sim \text{Pareto}(\alpha)$, i.e. $\mathbb{P}(R > x) = x^{-\alpha}$ for $x \geq 1$, be independent of Θ_i, $i = 1, 2$, and put

$$\mathbf{X}_i \stackrel{\text{d}}{=} \begin{pmatrix} R\cos(\Theta_i) \\ R\sin(\Theta_i) \end{pmatrix}.$$

Take $\mathbf{x} \in \mathbb{R}^2 \backslash \{\mathbf{0}\}$ and let $\beta \in [0, 2\pi)$ be given by

$$\frac{\mathbf{x}}{|\mathbf{x}|} = \begin{pmatrix} \cos(\beta) \\ \sin(\beta) \end{pmatrix}.$$

Then, for $u > |\mathbf{x}|$,

$$\begin{aligned}
&\mathbb{P}(\langle \mathbf{x}, \mathbf{X}_2 \rangle > u) - \mathbb{P}(\langle \mathbf{x}, \mathbf{X}_1 \rangle > u) \\
&\quad = \mathbb{P}(\langle \mathbf{x}/|\mathbf{x}|, \mathbf{X}_2 \rangle > u/|\mathbf{x}|) - \mathbb{P}(\langle \mathbf{x}/|\mathbf{x}|, \mathbf{X}_1 \rangle > u/|\mathbf{x}|) \\
&\quad = \delta \int_{u/|\mathbf{x}|}^{\infty} \int_{\beta - \arccos((u/|\mathbf{x}|)/r)}^{\beta + \arccos((u/|\mathbf{x}|)/r)} \alpha r^{-\alpha - 1} \sin((\alpha + 2)\theta) \mathrm{d}\theta \mathrm{d}r \\
&\quad = \delta \int_{u/|\mathbf{x}|}^{\infty} \alpha r^{-\alpha - 1} \int_{\beta - \arccos((u/|\mathbf{x}|)/r)}^{\beta + \arccos((u/|\mathbf{x}|)/r)} \sin((\alpha + 2)\theta)) \mathrm{d}\theta \mathrm{d}r \\
&\quad = -\frac{\delta \alpha}{\alpha + 2} \int_{u/|\mathbf{x}|}^{\infty} r^{-\alpha - 1} \Big(\cos((\alpha + 2)(\beta + \arccos((u/|\mathbf{x}|)/r))) \\
&\qquad\qquad - \cos((\alpha + 2)(\beta - \arccos((u/|\mathbf{x}|)/r))) \Big) \mathrm{d}r \\
&\quad = \frac{2\delta \alpha}{\alpha + 2} \sin((\alpha + 2)\beta) \int_{u/|\mathbf{x}|}^{\infty} r^{-\alpha - 1} \sin((\alpha + 2) \arccos((u/|\mathbf{x}|)/r)) \mathrm{d}r
\end{aligned}$$

Using standard variable substitutions and trigonometric formulas the integral can be rewritten as follows.

$$\begin{aligned}
&\int_{u/|\mathbf{x}|}^{\infty} r^{-\alpha - 1} \sin((\alpha + 2) \arccos((u/|\mathbf{x}|)/r)) \mathrm{d}r \\
&\quad = \frac{u^{-\alpha}}{|\mathbf{x}|^{-\alpha}} \int_0^1 v^{\alpha - 1} \sin((\alpha + 2) \arccos(v)) \mathrm{d}v \\
&\quad = \frac{u^{-\alpha}}{|\mathbf{x}|^{-\alpha}} \int_0^{\pi/2} \cos^{\alpha - 1}(s) \sin((\alpha + 2)s) \sin(s) \mathrm{d}s \\
&\quad = \frac{u^{-\alpha}}{|\mathbf{x}|^{-\alpha}} \int_0^{\pi/2} \cos^{\alpha - 1}(s) \cos((\alpha + 1)s) \mathrm{d}s - \frac{u^{-\alpha}}{|\mathbf{x}|^{-\alpha}} \int_0^{\pi/2} \cos^{\alpha}(s) \cos((\alpha + 2)s) \mathrm{d}s.
\end{aligned}$$

The two last integrals are zero for every integer $\alpha \in (0, \infty)$; these integrals can be found in Gradshteyn and Ryzhik [8] p. 392. Hence, for $u > |\mathbf{x}|$,

$$\mathbb{P}(\langle \mathbf{x}, \mathbf{X}_2 \rangle > u) = \mathbb{P}(\langle \mathbf{x}, \mathbf{X}_1 \rangle > u).$$

Remark 3.6. Theorem 3.4 (i) is proved in Basrak, Davis and Mikosch [3] by showing that the limit function w in (3.5) uniquely determines the spectral measure σ in (3.3) if $\alpha > 0$ is not an integer. The above example shows that the idea behind the proof can not be extended to integer-valued tail indices. Whether the result still holds in this case is not known.

3.3 Additive processes

In this section we introduce additive processes following Sato [15]. An *additive process* $\{\mathbf{X}_t : t \geq 0\}$ on $\mathbb{R}^d$ is a stochastically continuous stochastic process with independent increments, starting at zero. There exists a version of it which has right continuous sample paths with left limits. We will always choose such a version. If in addition $\{\mathbf{X}_t : t \geq 0\}$ has stationary increments, then it is called a *Lévy process*. For an additive process $\{\mathbf{X}_t : t \geq 0\}$ on $\mathbb{R}^d$, for every t, the distribution of $\mathbf{X}_t$ is *infinitely divisible*, i.e. for any positive integer n, there exist iid random vectors $\mathbf{Z}_{1,n,t}, \ldots, \mathbf{Z}_{n,n,t}$ such that $\mathbf{X}_t \stackrel{\mathrm{d}}{=} \sum_{i=1}^n \mathbf{Z}_{i,n,t}$. If $\{\mathbf{X}_t : t \geq 0\}$ is a Lévy process, then we can take $\mathbf{Z}_{i,n,t} = \mathbf{X}_{ti/n} - \mathbf{X}_{t(i-1)/n}$.

We denote by $\widehat{F}$ the characteristic function of a probability distribution F on $\mathbb{R}^d$,

$$\widehat{F}(\mathbf{z}) = \int_{\mathbb{R}^d} e^{i\langle \mathbf{z}, \mathbf{x}\rangle} F(\mathrm{d}\mathbf{x}).$$

For any infinitely divisible probability distribution F on $\mathbb{R}^d$ we have the *Lévy-Khintchine representation* (Sato [15] Theorem 8.1 p. 37)

$$\begin{aligned} \widehat{F}(\mathbf{z}) = \exp\Big(&-\frac{1}{2}\langle \mathbf{z}, A\mathbf{z}\rangle + i\langle \boldsymbol{\gamma}, \mathbf{z}\rangle \\ &+ \int_{\mathbb{R}^d\setminus\{\mathbf{0}\}} (e^{i\langle \mathbf{z},\mathbf{x}\rangle} - 1 - i\langle \mathbf{z}, \mathbf{x}\rangle 1_{\{|\mathbf{x}|\leq 1\}}(\mathbf{x}))\nu(\mathrm{d}\mathbf{x})\Big), \quad \mathbf{z} \in \mathbb{R}^d, \end{aligned} \tag{3.6}$$

where A is a symmetric nonnegative definite $d \times d$ matrix, ν is a measure on $\mathbb{R}^d\setminus\{\mathbf{0}\}$ satisfying

$$\int_{\mathbb{R}^d\setminus\{\mathbf{0}\}} (|\mathbf{x}|^2 \wedge 1)\nu(\mathrm{d}\mathbf{x}) < \infty, \tag{3.7}$$

and $\boldsymbol{\gamma} \in \mathbb{R}^d$. We call $(A, \nu, \boldsymbol{\gamma})$ the *generating triplet of* F. The matrix A and the measure ν are called, respectively, the *Gaussian covariance matrix* and the *Lévy measure* of F. When $A = 0$, F is called purely non-Gaussian.

An important result for additive processes is the Lévy-Itô decomposition which we will recall below. First some notation. Let

$$\begin{aligned} D_{a,b} &= \{\mathbf{x} \in \mathbb{R}^d : a < |\mathbf{x}| \leq b\}, \quad \text{for } 0 \leq a < b < \infty, \\ D_{a,\infty} &= \{\mathbf{x} \in \mathbb{R}^d : a < |\mathbf{x}| < \infty\}, \quad \text{for } 0 \leq a < \infty. \end{aligned}$$

Theorem 3.7 (Sato [15] Theorem 19.2 p. 120). *Let $\{\mathbf{X}_t : t \geq 0\}$ be an additive process on $\mathbb{R}^d$ defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with system of generating triplets $\{(A_t, \nu_t, \boldsymbol{\gamma}_t)\}$ and define the measure $\widetilde{\nu}$ on $(0,\infty) \times (\mathbb{R}^d \backslash \{\mathbf{0}\})$ by $\widetilde{\nu}((0,t] \times B) = \nu_t(B)$ for $B \in \mathcal{B}(\mathbb{R}^d \backslash \{\mathbf{0}\})$. Let $\Omega_0 \in \mathcal{F}$, $\mathbb{P}(\Omega_0) = 1$, be such that $t \mapsto \mathbf{X}_t(\omega)$ is right continuous in $t \geq 0$ with left limits in $t > 0$ for each $\omega \in \Omega_0$ and define, for $H \in \mathcal{B}((0,\infty) \times (\mathbb{R}^d \backslash \{\mathbf{0}\}))$,*

$$\xi(H, \omega) = \begin{cases} \#\{s : (s, \mathbf{X}_s(\omega) - \mathbf{X}_{s-}(\omega)) \in H\}, & \text{for } \omega \in \Omega_0 \\ 0, & \text{for } \omega \notin \Omega_0. \end{cases} \tag{3.8}$$

Then the following hold.

(i) $\{\xi(H) : H \in \mathcal{B}((0,\infty) \times (\mathbb{R}^d \backslash \{\mathbf{0}\}))\}$ *is a Poisson random measure on* $(0,\infty) \times (\mathbb{R}^d \backslash \{\mathbf{0}\})$ *with intensity measure* $\widetilde{\nu}$.

(ii) There is $\Omega_1 \in \mathcal{F}$ *with* $\mathbb{P}(\Omega_1) = 1$ *such that, for any* $\omega \in \Omega_1$,

$$\begin{aligned} \mathbf{X}_t^1(\omega) = \lim_{\epsilon \downarrow 0} \int_{(0,t] \times D_{\epsilon,1}} \{\mathbf{x}\xi(\mathrm{d}(s,\mathbf{x}), \omega) - \mathbf{x}\widetilde{\nu}(\mathrm{d}(s,\mathbf{x}))\} \\ + \int_{(0,t] \times D_{1,\infty}} \mathbf{x}\xi(\mathrm{d}(s,\mathbf{x}), \omega) \end{aligned} \tag{3.9}$$

is defined for all $t \in [0,\infty)$ *and the convergence is uniform in* t *on any bounded interval. The process* $\{\mathbf{X}_t^1\}$ *is an additive process on* $\mathbb{R}^d$ *with* $\{(0, \nu_t, 0)\}$ *as the system of generating triplets.*

(iii) Define

$$\mathbf{X}_t^2(\omega) = \mathbf{X}_t(\omega) - \mathbf{X}_t^1(\omega) \quad \text{for } \omega \in \Omega_1.$$

There is $\Omega_2 \in \mathcal{F}$ *with* $\mathbb{P}(\Omega_2) = 1$ *such that, for any* $\omega \in \Omega_2$, $\mathbf{X}_t^2(\omega)$ *is continuous in* t. *The process* $\{\mathbf{X}_t^2\}$ *is an additive process on* $\mathbb{R}^d$ *with* $\{(A_t, 0, \boldsymbol{\gamma}_t)\}$ *as the system of generating triplets.*

(iv) The two processes $\{\mathbf{X}_t^1\}$ *and* $\{\mathbf{X}_t^2\}$ *are independent.*

In connection to this theorem we give the supplementary result which says that the part of the additive process containing the small jumps will always have finite moments of all orders. This will be relevant when studying the tails of a regularly varying additive process.

Lemma 3.8. *Let* $\mathbf{Y}_t = \lim_{\epsilon \downarrow 0} \int_{(0,t] \times D_{\epsilon,1}} \{\mathbf{x}\xi(\mathrm{d}(s,\mathbf{x})) - \mathbf{x}\widetilde{\nu}(\mathrm{d}(s,\mathbf{x}))\}$. *Then for every* $t > 0$ *and* $m \in \mathbb{N}$, $\mathrm{E}\big(|\mathbf{Y}_t|^m\big) < \infty$.

Proof. Fix arbitrary $t > 0$ and integer $m \geq 1$ (the case $m = 0$ is trivial). For notational convenience, let z_j, x_j and $\mathbf{Y}_{t,j}$ denote the jth component of $\mathbf{z}$, $\mathbf{x}$ and $\mathbf{Y}_t$ respectively. Since any two norms on $\mathbb{R}^d$ are equivalent we may without loss of generality take $|\cdot|$ to be the standard Euclidean norm, i.e. the norm given by

$|\mathbf{x}| = (\sum_{i=1}^d x_i^2)^{1/2}$. Note that by the Lévy-Khintchine representation for infinitely divisible distributions

$$\int_{D_{0,1}} (e^{i\langle \mathbf{z},\mathbf{x}\rangle} - 1 - i\langle \mathbf{z},\mathbf{x}\rangle)\nu_t(\mathrm{d}\mathbf{x}) \quad \text{and} \quad \int_{D_{0,1}} |\mathbf{x}|^2 \nu_t(\mathrm{d}\mathbf{x})$$

exist finitely. Let $\mathbf{Y}_t^\epsilon = \int_{(0,t]\times D_{\epsilon,1}} \{\mathbf{x}\xi(\mathrm{d}(s,\mathbf{x})) - \mathbf{x}\widetilde{\nu}(\mathrm{d}(s,\mathbf{x}))\}$. Then $\mathbf{Y}_t^\epsilon \xrightarrow{d} \mathbf{Y}_t$ as $\epsilon \downarrow 0$ and hence the characteristic functions converges,

$$\begin{aligned} \mathrm{E}(e^{i\langle \mathbf{z},\mathbf{Y}_t^\epsilon\rangle}) &= \exp\Big\{ \int_{D_{\epsilon,1}} (e^{i\langle \mathbf{z},\mathbf{x}\rangle} - 1 - i\langle \mathbf{z},\mathbf{x}\rangle)\nu_t(\mathrm{d}\mathbf{x})\Big\} \\ &\to \exp\Big\{ \int_{D_{0,1}} (e^{i\langle \mathbf{z},\mathbf{x}\rangle} - 1 - i\langle \mathbf{z},\mathbf{x}\rangle)\nu_t(\mathrm{d}\mathbf{x})\Big\} = \mathrm{E}(e^{i\langle \mathbf{z},\mathbf{Y}_t\rangle}). \end{aligned}$$

If $n_1,\ldots,n_d \in \mathbb{N}$ with $n_j \geq 2$ for some $j \in \{1,\ldots,d\}$, then

$$\int_{D_{0,1}} x_1^{n_1}\cdot\ldots\cdot x_d^{n_d}\nu_t(\mathrm{d}\mathbf{x}) \leq \int_{D_{0,1}} x_j^{n_j}\nu_t(\mathrm{d}\mathbf{x}) \leq \int_{D_{0,1}} |\mathbf{x}|^2\nu_t(\mathrm{d}\mathbf{x}) < \infty.$$

Moreover,

$$\mathrm{E}(|\mathbf{Y}_t|^{2m}) = \sum_{n_1+\cdots+n_d=m} \frac{m!}{n_1!\ldots n_d!}\mathrm{E}(Y_{t,1}^{2n_1}\cdot\ldots\cdot Y_{t,d}^{2n_d}),$$

where since $2n_j \geq 2$ for some j,

$$\begin{aligned} \mathrm{E}(Y_{t,1}^{2n_1}\cdot\ldots\cdot Y_{t,d}^{2n_d}) &= \Big(\frac{1}{i}\frac{\partial}{\partial z_1}\Big)^{2n_1}\cdot\ldots\cdot\Big(\frac{1}{i}\frac{\partial}{\partial z_d}\Big)^{2n_d}\mathrm{E}(e^{i\langle \mathbf{z},\mathbf{Y}_t\rangle})\Big|_{\mathbf{z}=0} \\ &= \int_{D_{0,1}} x_1^{2n_1}\cdot\ldots\cdot x_d^{2n_d}\nu_t(\mathrm{d}\mathbf{x}) < \infty. \end{aligned}$$

Hence $\mathrm{E}(|\mathbf{Y}_t|^{2m}) < \infty$. However, $\mathrm{E}(|\mathbf{Y}_t|^m) \leq (\mathrm{E}(|\mathbf{Y}_t|^{2m}))^{1/2}$ from which the conclusion follows. □

3.4 Sums of regularly varying random vectors

In this section we will derive some useful results concerning sums of regularly varying random vectors. The results generalize known results in the univariate case to the multivariate setting but the techniques used in the proofs are quite different. Let us start with a result on linear transformations of a regularly varying random vector (see also Basrak, Davis and Mikosch [2]).

Proposition 3.9. *Let* $\mathbf{X}$ *be an* $\mathbb{R}^d$*-valued random vector and suppose there exist a sequence* $\{a_n\}, 0 < a_n \uparrow \infty$*, and a Radon measure* μ *on* $\mathcal{B}(\overline{\mathbb{R}}^d\backslash\{\mathbf{0}\})$ *with* $\mu(\overline{\mathbb{R}}^d\backslash\mathbb{R}^d) =$

0 *such that* $n\,\mathbb{P}(a_n^{-1}\mathbf{X} \in \cdot\,) \xrightarrow{v} \mu(\cdot)$ *on* $\mathcal{B}(\overline{\mathbb{R}}^d\backslash\{\mathbf{0}\})$. *If* $T : \mathbb{R}^d \to \mathbb{R}^p$, $p \le d$, *is a linear transformation of full rank, then*

$$n\,\mathbb{P}(a_n^{-1}T(\mathbf{X}) \in \cdot\,) \xrightarrow{v} \mu \circ T^{-1}(\cdot \cap \mathbb{R}^p) \quad \text{on } \mathcal{B}(\overline{\mathbb{R}}^p\backslash\{\mathbf{0}\}).$$

Remark 3.10. Note that the result above does not hold in general if $p > d$ or if $p \le d$ but T not of full rank. In either case we can only get convergence on the sub-σ-algebra $T(\mathcal{B}(\overline{\mathbb{R}}^d\backslash\{0\})) \subset \mathcal{B}(\overline{\mathbb{R}}^p\backslash\{0\})$.

Proof. Let $B \in \mathcal{B}(\overline{\mathbb{R}}^p\backslash\{\mathbf{0}\})$ be relatively compact. If $B \subset \overline{\mathbb{R}}^p\backslash\mathbb{R}^p$, then

$$n\,\mathbb{P}(a_n^{-1}T(\mathbf{X}) \in B) = \mu \circ T^{-1}(B \cap \mathbb{R}^p) = 0$$

for all n. Hence we can without loss of generality assume that $B \subset \mathbb{R}^p$. Since B is relatively compact there exists an $\epsilon > 0$ such that $\inf_{x\in B} |x| > \epsilon$. For $x \in B$ take y such that $T(y) = x$ (since T is onto such a y always exists). Since $|T(y)| \le \|T\||y|$ and $\|T\| < \infty$ for all linear transformations $T : \mathbb{R}^d \to \mathbb{R}^p$, $|y| \ge |x|/\|T\| > \epsilon/\|T\|$ for all $y \in T^{-1}(B)$. Hence $T^{-1}(B)$ is relatively compact in $\overline{\mathbb{R}}^d\backslash\{\mathbf{0}\}$. If $\mu(\partial T^{-1}(B)) = 0$, then

$$\begin{aligned} n\,\mathbb{P}(a_n^{-1}T(\mathbf{X}) \in B) &= n\,\mathbb{P}(a_n^{-1}T(\mathbf{X}) \in B) = n\,\mathbb{P}(T(a_n^{-1}\mathbf{X}) \in B) \\ &= n\,\mathbb{P}(a_n^{-1}\mathbf{X} \in T^{-1}(B)) \to \mu(T^{-1}(B)). \end{aligned}$$

Since $\mu(\partial T^{-1}(B)) = \mu \circ T^{-1}(\partial B)$ and $T^{-1}(B) \in \mathcal{B}(\overline{\mathbb{R}}^d\backslash\{\mathbf{0}\})$ the conclusion follows. □

We proceed by considering sums of a fixed number of independent regularly varying random vectors.

Proposition 3.11. *Let* $\mathbf{X}$ *be an* $\mathbb{R}^d$*-valued random vector and suppose that there exist a sequence* $\{a_n\}, 0 < a_n \uparrow \infty$, *and a Radon measure* μ *on* $\mathcal{B}(\overline{\mathbb{R}}^d\backslash\{\mathbf{0}\})$ *with* $\mu(\overline{\mathbb{R}}^d\backslash\mathbb{R}^d) = 0$ *such that* $n\,\mathbb{P}(a_n^{-1}\mathbf{X} \in \cdot\,) \xrightarrow{v} \mu(\cdot)$ *on* $\mathcal{B}(\overline{\mathbb{R}}^d\backslash\{\mathbf{0}\})$.

(i) If $\widetilde{\mathbf{X}}$ *is a random vector in* $\mathbb{R}^d$, *independent of* $\mathbf{X}$, *and if there exists a Radon measure* $\widetilde{\mu}$ *on* $\mathcal{B}(\overline{\mathbb{R}}^d\backslash\{\mathbf{0}\})$ *with* $\widetilde{\mu}(\overline{\mathbb{R}}^d\backslash\mathbb{R}^d) = 0$ *such that* $n\,\mathbb{P}(a_n^{-1}\widetilde{\mathbf{X}} \in \cdot\,) \xrightarrow{v} \widetilde{\mu}(\cdot)$ *on* $\mathcal{B}(\overline{\mathbb{R}}^d\backslash\{\mathbf{0}\})$, *then*

$$n\,\mathbb{P}(a_n^{-1}(\mathbf{X} + \widetilde{\mathbf{X}}) \in \cdot\,) \xrightarrow{v} \mu(\cdot) + \widetilde{\mu}(\cdot) \quad \text{on } \mathcal{B}(\overline{\mathbb{R}}^d\backslash\{\mathbf{0}\}).$$

(ii) If for some k, *there are iid random vectors* $\mathbf{X}_1, \ldots, \mathbf{X}_k$ *such that* $\mathbf{X} \stackrel{\mathrm{d}}{=} \mathbf{X}_1 + \cdots + \mathbf{X}_k$, *then*

$$n\,\mathbb{P}(a_n^{-1}\mathbf{X}_1 \in \cdot\,) \xrightarrow{v} \frac{1}{k}\mu(\cdot) \quad \text{on } \mathcal{B}(\overline{\mathbb{R}}^d\backslash\{\mathbf{0}\}).$$

Remark 3.12. A statement similar to (i), for $\mathbb{R}_+^d$-valued random vectors, is proved in Resnick [12] (Proposition 4.1 p. 85).

Proof. (i) Take $\epsilon > 0$ and note that, by Remark 3.3 (i), $\mu(\partial B_{\mathbf{0},\epsilon}) = \widetilde{\mu}(\partial B_{\mathbf{0},\epsilon}) = 0$. Since

$$n\,\mathbb{P}(a_n^{-1}(\mathbf{X},\widetilde{\mathbf{X}}) \in B_{\mathbf{0},\epsilon}^c \times B_{\mathbf{0},\epsilon}^c) = n\,\mathbb{P}(a_n^{-1}\mathbf{X} \in B_{\mathbf{0},\epsilon}^c)\,\mathbb{P}(a_n^{-1}\widetilde{\mathbf{X}} \in B_{\mathbf{0},\epsilon}^c) \to 0,$$

as $n \to \infty$, it follows that $n\,\mathbb{P}(a_n^{-1}(\mathbf{X},\widetilde{\mathbf{X}}) \in \cdot\,) \xrightarrow{v} \widehat{\mu}(\cdot)$ on $\mathcal{B}(\overline{\mathbb{R}}^{2d}\backslash\{\mathbf{0}\})$, where $\widehat{\mu}$ is a Radon measure which concentrates on $(\{\mathbf{0}\} \times \mathbb{R}^d) \cup (\mathbb{R}^d \times \{\mathbf{0}\})$. Let $T : \mathbb{R}^{2d} \to \mathbb{R}^d$ be the linear transformation $T(\mathbf{x},\widetilde{\mathbf{x}}) = \mathbf{x} + \widetilde{\mathbf{x}}$. By Proposition 3.9,

$$n\,\mathbb{P}(a_n^{-1}(\mathbf{X} + \widetilde{\mathbf{X}}) \in \cdot\,) \xrightarrow{v} \widehat{\mu} \circ T^{-1}(\cdot \cap \mathbb{R}^d) \quad \text{on } \mathcal{B}(\overline{\mathbb{R}}^d\backslash\{\mathbf{0}\}).$$

Hence for any $B \in \mathcal{B}(\overline{\mathbb{R}}^d\backslash\{\mathbf{0}\})$,

$$\begin{aligned}
\widehat{\mu} \circ T^{-1}(B \cap \mathbb{R}^d) &= \widehat{\mu}(\{(\mathbf{x},\widetilde{\mathbf{x}}) : \mathbf{x} + \widetilde{\mathbf{x}} \in B \cap \mathbb{R}^d\}) \\
&= \widehat{\mu}(\{(\mathbf{x},\mathbf{0}) : \mathbf{x} + \mathbf{0} \in B \cap \mathbb{R}^d\}) + \widehat{\mu}(\{(\mathbf{0},\widetilde{\mathbf{x}}) : \mathbf{0} + \widetilde{\mathbf{x}} \in B \cap \mathbb{R}^d\}) \\
&= \mu(B) + \widetilde{\mu}(B).
\end{aligned}$$

(ii) Since $n\,\mathbb{P}(a_n^{-1}\mathbf{X} \in \cdot\,) \xrightarrow{v} \mu(\cdot)$ on $\mathcal{B}(\overline{\mathbb{R}}^d\backslash\{\mathbf{0}\})$ it follows that for any subsequence $\{n_j\} \subset \mathbb{N}$ such that $n_j \to \infty$ as $j \to \infty$, $n_j\,\mathbb{P}(a_{n_j}^{-1}\mathbf{X} \in \cdot\,) \xrightarrow{v} \mu(\cdot)$ on $\mathcal{B}(\overline{\mathbb{R}}^d\backslash\{\mathbf{0}\})$. Hence, it follows by (i) that any subsequential vague limit μ_1 of $\{n\,\mathbb{P}(a_n^{-1}\mathbf{X}_1 \in \cdot\,)\}$ must satisfy $\mu_1 = \mu/k$. Hence, we only need to show that $\{n\,\mathbb{P}(a_n^{-1}\mathbf{X}_1 \in \cdot\,)\}$ is relatively compact in the vague topology. By Theorem 15.7.5 in Kallenberg [9], $\{n\,\mathbb{P}(a_n^{-1}\mathbf{X}_1 \in \cdot\,)\}$ is relatively compact in the vague topology if and only if $\sup_{n\geq 1} n\,\mathbb{P}(a_n^{-1}\mathbf{X}_1 \in B) < \infty$ for every relatively compact $B \in \mathcal{B}(\overline{\mathbb{R}}^d\backslash\{\mathbf{0}\})$. We prove this by contradiction. Suppose that there is a relatively compact set $B \in \mathcal{B}(\overline{\mathbb{R}}^d\backslash\{\mathbf{0}\})$ such that $\sup_{n\geq 1} n\,\mathbb{P}(a_n^{-1}\mathbf{X}_1 \in B) = \infty$. Then there is an $r > 0$ such that $\sup_{n\geq 1} n\,\mathbb{P}(a_n^{-1}\mathbf{X}_1 \in B_{\mathbf{0},r}^c) = \infty$. Since $n\,\mathbb{P}(a_n^{-1}\mathbf{X}_1 \in B_{\mathbf{0},r}^c) < \infty$ for all finite n this implies $\limsup_{n\to\infty} n\,\mathbb{P}(a_n^{-1}\mathbf{X}_1 \in B_{\mathbf{0},r}^c) = \infty$. Take $\epsilon \in (0, r/k)$. Then

$$\begin{aligned}
\mathbb{P}(a_n^{-1}\mathbf{X} \in B_{\mathbf{0},r-k\epsilon}^c) &= \mathbb{P}(a_n^{-1}(\mathbf{X}_1 + \cdots + \mathbf{X}_k) \in B_{\mathbf{0},r-k\epsilon}^c) \\
&\geq \mathbb{P}(a_n^{-1}\mathbf{X}_1 \in B_{\mathbf{0},r}^c, a_n^{-1}\mathbf{X}_j \in B_{\mathbf{0},\epsilon} \text{ for } j = 2,\ldots,k) \\
&= \mathbb{P}(a_n^{-1}\mathbf{X}_1 \in B_{\mathbf{0},r}^c)\,\mathbb{P}(a_n^{-1}\mathbf{X}_1 \in B_{\mathbf{0},\epsilon})^{k-1}.
\end{aligned}$$

Hence

$$\begin{aligned}
\limsup_{n\to\infty} n\,\mathbb{P}(a_n^{-1}\mathbf{X} \in B_{\mathbf{0},r-k\epsilon}^c) &\geq \limsup_{n\to\infty} n\,\mathbb{P}(a_n^{-1}\mathbf{X}_1 \in B_{\mathbf{0},r}^c)\,\mathbb{P}(a_n^{-1}\mathbf{X}_1 \in B_{\mathbf{0},\epsilon})^{k-1} \\
&= \infty.
\end{aligned}$$

This contradicts the assumption that $\{n\,\mathbb{P}(a_n^{-1}\mathbf{X} \in \cdot\,)\}$ is relatively compact and we conclude that $\{n\,\mathbb{P}(a_n^{-1}\mathbf{X}_1 \in \cdot\,)\}$ must be relatively compact in the vague topology. □

Above we considered a sum of a fixed number of terms. Let us now consider the case with a random number of terms N, where N is independent of the terms $\mathbf{X}_k$.

Proposition 3.13. *Let $\{\mathbf{X}_k\}_{k\geq 1}$ be a sequence of iid $\mathbb{R}^d$-valued random vectors and suppose that there exist a sequence $\{a_n\}, 0 < a_n \uparrow \infty$, and a nonzero Radon measure μ on $\mathcal{B}(\overline{\mathbb{R}}^d\backslash\{\mathbf{0}\})$ with $\mu(\overline{\mathbb{R}}^d\backslash\mathbb{R}^d) = 0$ such that $n\,\mathbb{P}(a_n^{-1}\mathbf{X}_1 \in \cdot\,) \xrightarrow{v} \mu(\cdot)$ on $\mathcal{B}(\overline{\mathbb{R}}^d\backslash\{\mathbf{0}\})$. If N is a nonnegative random variable with $\sum_{n=0}^{\infty}\mathbb{P}(N=n)(1+\epsilon)^n < \infty$ for some $\epsilon > 0$ and N is independent of $\{\mathbf{X}_k\}_{k\geq 1}$, then, as $n\to\infty$,*

$$n\,\mathbb{P}(a_n^{-1}\sum_{k=1}^{N}\mathbf{X}_k \in \cdot\,) \xrightarrow{v} \mathrm{E}(N)\mu(\cdot) \quad \text{on } \mathcal{B}(\overline{\mathbb{R}}^d\backslash\{\mathbf{0}\}).$$

If $\{\mathbf{Y}_k\}_{k\geq 1}$ is another sequence of iid $\mathbb{R}^d$-valued random vectors, independent of N, such that $n\,\mathbb{P}(a_n^{-1}\mathbf{Y}_1 \in \cdot\,) \xrightarrow{v} 0$ as $n\to\infty$, then $n\,\mathbb{P}(a_n^{-1}\sum_{k=1}^{N}\mathbf{Y}_k \in \cdot\,) \xrightarrow{v} 0$ as $n\to\infty$.

Proof. Take relatively compact $B \in \mathcal{B}(\overline{\mathbb{R}}^d\backslash\{\mathbf{0}\})$ with $\mu(\partial B) = 0$. Since for all n, $n\,\mathbb{P}(a_n^{-1}\sum_{k=1}^{N}\mathbf{X}_k \in \overline{\mathbb{R}}^d\backslash\mathbb{R}^d) = \mathrm{E}(N)\mu(\overline{\mathbb{R}}^d\backslash\mathbb{R}^d) = 0$, we may without loss of generality assume that $B \subset \mathbb{R}^d$. Let $\gamma = \inf_{\mathbf{x}\in B}|\mathbf{x}|$. Since

$$\begin{aligned} n\,\mathbb{P}(a_n^{-1}\sum_{k=1}^{N}\mathbf{X}_k \in B) &= \sum_{l=1}^{\infty} n\,\mathbb{P}(a_n^{-1}\sum_{k=1}^{l}\mathbf{X}_k \in B)\,\mathbb{P}(N=l) \\ &\leq \sum_{l=1}^{\infty} n\,\mathbb{P}(a_n^{-1}\sum_{k=1}^{l}|\mathbf{X}_k| > \gamma)\,\mathbb{P}(N=l) \qquad (3.10) \end{aligned}$$

and $|\mathbf{X}_k|$ is univariate regularly varying, it follows from Theorem 3 in Embrechts, Goldie and Veraverbeke [7] that (3.10) converges to $\mathrm{E}(N)\mu(B^c_{\mathbf{0},\gamma})$. Hence, using Pratt's Theorem (see Pratt [11]) and Proposition 3.11, we conclude that we may interchange the sum and the limit to obtain

$$\begin{aligned} \lim_{n\to\infty} n\,\mathbb{P}(a_n^{-1}\sum_{k=1}^{N}\mathbf{X}_k \in B) &= \lim_{n\to\infty}\sum_{l=1}^{\infty} n\,\mathbb{P}(a_n^{-1}\sum_{k=1}^{l}\mathbf{X}_k \in B)\,\mathbb{P}(N=l) \\ &= \sum_{l=1}^{\infty}\lim_{n\to\infty} n\,\mathbb{P}(a_n^{-1}\sum_{k=1}^{l}\mathbf{X}_k \in B)\,\mathbb{P}(N=l) \\ &= \mathrm{E}(N)\mu(B) \end{aligned}$$

for every relatively compact $B \in \mathcal{B}(\overline{\mathbb{R}}^d\backslash\{\mathbf{0}\})$ with $\mu(\partial B) = 0$. The second claim is proved similarly. □

Remark 3.14. The moment condition on N, $\sum_{n=0}^{\infty}\mathbb{P}(N=n)(1+\epsilon)^n < \infty$ for some $\epsilon > 0$, comes from the remark following Theorem 3 in Embrechts, Goldie and

Veraverbeke [7]. This result concerns random sums of univariate random variables with subexponential tails. Since we use this result for terms which are regularly varying (at ∞) the moment condition can in fact be substantially weakened.

We conclude this section with a short lemma. It can be used for instance in connection with Proposition 3.11 to show that if we add any random vector with all moments finite to an independent regularly varying random vector $\mathbf{X}$ with limit measure μ, then the sum is regularly varying with the same limit measure μ.

Lemma 3.15. *Let $\mathbf{X}$ be an $\mathbb{R}^d$-valued random vector. If $\mathrm{E}(|\mathbf{X}|^m) < \infty$ for every $m \in \mathbb{N}$, then for every regularly varying sequence $\{a_n\}$, $0 < a_n \uparrow \infty$, and every relatively compact set $B \in \mathcal{B}(\overline{\mathbb{R}}^d \backslash \{0\})$, $n\,\mathbb{P}(a_n^{-1}\mathbf{X} \in B) \to 0$ as $n \to \infty$.*

Proof. Fix $\alpha > 0$ and let $\{a_n\}$, $0 < a_n \uparrow \infty$, be a regularly varying sequence with index $1/\alpha$. Let $B \in \mathcal{B}(\overline{\mathbb{R}}^d \backslash \{\mathbf{0}\})$ be relatively compact. Then there is an $x > 0$ such that $B \subset \overline{\mathbb{R}}^d \backslash \overline{B}_{\mathbf{0},x}$. Hence

$$n\,\mathbb{P}(a_n^{-1}\mathbf{X} \in B) \leq n\,\mathbb{P}(a_n^{-1}\mathbf{X} \in \overline{\mathbb{R}}^d \backslash \overline{B}_{\mathbf{0},x}) = n\,\mathbb{P}(|\mathbf{X}| > a_n x).$$

Define f by $f(t) = \inf\{n \in \mathbb{N} : a_n > t\}$. Then by Theorem 1.5.12 p. 28 in Bingham, Goldie and Teugels [4] f is regularly varying with index α, i.e. $f(t) = t^\alpha L(t)$ for some slowly varying function L, and $f(a_n) \sim n$ as $n \to \infty$. Hence

$$n\,\mathbb{P}(|\mathbf{X}| > a_n x) \sim f(a_n)\,\mathbb{P}(|\mathbf{X}| > a_n x) \quad \text{as } n \to \infty.$$

By Markov's inequality we have that $\mathbb{P}(|\mathbf{X}| > tx) \leq \mathrm{E}(|\mathbf{X}|^m)/(tx)^m < \infty$ for every $t > 0$, $x > 0$ and $m > 0$. Hence, for every $m > 0$,

$$f(a_n)\,\mathbb{P}(|\mathbf{X}| > a_n x) \leq a_n^\alpha L(a_n) a_n^{-m} x^{-m} \mathrm{E}(|\mathbf{X}|^m) = C(m) a_n^{\alpha - m} L(a_n).$$

Taking $m > \alpha$ and letting $n \to \infty$ yields $n\,\mathbb{P}(|\mathbf{X}| > a_n x) \to 0$ from which the conclusion follows. □

3.5 Multivariate regular variation for additive processes and functionals

We now turn to the main topic of this paper; the tail behavior of multivariate additive processes and functionals of them. In the univariate case it is well known (see e.g. Embrechts, Goldie and Veraverbeke [7]) that for an infinitely divisible regularly varying (or even subexponential) random variable X with Lévy measure ν it holds that $\mathbb{P}(X > u) \sim \nu(\{x : x > u\})$ as $u \to \infty$. This property is sometimes referred to as *tail equivalence* of X and ν. For a univariate additive process $\{X_t : t \geq 0\}$ with sequence of generating triplets $\{(A_t, \nu_t, \gamma_t)\}$ this implies that if X_t is regularly varying (subexponential) for some $t > 0$, then $\mathbb{P}(X_t > u) \sim \nu_t(\{x : x >$

$u\}$) as $u \to \infty$. The intuition behind this result is explained by what is sometimes referred to as the "large deviations approach": unlikely events happen in the most likely way. In this case this is interpreted as follows. The most likely way the process becomes large at time t is due to one big jump of the process before time t. The Lévy measure of $\{x : x > u\}$ gives the intensity of jumps to this set during $(0, t]$ and the probability of exactly one jump to this set is asymptotically $\nu_t(\{x : x > u\})$. The same intuition holds also in the multivariate case as the next result shows. The tail equivalence should now be interpreted as equality of the limiting measures associated with multivariate regular variation.

Theorem 3.16. *Let $\{\mathbf{X}_t : t \geq 0\}$ be an additive process on $\mathbb{R}^d$ with system of generating triplets $\{(A_t, \nu_t, \boldsymbol{\gamma}_t)\}$. Fix an arbitrary $t > 0$. Then the following statements are equivalent.*

(i) There exist a sequence $\{a_n\}$, $0 < a_n \uparrow \infty$, and a nonzero Radon measure μ_t on $\mathcal{B}(\overline{\mathbb{R}}^d \backslash \{\mathbf{0}\})$ with $\mu_t(\overline{\mathbb{R}}^d \backslash \mathbb{R}^d) = 0$ such that

$$n\,\mathbb{P}(a_n^{-1}\mathbf{X}_t \in \cdot\,) \xrightarrow{v} \mu_t(\cdot) \quad \text{on } \mathcal{B}(\overline{\mathbb{R}}^d \backslash \{\mathbf{0}\}). \tag{3.11}$$

(ii) There exist a sequence $\{a_n\}$, $0 < a_n \uparrow \infty$, and a nonzero Radon measure μ_t on $\mathcal{B}(\overline{\mathbb{R}}^d \backslash \{\mathbf{0}\})$ with $\mu_t(\overline{\mathbb{R}}^d \backslash \mathbb{R}^d) = 0$ such that

$$n\nu_t(a_n \cdot\,) \xrightarrow{v} \mu_t(\cdot) \quad \text{on } \mathcal{B}(\overline{\mathbb{R}}^d \backslash \{\mathbf{0}\}). \tag{3.12}$$

Furthermore, the sequences $\{a_n\}$ and the measures μ_t in (i) and (ii) can be taken to be equal.

Remark 3.17. (i) Note that for any $t > 0$ and any infinitely divisible random vector $\mathbf{Y}$ there exists an additive process $\{\mathbf{X}_t : t \geq 0\}$ such that $\mathbf{Y} \stackrel{\mathrm{d}}{=} \mathbf{X}_t$. Hence Theorem 3.16 can be reformulated in terms of infinitely divisible random vectors.
(ii) Note also that if Theorem 3.4 (i) were true for arbitrary tails indices $\alpha > 0$, then Theorem 3.16 would be easily proved using the results in Embrechts, Goldie and Veraverbeke [7] for univariate subexponential infinitely divisible random variables.

Proof of Theorem 3.16. To prove this theorem we will make use of the Lévy-Itô decomposition, Theorem 3.7, which says that $\mathbf{X}_t$ has representation

$$\mathbf{X}_t = \mathbf{Y}_t + \mathbf{J}_t + \mathbf{X}_t^2 \quad \text{a.s.},$$

where $\mathbf{Y}_t$, $\mathbf{J}_t$ and $\mathbf{X}_t^2$ are independent and, with the notation of Theorem 3.7,

$$\begin{aligned}
\mathbf{Y}_t &= \lim_{\epsilon \downarrow 0} \int_{(0,t] \times D_{\epsilon,1}} \{\mathbf{x}\xi(\mathrm{d}(s, \mathbf{x})) - \mathbf{x}\widetilde{\nu}(\mathrm{d}(s, \mathbf{x}))\}, \\
\mathbf{J}_t &= \int_{(0,t] \times D_{1,\infty}} \mathbf{x}\xi(\mathrm{d}(s, \mathbf{x})), \\
\mathbf{X}_t^2 &= \mathbf{X}_t - \mathbf{Y}_t - \mathbf{J}_t.
\end{aligned}$$

$\mathbf{X}_t^2$ is Gaussian and hence has finite moments of all orders. By Lemma 3.8, $\mathbf{Y}_t$ has finite moments of all orders and will therefore, as we will see, not contribute to $\mathbf{X}_t$ being regularly varying. It will be sufficient to consider the part $\mathbf{J}_t$ being the accumulated big jumps of the process up to time t. Since ξ is a Poisson random measure, the characteristic function of $\mathbf{J}_t$ is

$$\mathrm{E}[e^{i\langle \mathbf{z}, \mathbf{J}_t\rangle}] = \exp\Big\{ \int_{D_{1,\infty}} (e^{i\langle \mathbf{z},\mathbf{x}\rangle} - 1)\nu_t(\mathrm{d}\mathbf{x})\Big\}$$

and hence $\mathbf{J}_t$ has representation as a compound Poisson random vector

$$\mathbf{J}_t \stackrel{\mathrm{d}}{=} \sum_{k=0}^{N_t} \mathbf{J}_{k,t},$$

where $N_t \sim \mathrm{Po}(\nu_t(D_{1,\infty}))$, $\mathbf{J}_{0,t} = 0$, $\mathbf{J}_{k,t} \sim \nu_t(\cdot \cap D_{1,\infty})/\nu_t(D_{1,\infty})$ and all vectors are independent.

(i) $\Rightarrow$ (ii) To prove this implication we will prove that

$$n\,\mathbb{P}(a_n^{-1}\mathbf{J}_{1,t} \in \cdot\,) \stackrel{v}{\to} \mu_t(\cdot)/\nu_t(D_{1,\infty}) \quad \text{as } n \to \infty. \tag{3.13}$$

Once this is done the implication follows since $\mathbf{J}_{k,t} \sim \nu_t(\cdot \cap D_{1,\infty})/\nu_t(D_{1,\infty})$ and for any relatively compact $B \in \mathcal{B}(\overline{\mathbb{R}}^d\backslash\{0\})$ and large enough n, $\nu_t(\{a_nB\} \cap D_{1,\infty}) = \nu_t(a_nB)$. So if (3.13) holds, then also

$$n\nu_t(a_n\cdot\,) \stackrel{v}{\to} \mu_t(\cdot) \quad \text{as } n \to \infty.$$

We prove this part in two steps. First we show that for (3.11) to hold it is necessary that

$$n\,\mathbb{P}(a_n^{-1}\mathbf{J}_t \in \cdot\,) \stackrel{v}{\to} \mu_t(\cdot) \quad \text{as } n \to \infty, \tag{3.14}$$

holds and for (3.14) to hold it is necessary that (3.13) holds. Let us start by showing that the sequence $\{n\,\mathbb{P}(a_n^{-1}\mathbf{J}_t \in \cdot\,)\}$ is relatively compact in the vague topology. As in the proof of Proposition 3.11 we prove this by contradiction. Assume that there exists a relatively compact $B \in \mathcal{B}(\overline{\mathbb{R}}^d\backslash\{0\})$ such that $\sup_{n\geq 1} n\,\mathbb{P}(a_n^{-1}\mathbf{J}_t \in B) = \infty$. Then there is an $r > 0$ such that $B \subset B_{0,r}^c$ and hence $\sup_{n\geq 1} n\,\mathbb{P}(a_n^{-1}\mathbf{J}_t \in B_{0,r}^c) = \infty$. Since $n\,\mathbb{P}(a_n^{-1}\mathbf{J}_t \in B_{0,r}^c)$ is finite for every n, $\limsup_{n\to\infty} n\,\mathbb{P}(a_n^{-1}\mathbf{J}_t \in B_{0,r}^c) = \infty$. Take $\epsilon \in (0, r/3)$. Then

$$\begin{aligned}
&\limsup_{n\to\infty}\ n\,\mathbb{P}(a_n^{-1}\mathbf{X}_t \in B_{0,r-2\epsilon}^c) = \limsup_{n\to\infty} n\,\mathbb{P}(a_n^{-1}(\mathbf{Y}_t + \mathbf{J}_t + \mathbf{X}_t^2) \in B_{0,r-2\epsilon}^c)\\
&\geq \limsup_{n\to\infty} n\,\mathbb{P}(a_n^{-1}\mathbf{J}_t \in B_{0,r}^c)\,\mathbb{P}(a_n^{-1}\mathbf{Y}_t \in B_{0,\epsilon})\,\mathbb{P}(a_n^{-1}\mathbf{X}_t^2 \in B_{0,\epsilon}) = \infty.
\end{aligned}$$

Thus $\{n\,\mathbb{P}(a_n^{-1}\mathbf{X}_t \in \cdot\,)\}$ is not relatively compact which is a contradiction. We conclude that $\{n\,\mathbb{P}(a_n^{-1}\mathbf{J}_t \in \cdot\,)\}$ is relatively compact. Let $\{n_i\}$ be a subsequence

such that $\lim_{i\to\infty} n_i = \infty$ and let $\mu_{1,t}$ be the vague limit of $\{n_i\,\mathbb{P}(a_{n_i}^{-1}\mathbf{J}_t \in \cdot\,)\}$ as $i \to \infty$. Since $\mathbf{Y}_t$ and $\mathbf{X}_t^2$ have finite moments of all orders, by Lemma 3.15, for every relatively compact $B \in \mathcal{B}(\overline{\mathbb{R}}^d\backslash\{0\})$,

$$n_i\,\mathbb{P}(a_{n_i}^{-1}\mathbf{Y}_t \in B) \to 0 \quad \text{and} \quad n_i\,\mathbb{P}(a_{n_i}^{-1}\mathbf{X}_t^2 \in B) \to 0 \quad \text{as } i \to \infty.$$

Then, by Proposition 3.11 (i), $n_i\,\mathbb{P}(a_{n_i}^{-1}\mathbf{X}_t \in \cdot\,) \xrightarrow{v} \mu_{1,t}(\cdot)$ as $i \to \infty$. However, we have assumed that (3.11) holds so $\mu_{1,t} = \mu_t$. Hence, (3.14) holds.

We continue with a similar argument to show that (3.13) holds. Let us start by showing that the sequence $\{n\,\mathbb{P}(a_n^{-1}\mathbf{J}_{1,t} \in \cdot\,)\}$ is relatively compact in the vague topology. Assume that there exists a relatively compact $B \in \mathcal{B}(\overline{\mathbb{R}}^d\backslash\{0\})$ such that $\sup_{n\geq 1} n\,\mathbb{P}(a_n^{-1}\mathbf{J}_{1,t} \in B) = \infty$. Then there is an $r > 0$ such that $B \subset B_{0,r}^c$ and hence $\sup_{n\geq 1} n\,\mathbb{P}(a_n^{-1}\mathbf{J}_{1,t} \in B_{0,r}^c) = \infty$. Since $n\,\mathbb{P}(a_n^{-1}\mathbf{J}_{1,t} \in B_{0,r}^c)$ is finite for every n, $\limsup_{n\to\infty} n\,\mathbb{P}(a_n^{-1}\mathbf{J}_{1,t} \in B_{0,r}^c) = \infty$. We have

$$\limsup_{n\to\infty} n\,\mathbb{P}(a_n^{-1}\mathbf{J}_t \in B_{0,r}^c) \geq \limsup_{n\to\infty} n\,\mathbb{P}(a_n^{-1}\mathbf{J}_{1,t} \in B_{0,r}^c)\,\mathbb{P}(N_t = 1) = \infty.$$

Thus $\{n\,\mathbb{P}(a_n^{-1}\mathbf{J}_t \in \cdot\,)\}$ is not relatively compact which is a contradiction. We conclude that $\{n\,\mathbb{P}(a_n^{-1}\mathbf{J}_{1,t} \in \cdot\,)\}$ is relatively compact. Let $\{n_j\}$ be a subsequence such that $\lim_{j\to\infty} n_j = \infty$ and let $\mu_{2,t}$ be the vague limit of $\{n_j\,\mathbb{P}(a_{n_j}^{-1}\mathbf{J}_{1,t} \in \cdot\,)\}$ as $j \to \infty$. By Proposition 3.13 it follows that

$$n_j\,\mathbb{P}(a_{n_j}^{-1}\mathbf{J}_t \in \cdot\,) \xrightarrow{v} \mathbb{E}(N_t)\mu_{2,t}(\cdot) = \nu_t(D_{1,\infty})\mu_{2,t}(\cdot) \quad \text{as } j \to \infty$$

and hence we must have $\mu_{2,t} = \mu_t/\nu_t(D_{1,\infty})$. Hence, (3.13) holds.

(ii) $\Rightarrow$ (i) Since $\mathbf{J}_{k,t} \sim \nu_t(\cdot \cap D_{1,\infty})/\nu_t(D_{1,\infty})$ and we have assumed that $n\nu_t(a_n\cdot\,) \xrightarrow{v} \mu_t(\cdot)$ as $n \to \infty$, it follows that

$$n\,\mathbb{P}(a_n^{-1}\mathbf{J}_{k,t} \in \cdot\,) \xrightarrow{v} \mu_t(\cdot)/\nu_t(D_{1,\infty}) \quad \text{as } n \to \infty$$

since for any relatively compact $B \in \mathcal{B}(\overline{\mathbb{R}}^d\backslash\{0\})$ and large enough n, $\nu_t(\{a_nB\} \cap D_{1,\infty}) = \nu_t(a_nB)$. Then, by Proposition 3.13,

$$n\,\mathbb{P}(a_n^{-1}\mathbf{J}_t \in \cdot\,) \xrightarrow{v} \mathbb{E}(N_t)\mu_t(\cdot)/\nu_t(D_{1,\infty}) = \mu_t(\cdot) \quad \text{as } n \to \infty.$$

Now, since $\mathbf{X}_t = \mathbf{Y}_t + \mathbf{J}_t + \mathbf{X}_t^2$ a.s. where the terms are independent, $\mathbf{Y}_t$ and $\mathbf{X}_t^2$ have finite moments of all orders and $n\,\mathbb{P}(a_n^{-1}\mathbf{J}_t \in \cdot\,) \xrightarrow{v} \mu_t(\cdot)$ as $n \to \infty$, the conclusion follows by combining Lemma 3.15 and the first part of Proposition 3.11. □

The next result (essentially a multivariate version of the result in Willekens [16]) will be of relevance for the subsequent study of functionals of additive processes, but is also interesting in itself. It says essentially the following. If the tails of the process are sufficiently heavy, then the probability that the process reaches a set far away from the origin before time $t > 0$ is asymptotically equal to the probability that the process ends up in that set at time t. Note that condition (3.15) is much weaker than that of multivariate regular variation.

Lemma 3.18. *Let $\{\mathbf{X}_t : t \geq 0\}$ be an additive process on $\mathbb{R}^d$ and let $A \in \mathcal{R}^d$ be bounded away from $\mathbf{0}$. For $z > 0$ let $T_z^A = \inf\{s : \mathbf{X}_s \in zA\}$. Fix $t > 0$, let $r > 0$ be arbitrary and put $S_r(zA) = \{\mathbf{x} \in \mathbb{R}^d : \inf_{\mathbf{y} \in zA} |\mathbf{x} - \mathbf{y}| < r\}$. Then*

(i) $\mathbb{P}(T_z^A \leq t)\, \mathbb{P}(\mathbf{X}_s \in B_{\mathbf{0},r/2} \text{ all } s \leq t) \leq \mathbb{P}(\mathbf{X}_t \in S_r(zA))$.

(ii) If

$$\lim_{z \to \infty} \frac{\mathbb{P}(\mathbf{X}_t \in S_r(zA))}{\mathbb{P}(\mathbf{X}_t \in zA)} = 1, \tag{3.15}$$

then

$$\lim_{z \to \infty} \frac{\mathbb{P}(T_z^A \leq t)}{\mathbb{P}(\mathbf{X}_t \in zA)} = 1.$$

Proof.

$$\begin{aligned}
\mathbb{P}(T_z^A \leq t) &= \mathbb{P}(T_z^A \leq t, \mathbf{X}_t \in S_r(zA)) + \mathbb{P}(T_z^A \leq t, \mathbf{X}_t \in S_r(zA)^c) \\
&\leq \mathbb{P}(\mathbf{X}_t \in S_r(zA)) + \mathbb{P}(T_z^A \leq t, \mathbf{X}_t \in S_r(zA)^c) \\
&\leq \mathbb{P}(\mathbf{X}_t \in S_r(zA)) + \mathbb{P}(T_z^A \leq t, \mathbf{X}_t - \mathbf{X}_{T_z^A} \in B_{\mathbf{0},r}^c) \\
&= \mathbb{P}(\mathbf{X}_t \in S_r(zA)) + \mathbb{P}(T_z^A \leq t)\, \mathbb{P}(\mathbf{X}_t - \mathbf{X}_{T_z^A} \in B_{\mathbf{0},r}^c \mid T_z^A \leq t) \\
&\leq \mathbb{P}(\mathbf{X}_t \in S_r(zA)) + \mathbb{P}(T_z^A \leq t)\, \mathbb{P}(\mathbf{X}_s \in B_{\mathbf{0},r/2}^c \text{ some } s \leq t) \\
&= \mathbb{P}(\mathbf{X}_t \in S_r(zA)) + \mathbb{P}(T_z^A \leq t)(1 - \mathbb{P}(\mathbf{X}_s \in B_{\mathbf{0},r/2} \text{ all } s \leq t))
\end{aligned}$$

Hence $\mathbb{P}(T_z^A \leq t)\, \mathbb{P}(\mathbf{X}_s \in B_{\mathbf{0},r/2} \text{ all } s \leq t) \leq \mathbb{P}(\mathbf{X}_t \in S_r(zA))$. It follows that

$$\begin{aligned}
1 &\leq \liminf_{z \to \infty} \frac{\mathbb{P}(T_z^A \leq t)}{\mathbb{P}(\mathbf{X}_t \in zA)} \leq \limsup_{z \to \infty} \frac{\mathbb{P}(T_z^A \leq t)}{\mathbb{P}(\mathbf{X}_t \in zA)} \\
&\leq \limsup_{z \to \infty} \frac{\mathbb{P}(\mathbf{X}_t \in S_r(zA))}{\mathbb{P}(\mathbf{X}_t \in zA)} \cdot \frac{1}{\mathbb{P}(\mathbf{X}_s \in B_{\mathbf{0},r/2} \text{ all } s \leq t)} \\
&= \frac{1}{\mathbb{P}(\mathbf{X}_s \in B_{\mathbf{0},r/2} \text{ all } s \leq t)}.
\end{aligned}$$

Since $r > 0$ was arbitrary we can let $r \to \infty$ and the conclusion follows. □

Let us now study vectors of functionals applied to each component of a multivariate additive process. We consider the implications of regular variation of the process at time t on the vector of the componentwise suprema of the process and the componentwise suprema of its jumps up to time t. We have the following result.

Theorem 3.19. *Let $\{\mathbf{X}_t : t \geq 0\}$ be an additive process on $\mathbb{R}^d$ with system of generating triplets $\{(A_t, \nu_t, \boldsymbol{\gamma}_t)\}$. Fix an arbitrary $t > 0$. Then the following statements are equivalent.*

(i) There exist a sequence $\{a_n\}$, $0 < a_n \uparrow \infty$, and a nonzero Radon measure μ_t on $\mathcal{B}(\overline{\mathbb{R}}^d \backslash \{\mathbf{0}\})$ with $\mu_t(\overline{\mathbb{R}}^d \backslash \mathbb{R}^d) = 0$ such that

$$n\, \mathbb{P}(a_n^{-1} \mathbf{X}_t \in \cdot\,) \xrightarrow{v} \mu_t(\cdot) \quad \text{on } \mathcal{B}(\overline{\mathbb{R}}^d \backslash \{\mathbf{0}\}).$$

(ii) There exist a sequence $\{a_n\}$, $0 < a_n \uparrow \infty$, *and a nonzero Radon measure* μ_t *on* $\mathcal{B}(\overline{\mathbb{R}}^d \backslash \{\mathbf{0}\})$ *with* $\mu_t(\overline{\mathbb{R}}^d \backslash \mathbb{R}^d) = 0$ *such that*

$$n\,\mathbb{P}(a_n^{-1}\mathbf{X}_s \in \cdot \text{ some } s \le t) \xrightarrow{v} \mu_t(\cdot) \quad \text{on } \mathcal{B}(\overline{\mathbb{R}}^d \backslash \{\mathbf{0}\}),$$

The sequences $\{a_n\}$ *and the measures* μ_t *in (i) and (ii) can be taken to be equal. Moreover, if any of the statements (i) or (ii) hold, then*

(iii) $n\,\mathbb{P}(a_n^{-1}\mathbf{X}_t^{\Delta} \in \cdot\,) \xrightarrow{v} \mu_t(\cdot)$ *on* $\mathcal{B}(\overline{\mathbb{R}}_+^d \backslash \{\mathbf{0}\})$,
where $\mathbf{X}_t^{\Delta} = (\sup_{0<s\le t} \Delta X_s^{(1)}, \dots, \sup_{0<s\le t} \Delta X_s^{(d)})$ *and*

(iv) $n\,\mathbb{P}(a_n^{-1}\mathbf{X}_t^{*} \in \cdot\,) \xrightarrow{v} \mu_t(\cdot)$ *on* $\mathcal{B}(\overline{\mathbb{R}}_+^d \backslash \{\mathbf{0}\})$,
where $\mathbf{X}_t^{*} = (\sup_{0\le s\le t} X_s^{(1)}, \dots, \sup_{0\le s\le t} X_s^{(d)})$.

Proof. (i) $\Rightarrow$ (ii) Let $V_{u,S} = \{\mathbf{x} \in \mathbb{R}^d \backslash \{\mathbf{0}\} : |\mathbf{x}| > u, \mathbf{x}/|\mathbf{x}| \in S\}$ for $u > 0$ and $S \in \mathcal{B}(\mathbb{S}^{d-1})$ be a μ_t-continuity set. Suppose first that $\mu_t(V_{u,S}) = 0$. Take $r > 0$ such that $\mathbb{P}(\mathbf{X}_s \in B_{\mathbf{0},r/2} \text{ all } s \le t) > 0$. For every $\delta \in (0,1)$ there is $n_{r,\delta}$ such that for $n > n_{r,\delta}$

$$S_r(a_n V_{u,S}) \subset a_n(1-\delta)V_{u,S} \cup a_n(1-\delta)V_{u,S_\delta(S)\backslash S},$$

where $S_r(a_n V_{u,S}) = \{\mathbf{x} \in \mathbb{R}^d : \inf_{\mathbf{y}\in a_n V_{u,S}} |\mathbf{x}-\mathbf{y}| < r\}$ and $S_\delta(S) = \{\mathbf{x} \in \mathbb{S}^{d-1} : \inf_{\mathbf{y}\in S} |\mathbf{x}-\mathbf{y}| < \delta\}$. Furthermore, $V_{u,S_\delta(S)\backslash S}$ fails to be a μ_t-continuity set for at most countably many δ. Hence, by Lemma 3.18 (i), for $n > n_{r,\delta}$,

$$\begin{aligned}
&n\,\mathbb{P}(a_n^{-1}\mathbf{X}_s \in V_{u,S} \text{ some } s \le t)\\
&\quad\le \frac{n\,\mathbb{P}(a_n^{-1}\mathbf{X}_t \in (1-\delta)V_{u,S})}{\mathbb{P}(\mathbf{X}_s \in B_{\mathbf{0},r/2} \text{ all } s \le t)} + \frac{n\,\mathbb{P}(a_n^{-1}\mathbf{X}_t \in (1-\delta)V_{u,S_\delta(S)\backslash S})}{\mathbb{P}(\mathbf{X}_s \in B_{\mathbf{0},r/2} \text{ all } s \le t)}\\
&\quad\to 0 + \frac{(1-\delta)^{-\alpha}\mu_t(V_{u,S_\delta(S)\backslash S})}{\mathbb{P}(\mathbf{X}_s \in B_{\mathbf{0},r/2} \text{ all } s \le t)}
\end{aligned}$$

for some $\alpha > 0$. Since $\delta \in (0,1)$ was arbitrary, by letting $\delta \to 0$ the conclusion follows.

Now, suppose $\mu_t(V_{u,S}) > 0$. We first show that the condition in Lemma 3.18 (ii) is satisfied. Take $\epsilon > 0$. For every $\delta \in (0,1)$ there is $z_0 > 0$ such that for $z > z_0$

$$S_\epsilon(zV_{u,S}) \subset (1-\delta)zV_{u,S} \cup (1-\delta)zV_{u,S_\delta(S)\backslash S}.$$

Hence

$$\begin{aligned}
1 \;&\le\; \frac{\mathbb{P}(\mathbf{X}_t \in S_\epsilon(zV_{u,S}))}{\mathbb{P}(\mathbf{X}_t \in zV_{u,S})} \le \frac{\mathbb{P}(\mathbf{X}_t \in (1-\delta)zV_{u,S})}{\mathbb{P}(\mathbf{X}_t \in zV_{u,S})} + \frac{\mathbb{P}(\mathbf{X}_t \in (1-\delta)zV_{u,S_\delta(S)\backslash S})}{\mathbb{P}(\mathbf{X}_t \in zV_{u,S})}\\
&\to\; \frac{\mu_t((1-\delta)V_{u,S})}{\mu_t(V_{u,S})} + \frac{\mu_t((1-\delta)V_{u,S_\delta(S)\backslash S})}{\mu_t(V_{u,S})} = (1-\delta)^{-\alpha}\left(1 + \frac{\mu_t(V_{u,S_\delta(S)\backslash S})}{\mu_t(V_{u,S})}\right)
\end{aligned}$$

for some $\alpha > 0$. Since $\delta \in (0,1)$ was arbitrary, by letting $\delta \to 0$ it follows that the condition in Lemma 3.18 is satisfied, and hence, by Lemma 3.18,

$$\lim_{z\to\infty} \frac{\mathbb{P}(\mathbf{X}_s \in zV_{u,S} \text{ some } s \le t)}{\mathbb{P}(\mathbf{X}_t \in zV_{u,S})} = 1.$$

Hence

$$\begin{aligned} n\,\mathbb{P}(\mathbf{X}_s \in a_n V_{u,S} \text{ some } s \le t) &= n\,\mathbb{P}(\mathbf{X}_t \in a_n V_{u,S}) \frac{\mathbb{P}(\mathbf{X}_s \in a_n V_{u,S} \text{ some } s \le t)}{\mathbb{P}(\mathbf{X}_t \in a_n V_{u,S})} \\ &\to \mu_t(V_{u,S}), \end{aligned}$$

as $n \to \infty$. Since convergence of every such set $V_{u,S}$ implies vague convergence on $\mathcal{B}(\overline{\mathbb{R}}^d \backslash \{\mathbf{0}\})$ the conclusion follows.
(ii) $\Rightarrow$ (i) For any $A \in \mathcal{B}(\overline{\mathbb{R}}^d \backslash \{\mathbf{0}\})$ we have $\mathbb{P}(a_n^{-1}\mathbf{X}_t \in A) \le \mathbb{P}(a_n^{-1}\mathbf{X}_s \in A \text{ some } s \le t)$ and hence

$$\limsup_{n\to\infty} n\,\mathbb{P}(a_n^{-1}\mathbf{X}_t \in A) \le \mu_t(A).$$

A lower bound is constructed as follows. Let $V_{u,S} = \{\mathbf{x} \in \mathbb{R}^d \backslash \{\mathbf{0}\} : |\mathbf{x}| > u, \mathbf{x}/|\mathbf{X}| \in S\}$, for $u > 0$ and S of the form $\{\mathbf{x} \in \mathbb{S}^{d-1} : |\mathbf{x} - \mathbf{x}_0| < r_0\}$ for some $\mathbf{x}_0 \in \mathbb{S}^{d-1}$ and $r_0 \in (0,1)$, such that $V_{u,S}$ is a μ_t-continuity set. For such an S and small enough $\delta > 0$, let $S^\delta(S) = \{\mathbf{x} \in S : \inf_{\mathbf{y}\in\mathbb{S}^{d-1}\backslash S} |\mathbf{x} - \mathbf{y}| \ge \delta\}$ and for $r > 0$ let $S^r(a_n V_{u,S}) = \{\mathbf{x} \in a_n V_{u,S} : \inf_{\mathbf{y}\in a_n V_{u,S}^c} |\mathbf{x} - \mathbf{y}| \ge r\}$. For every $\delta \in (0,1)$ there is $n_{r,\delta}$ such that for $n \ge n_{r,\delta}$,

$$S^r(a_n V_{u,S}) \supset (1+\delta) a_n V_{u,S^\delta(S)}.$$

Hence, for $n \ge n_{r,\delta}$,

$$\begin{aligned} &n\,\mathbb{P}(a_n^{-1}\mathbf{X}_t \in V_{u,S}) \\ &\quad \ge n\,\mathbb{P}(a_n^{-1}\mathbf{X}_s \in (1+\delta)V_{u,S^\delta(S)} \text{ some } s \le t, \mathbf{X}_q - \mathbf{X}_s \in a_n B_{\mathbf{0},\delta} \text{ all } s \le q \le t) \\ &\quad \ge n\,\mathbb{P}(a_n^{-1}\mathbf{X}_s \in (1+\delta)V_{u,S^\delta(S)} \text{ some } s \le t)(1 - \mathbb{P}(|\mathbf{X}|_q > a_n\delta \text{ some } 0 \le q \le t)) \\ &\quad \to (1+\delta)^{-\alpha}\mu_t(V_{u,S^\delta(S)}). \end{aligned}$$

Since $\delta > 0$ was arbitrary we can let $\delta \to 0$ and hence

$$\liminf_{n\to\infty} n\,\mathbb{P}(a_n^{-1}\mathbf{X}_t \in V_{u,S}) \ge \mu_t(V_{u,S}).$$

Since convergence of every such set $V_{u,S}$ implies vague convergence on $\mathcal{B}(\overline{\mathbb{R}}^d \backslash \{\mathbf{0}\})$ the conclusion follows.
(i) $\Rightarrow$ (iii) Since $\{\mathbf{X}_s : s \ge 0\}$ is an additive process there is $\Omega_0 \in \mathcal{F}$ with $\mathbb{P}(\Omega_0) = 1$ such that, for every $\omega \in \Omega_0$, $\mathbf{X}_s(\omega)$ is right continuous in $s \ge 0$ and has left limits

in $s > 0$. Hence $\mathbf{X}_s^\Delta \geq 0$ for every $\omega \in \Omega_0$ and $s > 0$. Let ξ be the Poisson random measure given by (3.8). Then, for any $\mathbf{x} \in \mathbb{R}_+^d \backslash \{\mathbf{0}\}$,

$$\mathbb{P}(\mathbf{X}_t^\Delta < \mathbf{x}) = \mathbb{P}(\xi((0,t] \times [-\boldsymbol{\infty}, \mathbf{x})^c) = 0) = \exp(-\nu_t([-\boldsymbol{\infty}, \mathbf{x})^c)).$$

Hence

$$\mathbb{P}(\mathbf{X}_t^\Delta < \mathbf{x}) = \begin{cases} \exp(-\nu_t([-\boldsymbol{\infty}, \mathbf{x})^c)) & \text{for } \mathbf{x} \in \mathbb{R}_+^d \backslash \{\mathbf{0}\}, \\ 0 & \text{otherwise.} \end{cases}$$

By Remark 3.3 (i), $\mu_t(\partial[-\boldsymbol{\infty}, \mathbf{x})^c) = 0$ for every $\mathbf{x} \in \mathbb{R}_+^d \backslash \{\mathbf{0}\}$ (boundaries of spheres centered at $\mathbf{0}$ have zero μ_t-measure). Hence, by Theorem 3.16, for every $\mathbf{x} \in \mathbb{R}_+^d \backslash \{\mathbf{0}\}$

$$\begin{aligned} n\,\mathbb{P}(a_n^{-1}\mathbf{X}_t^\Delta \in [-\boldsymbol{\infty}, \mathbf{x})^c) &= n(1 - \exp(-\nu_t(a_n[-\boldsymbol{\infty}, \mathbf{x})^c))) \\ &= n\nu_t(a_n[-\boldsymbol{\infty}, \mathbf{x})^c)(1 + O(\nu_t(a_n[-\boldsymbol{\infty}, \mathbf{x})^c))) \\ &\to \mu_t([-\boldsymbol{\infty}, \mathbf{x})^c). \end{aligned}$$

Take $\mathbf{a}, \mathbf{b} \in \mathbb{R}_+^d \backslash \{\mathbf{0}\}$ with $\mathbf{a} < \mathbf{b}$. Then

$$\begin{aligned} &n\,\mathbb{P}(a_n^{-1}\mathbf{X}_t^\Delta \in [\mathbf{a}, \mathbf{b})) \\ &= \sum_{i_1=1}^{2} \cdots \sum_{i_d=1}^{2} (-1)^{i_1+\cdots+i_d+1} n\,\mathbb{P}(a_n^{-1}\mathbf{X}_t^\Delta \in \{[-\infty, x_{1i_1}) \times \cdots \times [-\infty, x_{di_d})\}^c), \end{aligned}$$

where $x_{j1} = a^{(j)}$ and $x_{j2} = b^{(j)}$ for every $j \in \{1, \ldots, d\}$. Hence

$$n\,\mathbb{P}(a_n^{-1}\mathbf{X}_t^\Delta \in [\mathbf{a}, \mathbf{b})) \to \mu_t([\mathbf{a}, \mathbf{b})).$$

Since convergence of every such set $[\mathbf{a}, \mathbf{b})$ implies vague convergence on $\mathcal{B}(\overline{\mathbb{R}}_+^d \backslash \{\mathbf{0}\})$ the conclusion follows.
(i) $\Rightarrow$ (iv) Take $\mathbf{x} \in \mathbb{R}_+^d \backslash \{\mathbf{0}\}$. Note that if we would only take $\mathbf{x} \in (0, \infty)^d$ then we would not end up with a convergence determining class. Since $\mathbf{X}_t^* \geq \mathbf{X}_t$ a.s. and $\mu_t(\partial[\mathbf{x}, \boldsymbol{\infty})) = 0$ (boundaries of spheres centered at $\mathbf{0}$ have zero μ_t-measure)

$$n\,\mathbb{P}(a_n^{-1}\mathbf{X}_t^* \in [\mathbf{x}, \boldsymbol{\infty})) \geq n\,\mathbb{P}(a_n^{-1}\mathbf{X}_t \in [\mathbf{x}, \boldsymbol{\infty})) \to \mu_t([\mathbf{x}, \boldsymbol{\infty})),$$

i.e. $\liminf_{n\to\infty} n\,\mathbb{P}(a_n^{-1}\mathbf{X}_t^* \in [\mathbf{x}, \boldsymbol{\infty})) \geq \mu_t([\mathbf{x}, \boldsymbol{\infty}))$. To complete the proof it remains to show that $\limsup_{n\to\infty} n\,\mathbb{P}(a_n^{-1}\mathbf{X}_t^* \in [\mathbf{x}, \boldsymbol{\infty})) \leq \mu_t([\mathbf{x}, \boldsymbol{\infty}))$. First, define

$$\begin{aligned} A_{(1-\epsilon)\mathbf{x}} &= \{\mathbf{z} \in \mathbb{R}^d : z^{(j)} \geq (1-\epsilon)x^{(j)}, j = 1, \ldots, d\} \\ C_{\mathbf{x},\epsilon}^{(k)} &= \{\mathbf{z} \in \mathbb{R}^d : \mathbf{z} \in [-\boldsymbol{\infty}, \mathbf{x})^c, z^{(k)} \in [-\infty, (1-\epsilon)x^{(k)})\} \\ D_{\mathbf{x},\epsilon}^{(k)} &= \{\mathbf{z} \in \mathbb{R}^d : \mathbf{z} \in [-\boldsymbol{\infty}, \mathbf{x})^c \cap A_{(1-\epsilon)\mathbf{x}}^c, z^{(k)} \in [x^{(k)}, \infty]\} \end{aligned}$$

and note that for each $k \in \{1, \ldots, d\}$, $[-\boldsymbol{\infty}, \mathbf{x})^c \subset A_{(1-\epsilon)\mathbf{x}} \cup C_{\mathbf{x},\epsilon}^{(k)} \cup D_{\mathbf{x},\epsilon}^{(k)}$ and that the sets on the right hand side are disjoint. If $\mathbf{X}_t^* \in A_{\mathbf{x}}$, then either $\mathbf{X}_s \in A_{(1-\epsilon)\mathbf{x}}$ for

some $s \le t$ or, for some $k \in \{1, \dots, d\}$, $\mathbf{X}_{s_1} \in C^{(k)}_{\mathbf{x},\epsilon}$ for some $s_1 \le t$ and $\mathbf{X}_{s_2} \in D^{(k)}_{\mathbf{x},\epsilon}$ for some $s_2 \le t, s_2 \neq s_1$. Assume without loss of generality that

$$\begin{aligned}&\mathbb{P}(\mathbf{X}_{s_1} \in C^{(1)}_{a_n\mathbf{x},\epsilon}, \mathbf{X}_{s_2} \in D^{(1)}_{a_n\mathbf{x},\epsilon} \text{ some } s_1, s_2 \le t)\\ &\quad \ge \mathbb{P}(\mathbf{X}_{s_1} \in C^{(k)}_{a_n\mathbf{x},\epsilon}, \mathbf{X}_{s_2} \in D^{(k)}_{a_n\mathbf{x},\epsilon} \text{ some } s_1, s_2 \le t)\end{aligned}$$

for $k = 2, \dots, d$. Then

$$\begin{aligned}&n\,\mathbb{P}(\mathbf{X}^*_t \in a_n[\mathbf{x}, \boldsymbol{\infty}))\\ &\quad \le n\,\mathbb{P}(\mathbf{X}_s \in A_{a_n(1-\epsilon)\mathbf{x}} \text{ some } s \le t)\\ &\qquad + n\,\mathbb{P}(\cup_{k=1}^d \{\mathbf{X}_{s_1} \in C^{(k)}_{a_n\mathbf{x},\epsilon}, \mathbf{X}_{s_2} \in D^{(k)}_{a_n\mathbf{x},\epsilon} \text{ some } s_1, s_2 \le t\})\\ &\quad \le n\,\mathbb{P}(\mathbf{X}_s \in A_{a_n(1-\epsilon)\mathbf{x}} \text{ some } s \le t)\\ &\qquad + nd\,\mathbb{P}(\mathbf{X}_{s_1} \in C^{(1)}_{a_n\mathbf{x},\epsilon}, \mathbf{X}_{s_2} \in D^{(1)}_{a_n\mathbf{x},\epsilon} \text{ some } s_1, s_2 \le t)\end{aligned}$$

By Lemma 3.18

$$\begin{aligned}n\,\mathbb{P}(\mathbf{X}_s \in A_{a_n(1-\epsilon)\mathbf{x}} \text{ some } s \le t) &\sim n\,\mathbb{P}(\mathbf{X}_t \in A_{a_n(1-\epsilon)\mathbf{x}})\\ &\to \mu_t((1-\epsilon)[\mathbf{x}, \boldsymbol{\infty})) = (1-\epsilon)^{-\alpha}\mu_t([\mathbf{x}, \boldsymbol{\infty})),\end{aligned}$$

as $n \to \infty$. Let $\gamma = \inf_{u \in C^{(1)}_{\mathbf{x},\epsilon}, v \in D^{(1)}_{\mathbf{x},\epsilon}} |\mathbf{u} - \mathbf{v}|$. Then $\gamma > 0$ and

$$\begin{aligned}&nd\,\mathbb{P}(\mathbf{X}_{s_1} \in C^{(1)}_{a_n\mathbf{x},\epsilon}, \mathbf{X}_{s_2} \in D^{(1)}_{a_n\mathbf{x},\epsilon} \text{ some } s_1, s_2 \le t)\\ &\quad \le nd\,\mathbb{P}(\mathbf{X}_{s_1} \in C^{(1)}_{a_n\mathbf{x},\epsilon}, \mathbf{X}_{s_2} \in D^{(1)}_{a_n\mathbf{x},\epsilon} \text{ some } s_1 < s_2 \le t)\\ &\qquad + nd\,\mathbb{P}(\mathbf{X}_{s_1} \in C^{(1)}_{a_n\mathbf{x},\epsilon}, \mathbf{X}_{s_2} \in D^{(1)}_{a_n\mathbf{x},\epsilon} \text{ some } s_2 < s_1 \le t)\\ &\quad \le nd\,\mathbb{P}(\mathbf{X}_{s_1} \in C^{(1)}_{a_n\mathbf{x},\epsilon}, \mathbf{X}_{s_2} - \mathbf{X}_{s_1} \in B^c_{\mathbf{0},a_n\gamma/2} \text{ some } s_1 < s_2 \le t)\\ &\qquad + nd\,\mathbb{P}(\mathbf{X}_{s_1} \in D^{(1)}_{a_n\mathbf{x},\epsilon}, \mathbf{X}_{s_2} - \mathbf{X}_{s_1} \in B^c_{\mathbf{0},a_n\gamma/2} \text{ some } s_1 < s_2 \le t)\\ &\quad \le nd\,\mathbb{P}(\mathbf{X}_s \in C^{(1)}_{a_n\mathbf{x},\epsilon} \text{ some } s \le t)\,\mathbb{P}(\mathbf{X}_s \in B^c_{\mathbf{0},a_n\gamma/4} \text{ some } s \le t)\\ &\qquad + nd\,\mathbb{P}(\mathbf{X}_s \in D^{(1)}_{a_n\mathbf{x},\epsilon} \text{ some } s \le t)\,\mathbb{P}(\mathbf{X}_s \in B^c_{\mathbf{0},a_n\gamma/4} \text{ some } s \le t)\\ &\quad \to 0,\end{aligned}$$

as $n \to \infty$. Since $\epsilon > 0$ was arbitrary it follows that $\limsup_{n\to\infty} n\,\mathbb{P}(a_n^{-1}\mathbf{X}^*_t \in [\mathbf{x}, \boldsymbol{\infty})) \le \mu_t([\mathbf{x}, \boldsymbol{\infty}))$. Hence

$$n\,\mathbb{P}(a_n^{-1}\mathbf{X}^*_t \in [\mathbf{x}, \boldsymbol{\infty})) \to \mu_t([\mathbf{x}, \boldsymbol{\infty})).$$

Take $\mathbf{a}, \mathbf{b} \in \mathbb{R}^d_+ \setminus \{\mathbf{0}\}$ with $\mathbf{a} < \mathbf{b}$. Then

$$\begin{aligned}&n\,\mathbb{P}(a_n^{-1}\mathbf{X}^*_t \in [\mathbf{a}, \mathbf{b}))\\ &= \sum_{i_1=1}^{2} \cdots \sum_{i_d=1}^{2} (-1)^{i_1+\cdots+i_d} n\,\mathbb{P}(a_n^{-1}\mathbf{X}^*_t \in [x_{1i_1}, \infty) \times \cdots \times [x_{di_d}, \infty)),\end{aligned}$$

where $x_{j1} = a^{(j)}$ and $x_{j2} = b^{(j)}$ for every $j \in \{1, \dots, d\}$. Hence

$$n\,\mathbb{P}(a_n^{-1}\mathbf{X}_t^* \in [\mathbf{a}, \mathbf{b})) \to \mu_t([\mathbf{a}, \mathbf{b})).$$

Since convergence of every such set $[\mathbf{a}, \mathbf{b})$ implies vague convergence on $\mathcal{B}(\overline{\mathbb{R}}_+^d \backslash \{\mathbf{0}\})$ the conclusion follows.

□

Assertion (ii) of the last theorem gives the asymptotics for the probability that the process reaches a certain set during the time interval $[0, t]$. What if we would allow the set of interest to vary over time? i.e. we are looking for the asymptotics of the probability that the graph of the process $\{(s, \mathbf{X}_s) : 0 \le s \le t\}$ intersects a relatively compact set in $\mathcal{B}([0, t] \times (\overline{\mathbb{R}}^d \backslash \{\mathbf{0}\}))$, the σ-algebra generated by the sets of the form $T \times B$ with $T \in \mathcal{B}([0, t])$ and $B \in \mathcal{B}(\overline{\mathbb{R}}^d \backslash \{\mathbf{0}\})$. This requires knowledge of the tail of $\mathbf{X}_s$ for all $s \in [0, t]$, not only of $\mathbf{X}_t$ as has been the case so far. To obtain this kind of results we will work with the measure $\widetilde{\nu}$ (see Theorem 3.7) instead of simply ν_t. In order to fit into the vague convergence framework we extend the measure $\widetilde{\nu}$ to $[0, \infty) \times (\overline{\mathbb{R}}^d \backslash \{\mathbf{0}\})$ by requiring that $\widetilde{\nu}(\{(s, \mathbf{x}) : s = 0 \text{ or } \mathbf{x} \in \overline{\mathbb{R}}^d \backslash \mathbb{R}^d\}) = 0$. This extension is unique. Let us introduce the operation $*$ such that for $a \in (0, \infty)$ and sets of the form $A \times B$, $A \in \mathcal{B}([0, t])$ and $B \in \mathcal{B}(\overline{\mathbb{R}}^d \backslash \{\mathbf{0}\})$, we have $a * A \times B = A \times aB$. Clearly this operation can be extended to all sets in $\mathcal{B}([0, t] \times \overline{\mathbb{R}}^d \backslash \{\mathbf{0}\})$.

Theorem 3.20. *Let $\{\mathbf{X}_t : t \ge 0\}$ be an additive process on $\mathbb{R}^d$ with system of generating triplets $\{(A_t, \nu_t, \boldsymbol{\gamma}_t)\}$. Fix an arbitrary $t > 0$. Then the following statements are equivalent.*

(i) There exist a sequence $\{a_n\}$, $0 < a_n \uparrow \infty$, and a collection of Radon measures $\{\mu_s : 0 \le s \le t\}$ on $\mathcal{B}(\overline{\mathbb{R}}^d \backslash \{\mathbf{0}\})$ such that for every $0 \le s \le t$, $\mu_s(\overline{\mathbb{R}}^d \backslash \mathbb{R}^d) = 0$, $\mu_0 \equiv 0$, μ_t is nonzero, and

$$n\,\mathbb{P}(a_n^{-1}\mathbf{X}_s \in \cdot\,) \xrightarrow{v} \mu_s(\cdot) \quad \text{on } \mathcal{B}(\overline{\mathbb{R}}^d \backslash \{\mathbf{0}\}). \tag{3.16}$$

(ii) There exist a sequence $\{a_n\}$, $0 < a_n \uparrow \infty$, and a collection of Radon measures $\{\mu_s : 0 < s \le t\}$ on $\mathcal{B}(\overline{\mathbb{R}}^d \backslash \{\mathbf{0}\})$ such that for every $0 \le s \le t$, $\mu_s(\overline{\mathbb{R}}^d \backslash \mathbb{R}^d) = 0$, $\mu_0 \equiv 0$, μ_t is nonzero, and

$$n\nu_s(a_n \cdot\,) \xrightarrow{v} \mu_s(\cdot) \quad \text{on } \mathcal{B}(\overline{\mathbb{R}}^d \backslash \{\mathbf{0}\}). \tag{3.17}$$

(iii) There exist a sequence $\{a_n\}$, $0 < a_n \uparrow \infty$, and a nonzero Radon measure $\widetilde{\mu}$ on $\mathcal{B}([0, t] \times (\overline{\mathbb{R}}^d \backslash \{\mathbf{0}\}))$ with $\widetilde{\mu}([0, t] \times \overline{\mathbb{R}}^d \backslash \mathbb{R}^d) = 0$ such that, as $n \to \infty$,

$$n\widetilde{\nu}(a_n * \cdot\,) \xrightarrow{v} \widetilde{\mu}(\cdot) \quad \text{on } \mathcal{B}([0, t] \times (\overline{\mathbb{R}}^d \backslash \{\mathbf{0}\})). \tag{3.18}$$

(iv) There exist a sequence $\{a_n\}$, $0 < a_n \uparrow \infty$, *and a nonzero Radon measure* $\widetilde{\mu}$ *on* $\mathcal{B}([0,t] \times (\overline{\mathbb{R}}^d \backslash \{\mathbf{0}\}))$ *with* $\widetilde{\mu}([0,t] \times \overline{\mathbb{R}}^d \backslash \mathbb{R}^d) = 0$ *such that, as* $n \to \infty$,

$$n\,\mathbb{P}(\{(s, \mathbf{X}_s) : 0 \le s \le t\} \cap (a_n * \cdot\,) \neq \emptyset) \xrightarrow{v} \widetilde{\mu}(\cdot) \tag{3.19}$$

on $\mathcal{B}([0,t] \times (\overline{\mathbb{R}}^d \backslash \{\mathbf{0}\}))$.

Furthermore, the sequences $\{a_n\}$ *in (i)-(iv) can be taken to be equal and then the measures* μ_s *and* $\widetilde{\mu}([0,s] \times \cdot\,)$ *in (i)-(iv) coincide on* $\mathcal{B}(\overline{\mathbb{R}}^d \backslash \{\mathbf{0}\})$ *for every* $0 \le s \le t$.

Lemma 3.21. *Let* $\{\mathbf{X}_t : t \ge 0\}$ *be an additive process on* $\mathbb{R}^d$. *Fix an arbitrary* $t > 0$. *Then, for any sequence* $\{a_n\}$, $0 < a_n \uparrow \infty$, *and* $t > 0$,

$$n\,\mathbb{P}(a_n^{-1}\mathbf{X}_t \in \cdot\,) \xrightarrow{v} 0 \quad \text{on } \mathcal{B}(\overline{\mathbb{R}}^d \backslash \{\mathbf{0}\}) \tag{3.20}$$

if and only if

$$n\,\mathbb{P}(a_n^{-1}\mathbf{X}_s \in \cdot \text{ some } s \le t) \xrightarrow{v} 0 \quad \text{on } \mathcal{B}(\overline{\mathbb{R}}^d \backslash \{\mathbf{0}\}). \tag{3.21}$$

Proof. Fix a sequence $\{a_n\}$, $t > 0$ and relatively compact $B \in \mathcal{B}(\overline{\mathbb{R}}^d \backslash \{\mathbf{0}\})$.
Suppose that (3.21) holds. Since

$$\mathbb{P}(\mathbf{X}_t \in a_n B) \le \mathbb{P}(\mathbf{X}_s \in a_n B \text{ some } s \le t)$$

it follows immediately that (3.20) holds.
Suppose that (3.20) holds and, without loss of generality, that $B \subset \mathbb{R}^d$. Fix $r > 0$ and let $\gamma = \inf_{\mathbf{x} \in B} |\mathbf{x}|$. By Lemma 3.18 (i),

$$\begin{aligned} \mathbb{P}(\mathbf{X}_s \in a_n B \text{ some } s \le t) &\le \mathbb{P}(\mathbf{X}_s \in a_n B^c_{\mathbf{0},\gamma} \text{ some } s \le t) \\ &\le \frac{\mathbb{P}(\mathbf{X}_t \in S_r(a_n B^c_{\mathbf{0},\gamma}))}{\mathbb{P}(\mathbf{X}_s \in B_{\mathbf{0},r/2} \text{ all } s \le t)}. \end{aligned}$$

For every $r > 0$ there is an $n_{r,\gamma}$ such that $S_r(a_n B^c_{\mathbf{0},\gamma}) \subset a_n B^c_{\mathbf{0},\gamma/2}$ for $n > n_{r,\gamma}$. Hence

$$\limsup_{n\to\infty} n\,\mathbb{P}(\mathbf{X}_s \in a_n B \text{ some } s \le t) \le \limsup_{n\to\infty} \frac{n\,\mathbb{P}(\mathbf{X}_t \in a_n B^c_{\mathbf{0},\gamma/2})}{\mathbb{P}(\mathbf{X}_s \in B_{\mathbf{0},r/2} \text{ all } s \le t)} = 0,$$

from which (3.20) follows. □

Proof of Theorem 3.20. (i) $\Leftrightarrow$ (ii) For every $s \in [0,t]$ for which μ_s is nonzero this equivalence was established in Theorem 3.16. We need to establish the equivalence also if $\mu_s \equiv 0$. Since μ_t is nonzero, by Remark 3.3 (ii), the sequence $\{a_n\}$ has to be regularly varying. Hence equivalence can be proved by exactly the same arguments as in the proof of Theorem 3.16.

(ii) $\Rightarrow$ (iii) Define the Radon measure $\widetilde{\mu}$ on $\mathcal{B}([0,t]\times(\overline{\mathbb{R}}^d\backslash\{\mathbf{0}\}))$ by $\widetilde{\mu}([0,s]\times B)=\mu_s(B)$ for every $s\in[0,t]$ and $B\in\mathcal{B}(\overline{\mathbb{R}}^d\backslash\{\mathbf{0}\})$. Let us show that the sequence $\{n\widetilde{\nu}(a_n\cdot)\}$ is relatively compact in the vague topology. For each bounded set $A\in\mathcal{B}([0,t]\times\overline{\mathbb{R}}^d\backslash\{\mathbf{0}\})$ there exists a bounded set $B\in\mathcal{B}(\overline{\mathbb{R}}^d\backslash\{\mathbf{0}\})$ such that $A\subset[0,t]\times B$. Hence,

$$\sup_n n\widetilde{\nu}(a_n*A)\le\sup_n n\nu_t(a_nB)<\infty$$

and it follows that $\{n\widetilde{\nu}(a_n\cdot)\}$ is relatively compact. Since the sets of the form $[0,s]\times B$ with $\mu_t(\partial B)=0$ form a π-system which generates $\mathcal{B}([0,t]\times(\overline{\mathbb{R}}^d\backslash\{\mathbf{0}\}))$ it is sufficient to show that if $\widetilde{\mu}'$ is the limit of the subsequence $\{n'\widetilde{\nu}(a_{n'}\cdot)\}$, then $\widetilde{\mu}'$ and $\widetilde{\mu}$ coincide on sets of this form. Fix an arbitrary $s\in(0,t]$ and relatively compact $B\in\mathcal{B}(\overline{\mathbb{R}}^d\backslash\{\mathbf{0}\})$ with $\mu_t(\partial B)=0$. Then, $\mu_s(\partial B)\le\mu_t(\partial B)=0$ and

$$n'\widetilde{\nu}(a_{n'}*[0,s]\times B)=n'\nu_s(a_{n'}B)\to\mu_s(B)=\widetilde{\mu}([0,s]\times B)$$

and the conclusion follows.
(iii) $\Rightarrow$ (ii) Put $\mu_s(\cdot)=\widetilde{\mu}([0,s]\times\,\cdot\,)$ for $s\in[0,t]$. Let $F\in\mathcal{B}(\overline{\mathbb{R}}^d\backslash\{\mathbf{0}\})$ be closed and bounded. Then $[0,s]\times F$ is closed and bounded and by the Portmanteau theorem (see Kallenberg [9] Theorem 15.7.2 p. 169)

$$\limsup_{n\to\infty}n\nu_s(a_nF)=\limsup_{n\to\infty}n\widetilde{\nu}(a_n*[0,s]\times F)\le\widetilde{\mu}([0,s]\times F)=\mu_s(F)$$

and hence $n\nu_s(a_n\cdot)\xrightarrow{v}\mu_s(\cdot)$.
(i) $\Rightarrow$ (iv) Define the Radon measure $\widetilde{\mu}$ on $\mathcal{B}([0,t]\times(\overline{\mathbb{R}}^d\backslash\{\mathbf{0}\}))$ by $\widetilde{\mu}([0,s]\times B)=\mu_s(B)$ for every $s\in[0,t]$ and $B\in\mathcal{B}(\overline{\mathbb{R}}^d\backslash\{\mathbf{0}\})$. Let us show that the sequence $\{n\,\mathbb{P}(\{(s,\mathbf{X}_s):0\le s\le t\}\cap(a_n*\,\cdot\,)\ne\emptyset)\}$ is relatively compact in the vague topology. For each bounded set $A\in\mathcal{B}([0,t]\times\overline{\mathbb{R}}^d\backslash\{\mathbf{0}\})$ there exists a bounded set $B\in\mathcal{B}(\overline{\mathbb{R}}^d\backslash\{\mathbf{0}\})$ such that $A\subset[0,t]\times B$. Hence, by Theorem 3.19

$$\begin{aligned}&\sup_n n\,\mathbb{P}(\{(u,\mathbf{X}_u):0\le u\le s\}\cap(a_n*A)\ne\emptyset)\\&\quad\le\sup_n n\,\mathbb{P}(a_n^{-1}\mathbf{X}_u\in B\text{ some }u\le t)<\infty\end{aligned}$$

and it follows that $\{n\,\mathbb{P}(\{(s,\mathbf{X}_s):0\le s\le t\}\cap(a_n*\,\cdot\,)\ne\emptyset)\}$ is relatively compact. Since the sets of the form $[0,s]\times B$ with $\mu_t(\partial B)=0$ form a π-system which generates $\mathcal{B}([0,t]\times(\overline{\mathbb{R}}^d\backslash\{\mathbf{0}\}))$ it is sufficient to show that if $\widetilde{\mu}'$ is the limit of the subsequence $\{n'\widetilde{\nu}(a_{n'}\cdot)\}$, then $\widetilde{\mu}'$ and $\widetilde{\mu}$ coincide on sets of this form. Fix an arbitrary $s\in(0,t]$ and relatively compact $B\in\mathcal{B}(\overline{\mathbb{R}}^d\backslash\{\mathbf{0}\})$ with $\mu_t(\partial B)=0$. Then, $\mu_s(\partial B)\le\mu_t(\partial B)=0$ and by Theorem 3.19 and Lemma 3.21

$$\begin{aligned}&n'\,\mathbb{P}(\{(u,\mathbf{X}_u):0\le u\le t\}\cap(a_{n'}*[0,s]\times B)\ne\emptyset)=n'\,\mathbb{P}(a_{n'}^{-1}\mathbf{X}_u\in B\text{ some }u\le s)\\&\to\mu_s(B)=\widetilde{\mu}([0,s]\times B)\end{aligned}$$

and the conclusion follows.
(iv) $\Rightarrow$ (i) Put $\mu_s(\cdot) = \widetilde{\mu}([0,s] \times \ \cdot\ )$ for $s \in [0,t]$. Let $F \in \mathcal{B}(\overline{\mathbb{R}}^d \backslash \{\mathbf{0}\})$ be closed and bounded. Then $[0,s] \times F$ is closed and bounded and by the Portmanteau theorem (see Kallenberg [9] Theorem 15.7.2 p. 169), Theorem 3.19 and Lemma 3.21

$$\begin{aligned}
\limsup_{n\to\infty} n' \, \mathbb{P}(a_{n'}^{-1}\mathbf{X}_s \in F) &= \limsup_{n\to\infty} n' \, \mathbb{P}(a_{n'}^{-1}\mathbf{X}_u \in F \text{ some } u \le s) \\
&= \limsup_{n\to\infty} n \, \mathbb{P}(\{(s, \mathbf{X}_s) : 0 \le s \le t\} \cap (a_n * [0,s] \times F\ ) \neq \emptyset) \le \widetilde{\mu}([0,s] \times F) \\
&= \mu_s(F)
\end{aligned}$$

and hence $n \, \mathbb{P}(a_n^{-1}\mathbf{X}_s \in \cdot\,) \xrightarrow{v} \mu_s(\cdot)$. □

Working on the product space $[0,t] \times (\overline{\mathbb{R}}^d \backslash \{\mathbf{0}\})$ allows us to consider interesting functionals such as the integral of each component of the process. To derive the tail behavior for this functional we use the intuition described in the introduction which says that if the process takes a big jump, then it varies very little before and after the jump, compared to the size of the jump. Then the intuition tells us that the integral of component j, say, becomes bigger than u if $X_s^{(j)}$ exceeds $u/(t-s)$ for some $s \in [0,t]$. That is, the integral of $X^{(j)}$ becomes bigger than u if the graph $\{(s, X_s^{(j)}) : 0 \le s \le t\}$ intersects with the set $A_t^{(j)} = \{(s,x) \in [0,t] \times (\mathbb{R} \backslash \{0\}) : x > u/(t-s)\}$. This intuition is made precise in the next result, which also finishes the paper.

Theorem 3.22. *Let $\{\mathbf{X}_t : t \ge 0\}$ be an additive process on $\mathbb{R}^d$ with system of generating triplets $\{(A_t, \nu_t, \boldsymbol{\gamma}_t)\}$. Fix an arbitrary $t > 0$ and suppose there exist a sequence $\{a_n\}$, $0 < a_n \uparrow \infty$, and a nonzero Radon measure $\widetilde{\mu}$ on $\mathcal{B}([0,t] \times (\overline{\mathbb{R}}^d \backslash \{\mathbf{0}\}))$ with $\widetilde{\mu}([0,t] \times \overline{\mathbb{R}}^d \backslash \mathbb{R}^d) = 0$ such that*

$$n\widetilde{\nu}(a_n * \cdot\ ) \xrightarrow{v} \widetilde{\mu}(\cdot) \quad \text{on } \mathcal{B}([0,t] \times (\overline{\mathbb{R}}^d \backslash \{\mathbf{0}\})).$$

If $\mathbf{I}_t = \big(\int_0^t X_s^{(1)} \mathrm{d}s, \ldots, \int_0^t X_s^{(d)} \mathrm{d}s \big)$, then

$$n \, \mathbb{P}(a_n^{-1}\mathbf{I}_t \in \cdot\ ) \xrightarrow{v} \widetilde{\mu}(\{(s,\mathbf{x}) \in [0,t] \times (\overline{\mathbb{R}}^d \backslash \{\mathbf{0}\}) : \mathbf{x} \in \frac{1}{t-s} \cdot \}) \quad \text{on } \mathcal{B}(\overline{\mathbb{R}}^d \backslash \{\mathbf{0}\}).$$

In particular, if $\{\mathbf{X}_s : s \ge 0\}$ is a Lévy process, then

$$n \, \mathbb{P}(a_n^{-1}\mathbf{I}_t \in \cdot\ ) \xrightarrow{v} \frac{t^{\alpha+1}}{\alpha+1} \mu(\cdot) \quad \text{on } \mathcal{B}(\overline{\mathbb{R}}^d \backslash \{\mathbf{0}\}),$$

where $\mu(\cdot) = \widetilde{\mu}([0,1] \times \cdot\ )$ and $\alpha > 0$ is such that $\mu(u\ \cdot\ ) = u^{-\alpha}\mu(\cdot)$ for all $u > 0$.

Proof. By the Lévy-Itô decomposition we may write the process $\{\mathbf{X}_s\}$ as the sum of three independent processes. With the notation of Theorem 3.7,

$$\begin{aligned}\mathbf{X}_s(\omega) &= \mathbf{X}_s^2(\omega) \\ &+ \underbrace{\lim_{\epsilon\downarrow 0}\int_{(0,s]\times D_{\epsilon,1}} \{\mathbf{x}\xi(\mathrm{d}(u,\mathbf{x}),\omega) - \mathbf{x}\widetilde{\nu}(\mathrm{d}(u,\mathbf{x}))\}}_{\mathbf{Y}_s} + \underbrace{\int_{(0,s]\times D_{1,\infty}} \mathbf{x}\xi(\mathrm{d}(u,\mathbf{x}),\omega)}_{\mathbf{J}_s},\end{aligned}$$

on Ω_1, where $\{\mathbf{X}_s^2\}$ has a version with continuous sample paths, $\{\mathbf{Y}_s\}$ is a jump process with small jumps, and $\{\mathbf{J}_s\}$ is a jump process with big jumps. For $j = 1,\dots,m$, let $\mathbf{V}_j^m = \mathbf{J}_{tj/m} - \mathbf{J}_{t(j-1)/m}$. For $\omega \in \Omega_0 \cap \Omega_1$,

$$\sum_{j=1}^m \frac{t}{m}\mathbf{J}_{tj/m}(\omega) \to \int_0^t \mathbf{J}_s(\omega)\mathrm{d}s \quad \text{as } m\to\infty,$$

and

$$\sum_{j=1}^m |\mathbf{V}_j^m(\omega)| \to \sum_{u\in(0,t)} |\Delta\mathbf{J}_u(\omega)| \quad \text{as } m\to\infty,$$

where the sum extends over all (finitely many) u such that $|\Delta\mathbf{J}_u(\omega)|$ is nonzero. Take $\epsilon \in (0,1/2)$ and let A be a relatively compact μ_t-continuity set of the form $\{\mathbf{x}\in\mathbb{R}^d\backslash\{\mathbf{0}\} : |\mathbf{x}| > u, \mathbf{x}/|\mathbf{x}| \in S\}$ for $u>0$ and $S\in\mathcal{B}(\mathbb{S}^{d-1})$. Convergence of every such set implies vague convergence on $\mathcal{B}(\overline{\mathbb{R}}^d\backslash\{\mathbf{0}\})$. To begin with, we study the behavior of $n\,\mathbb{P}(\int_0^t \mathbf{J}_s\mathrm{d}s \in a_nA)$ as $n\to\infty$. Let $\gamma = \inf_{\mathbf{x}\in A}|\mathbf{x}|$ and set $A_t = \{(s,\mathbf{x}) : \mathbf{x}\in\frac{1}{t-s}A, s\in[0,t]\} \in \mathcal{B}([0,t]\times(\overline{\mathbb{R}}^d\backslash\{\mathbf{0}\}))$. We begin by constructing a lower bound.

$$\begin{aligned}&n\,\mathbb{P}(\sum_{j=1}^m \frac{t}{m}\mathbf{J}_{tj/m}\in a_nA) \\ &\geq n\,\mathbb{P}(\cup_{j=1}^m\{\mathbf{V}_j^m \in \frac{1}{t-tj/m}a_n(1+\epsilon)A, \sum_{i\neq j}|\mathbf{V}_i^m| \leq a_n\epsilon\gamma/t\}) \\ &= \sum_{j=1}^m \mathbb{P}(\sum_{i\neq j}|\mathbf{V}_i^m| \leq a_n\epsilon\gamma/t)\; n\,\mathbb{P}(\mathbf{V}_j^m \in \frac{1}{t-tj/m}a_n(1+\epsilon)A) \\ &\geq \mathbb{P}(\sum_{j=1}^m|\mathbf{V}_j^m| \leq a_n\epsilon\gamma/t)\sum_{j=1}^m n\,\mathbb{P}(\mathbf{V}_j^m \in \frac{1}{t-tj/m}a_n(1+\epsilon)A) \\ &\geq \mathbb{P}(\sum_{j=1}^m|\mathbf{V}_j^m| \leq a_n\epsilon\gamma/t)\; n\,\mathbb{P}(\cup_{j=1}^m\{\mathbf{V}_j^m \in \frac{1}{t-tj/m}a_n(1+\epsilon)A\}).\end{aligned}$$

Letting $m\to\infty$, we arrive at

$$n\,\mathbb{P}(\int_0^t \mathbf{J}_s\mathrm{d}s \in a_nA) \geq \mathbb{P}(\sum_{u\in(0,t)}|\Delta\mathbf{J}_u| \leq a_n\epsilon\gamma/t)\; n\widetilde{\nu}((1+\epsilon)a_n * A_t).$$

We now construct an upper bound.

$$
\begin{aligned}
& n\,\mathbb{P}(\sum_{j=1}^{m}\frac{t}{m}\mathbf{J}_{tj/m}\in a_nA)\\
&\quad\le n\,\mathbb{P}(\cup_{j=1}^{m}\{\mathbf{V}_j^m\in\frac{1}{t-tj/m}a_n(1-\epsilon)A,\sum_{i\neq j}|\mathbf{V}_i^m|\le a_n\epsilon\gamma/t\})\\
&\qquad+n\,\mathbb{P}(\mathbf{J}_{ti/m}\in B^c_{\mathbf{0},a_n\epsilon\gamma/t},\mathbf{J}_{tj/m}-\mathbf{J}_{ti/m}\in B^c_{\mathbf{0},a_n\epsilon\gamma/t}\text{ some }i<j\le m)\\
&\quad\le n\,\mathbb{P}(\cup_{j=1}^{m}\{\mathbf{V}_j^m\in\frac{1}{t-tj/m}a_n(1-\epsilon)A\})\\
&\qquad+n\,\mathbb{P}(\mathbf{J}_{ti/m}\in B^c_{\mathbf{0},a_n\epsilon\gamma/t},\mathbf{J}_{tj/m}-\mathbf{J}_{ti/m}\in B^c_{\mathbf{0},a_n\epsilon\gamma/t}\text{ some }i<j\le m)\\
&\quad\le n\,\mathbb{P}(\cup_{j=1}^{m}\{\mathbf{V}_j^m\in\frac{1}{t-tj/m}a_n(1-\epsilon)A\})\\
&\qquad+n\,\mathbb{P}(\mathbf{J}_{tj/m}\in B^c_{\mathbf{0},a_n\epsilon\gamma/t}\text{ some }j\le m)^2.
\end{aligned}
$$

Letting $m\to\infty$, we arrive at

$$
n\,\mathbb{P}(\int_0^t\mathbf{J}_s\mathrm{d}s\in a_nA)\le n\widetilde{\nu}((1-\epsilon)a_n*A_t)+n\,\mathbb{P}(\mathbf{J}_s\in B^c_{\mathbf{0},a_n\epsilon\gamma/t}\text{ some }s\in(0,t])^2.
$$

Since $\lim_{n\to\infty}\mathbb{P}(\sum_{u\in(0,t)}|\Delta\mathbf{J}_u|\le a_n\epsilon\gamma/t)=1$ and by the second part of Lemma 3.18 $\lim_{n\to\infty}n\,\mathbb{P}(\mathbf{J}_s\in B^c_{\mathbf{0},a_n\epsilon\gamma/t}\text{ some }s\in(0,t])^2=0$, we get

$$
\begin{aligned}
\liminf_{n\to\infty}n\,\mathbb{P}(\int_0^t\mathbf{J}_s\mathrm{d}s\in a_nA) &\ge \widetilde{\mu}((1+\epsilon)*A_t),\\
\limsup_{n\to\infty}n\,\mathbb{P}(\int_0^t\mathbf{J}_s\mathrm{d}s\in a_nA) &\le \widetilde{\mu}((1-\epsilon)*A_t).
\end{aligned}
$$

Since $\epsilon>0$ was arbitrary and $\widetilde{\mu}(u*\cdot\,)=u^{-\alpha}\widetilde{\mu}(\cdot)$ for all $u>0$, it follows that the lower and upper bound coincide, i.e.

$$
\lim_{n\to\infty}n\,\mathbb{P}(\int_0^t\mathbf{J}_s\mathrm{d}s\in a_nA)=\widetilde{\mu}(A_t).
$$

Since as $n\to\infty$

$$
n\,\mathbb{P}(\int_0^t\mathbf{X}_s^2\mathrm{d}s\in a_nA)\le n\,\mathbb{P}(\mathbf{X}_s^2\in\frac{1}{t}a_nA\text{ some }s\le t)\to 0
$$

and

$$
n\,\mathbb{P}(\int_0^t\mathbf{Y}_s\mathrm{d}s\in a_nA)\le n\,\mathbb{P}(\mathbf{Y}_s\in\frac{1}{t}a_nA\text{ some }s\le t)\to 0,
$$

by Proposition 3.11,

$$
\lim_{n\to\infty}n\,\mathbb{P}(\int_0^t\mathbf{X}_s\mathrm{d}s\in a_nA)=\lim_{n\to\infty}n\,\mathbb{P}(\int_0^t\mathbf{J}_s\mathrm{d}s\in a_nA).
$$

Hence

$$\lim_{n\to\infty} n\,\mathbb{P}(\int_0^t \mathbf{X}_s \mathrm{d}s \in a_n A) = \widetilde{\mu}(A_t).$$

In the case of a Lévy process, i.e. if $\widetilde{\mu}([0,s]\times\cdot\,) = s\mu(\cdot)$ for $s\in[0,t]$, then

$$\widetilde{\mu}(A_t) = \int_0^t \mu(\frac{1}{t-s}A)\mathrm{d}s = \mu(A)\int_0^t (t-s)^{\alpha}\mathrm{d}s = \frac{t^{\alpha+1}}{\alpha+1}\mu(A).$$

□

References

[1] Basrak, B. (2000) *The Sample Autocorrelation Function of Non-Linear Time Series*, Ph.D. Thesis, University of Groningen, Department of Mathematics.

[2] Basrak, B., Davis, R.A. and Mikosch, T. (2002) Regular variation of GARCH processes, *Stochastic Process. Appl.*, Vol. 99, 95–115.

[3] Basrak, B., Davis, R.A. and Mikosch, T. (2002) A Characterization of Multivariate Regular Variation, *Ann. Appl. Probab.*, Vol. 12, 908–920.

[4] Bingham, N.H., Goldie, C.M. and Teugels, J.L. (1987) *Regular variation*, Encyclopedia of mathematics and its applications 27, Cambridge University Press.

[5] Braverman, M., Mikosch, T. and Samorodnitsky, G. (2002) Tail probabilities of subadditive functionals acting on Lévy processes, *Ann. Appl. Probab.*, Vol. 12, 69–100.

[6] Davis, R.A. and Hsing, T. (1995) Point process and partial sum convergence for weakly dependent random variables with infinite variance, *Ann. Probab.*, Vol. 23, 879–917.

[7] Embrechts, P., Goldie, C.M. and Veraverbeke, N. (1979) Subexponentiality and Infinite Divisibility, *Z. Wahrsch. Verw. Gebiete*, Vol. 49, 335–347.

[8] Gradshteyn, I.S. and Ryzhik, I.M. (2000) *Table of Integrals, Series, and Products*, 6th edition, Academic Press.

[9] Kallenberg, O. (1983) *Random Measures*, 3rd edition, Akademie Verlag, Berlin.

[10] Kesten, H. (1973) Random difference equations and renewal theory for products of random matrices, *Acta Math.*, Vol. 131, 207–248.

[11] Pratt, J. (1960) On interchanging limits and integrals, *Ann. Math. Statist.*, Vol. 31, 74–77.

[12] Resnick, S.I. (1986) Point processes, regular variation and weak convergence, *Adv. in Appl. Probab.*, Vol. 18, 66–138.

[13] Resnick, S.I. (1987) *Extreme Values, Regular Variation, and Point Processes*, Springer Verlag, New York.

[14] Rosiński, J. and Samorodnitsky, G. (1993) Distributions of subadditive functionals of sample paths of infinitely divisible processes, *Ann. Probab.*, Vol. 21, 996–1014.

[15] Sato, K.-I. (1999) *Lévy Processes and Infinitely Divisible Distributions*, Cambridge University Press.

[16] Willekens, E. (1987) On the supremum of an infinitely divisible process, *Stochastic Process. Appl.*, Vol. 26, 173–175.

Chapter 4

On regular variation for stochastic processes

Henrik Hult and Filip Lindskog (2003), On regular variation for stochastic processes, *submitted*.

Abstract. We study a formulation of regular variation on the space of $\mathbb{R}^d$-valued right-continuous functions on $[0,1]$ with left limits and provide necessary and sufficient conditions for a stochastic process with sample paths in this space to be regularly varying. A version of the Continuous Mapping Theorem is proved which enables the derivation of the tail behavior of rather general mappings of the regularly varying stochastic process. For a wide class of Markov processes with asymptotically independent increments we obtain simplified sufficient conditions for regular variation. For such processes we show that the possible regular variation limit measures concentrate on step functions with one step, from which we conclude that extremes for such processes are due to one big jump in $(0,1]$ or an extreme starting point. Finally, using the Continuous Mapping Theorem we derive the tail behavior of filtered regularly varying Markov processes with asymptotically independent increments.

2000 *Mathematics Subject Classification.* 60F17, 60G17 (primary); 60G07, 60G70 (secondary).
Keywords and phrases. Regular variation; Extremes; Functional limit theorem; Markov processes.
Acknowledgments. The authors want to thank Boualem Djehiche for comments on the manuscript.

4.1 Introduction

In applications one sometimes encounters data sets with a few extremely large observations. For such data sets it is suggested to use heavy-tailed probability distributions to model the underlying uncertainty. This is the case for instance in so-called catastrophe insurance (fire, wind-storm, flooding) where the occurence of large claims may lead to large fluctuations in the cash-flow process faced by the insurance company. The situation is similar in finance where extremely large losses sometimes occur, indicating heavy tails of the return distributions. The probability of extreme stock price movements has to be accounted for when analyzing the risk of a portfolio. Another application is telecommunications networks where long service times may result in large variablility in the workload process. In many applications it is appropriate to use a stochastic process $\{X_t : t \geq 0\}$ to model the evolution of the quantity of interest over time. The notion of heavy tails enters naturally in this context either as an assumption on the marginals X_t or as an assumption on the increments $X_{t+h} - X_t$ of the process. However, it is often the case that the marginals or the increments of the process are not the main concern, but rather some functional of the process. A natural example is the supremum of the process during a time interval, $\sup_{0\leq t\leq T} X_t$, i.e. the largest value reached by the process during $[0, T]$. Another example is the mean of the process during a time interval, $T^{-1}\int_0^T X_t \mathrm{d}t$. We are then typically interested in the probability that the functional exceeds some high level, e.g. –What is the probability that the sea level exceeds a high barrier sometime during $[0, T]$? It may therefore be important to know how the tail behavior of the marginals X_t (or the increments) is related to the tail behavior of functionals of the process. For univariate infinitely divisible processes results on the tail behavior for subadditive functionals are derived in Rosiński and Samorodnitsky [13] under assumptions of subexponentiality. See also Braverman, Mikosch and Samorodnitsky [5] for further results on the tail behavior of subadditive functionals of univariate regularly varying Lévy processes. In the multivariate case one typically studies a d-dimensional stochastic process $\{\mathbf{X}_t : t \geq 0\}$. The process could be interpreted for instance as the consecutive measurements of sea levels at d different locations, the daily losses of d different stocks or the weekly amount of claims in d different insurance lines. A notable difference between the multivariate case and the univariate case when analyzing extremes is the possibility to have dependence between the components of the random vector. Large values may for instance tend to occur simultaneously in the different components. To have a good understanding of the dependence between extreme events in the multivariate case may be of great importance in applications. Similar to the univariate case some functional or vector of functionals of the process may be the primary concern. Natural examples are for instance the componentwise suprema of the process, $(\sup_{0\leq t\leq T} X_t^{(1)}, \ldots, \sup_{0\leq t\leq T} X_t^{(d)})$ i.e. the largest value reached by each component of the process during $[0, T]$. Another example is the componentwise mean of the process but other functionals or combinations of functionals may

also be of interest. We are typically interested in the probability that the vector of functionals belongs to some extreme set far away from the origin, e.g. –What is the probability that the sea level exceeds a high barrier at some (or all) locations sometime during $[0,T]$? To answer this type of questions we need to know how the tail behavior of the marginals $\mathbf{X}_t$ is related to the tail behavior of vectors of functionals of the process. In this paper we provide a natural framework for addressing such questions and illustrate how it can be applied.

Multivariate regular variation provides a natural way of understanding the tail behavior of heavy-tailed random vectors. A similar construction is possible for stochastic processes with sample paths in $D([0,1],\mathbb{R}^d)$; the space of $\mathbb{R}^d$-valued right-continuous functions on $[0,1]$ with left limits. This formulation seems to be well suited for understanding the tail behavior of heavy-tailed stochastic processes. We will exemplify this in various forms throughout the paper. A random vector $\mathbf{X}$ on $\mathbb{R}^d$ is said to be regularly varying if there exist an $\alpha > 0$ and a probability measure σ on the unit sphere $\mathbb{S}^{d-1} = \{\mathbf{x} \in \mathbb{R}^d : |\mathbf{x}| = 1\}$ such that, for every $x > 0$, as $u \to \infty$,

$$\frac{\mathbb{P}(|\mathbf{X}| > ux, \mathbf{X}/|\mathbf{X}| \in \cdot\,)}{\mathbb{P}(|\mathbf{X}| > u)} \xrightarrow{w} x^{-\alpha}\sigma(\cdot) \quad \text{on } \mathcal{B}(\mathbb{S}^{d-1}),$$

where $\mathcal{B}(\mathbb{S}^{d-1})$ denotes the Borel σ-algebra on $\mathbb{S}^{d-1}$ and $\xrightarrow{w}$ denotes weak convergence. The probability measure σ is referred to as the spectral measure of $\mathbf{X}$. It describes in which directions we are likely to find extreme realizations of $\mathbf{X}$. Similarly, we say that a stochastic process $\mathbf{X} = \{\mathbf{X}_t : t \in [0,1]\}$ with sample paths in $D([0,1],\mathbb{R}^d)$ is regularly varying if there exist an $\alpha > 0$ and a probability measure σ on $D_1([0,1],\mathbb{R}^d) = \{\mathbf{x} \in D([0,1],\mathbb{R}^d) : \sup_{t\in[0,1]} |\mathbf{x}_t| = 1\}$ such that, for every $x > 0$, as $u \to \infty$,

$$\frac{\mathbb{P}(|\mathbf{X}|_\infty > ux, \mathbf{X}/|\mathbf{X}|_\infty \in \cdot\,)}{\mathbb{P}(|\mathbf{X}|_\infty > u)} \xrightarrow{w} x^{-\alpha}\sigma(\cdot) \quad \text{on } \mathcal{B}(D_1([0,1],\mathbb{R}^d)),$$

where $\mathcal{B}(D_1([0,1],\mathbb{R}^d))$ denotes the Borel σ-algebra on $D_1([0,1],\mathbb{R}^d)$ and $|\mathbf{x}|_\infty = \sup_{t\in[0,1]} |\mathbf{x}_t|$. The spectral measure σ contains essentially all relevant information for understanding the extremal behavior of the process $\mathbf{X}$. For example, it might be interesting to know under which conditions the extremes of $\mathbf{X}$ are due to (at most) one single extreme jump (we allow also an extreme starting point). This can be formalized in terms of the support of the spectral measure by showing that the spectral measure concentrates on step functions, i.e. on the set

$$\{\mathbf{x} \in D_1([0,1],\mathbb{R}^d) : \mathbf{x} = \mathbf{y}1_{[v,1]}, v \in [0,1], \mathbf{y} \in \mathbb{S}^{d-1}\}.$$

We show that this is the case for a large class of regularly varying Markov processes, including all regularly varying additive processes (and hence also all regularly varying Lévy processes).

A natural question is why one would prefer formulating regular variation on $D([0,1],\mathbb{R}^d)$ rather than on, say, $(\mathbb{R}^d)^{[0,1]}$. The main reason is that many interesting

mappings from $D([0,1],\mathbb{R}^d)$ to $D([0,1],\mathbb{R}^d)$ (or to $\mathbb{R}^k$) are continuous whereas the corresponding mappings from $(\mathbb{R}^d)^{[0,1]}$ are not. However, with constructions similar to the one used in this paper, regular variation can be formulated on other complete separable metric spaces. In this paper we prefer to work on $D([0,1],\mathbb{R}^d)$.

An equivalent definition of regular variation on $D([0,1],\mathbb{R}^d)$ is the following; a stochastic process $\mathbf{X}$ with sample paths in $D([0,1],\mathbb{R}^d)$ is regularly varying if there exist a sequence $\{a_n\}$, $0 < a_n \uparrow \infty$, and a nonzero boundedly finite measure m on $\mathcal{B}(\overline{D}([0,1],\mathbb{R}^d))$ with $m(\overline{D}([0,1],\mathbb{R}^d)\backslash D([0,1],\mathbb{R}^d)) = 0$ such that, as $n \to \infty$,

$$n\,\mathbb{P}(a_n^{-1}\mathbf{X} \in \cdot\,) \xrightarrow{\hat{w}} m(\cdot) \quad \text{on } \mathcal{B}(\overline{D}([0,1],\mathbb{R}^d)), \tag{4.1}$$

where $\xrightarrow{\hat{w}}$ denotes so-called $\hat{w}$-convergence. (The precise meaning of $\overline{D}([0,1],\mathbb{R}^d)$ is explained in Section 4.2. At this point it may be viewed as only a slight modification of $D([0,1],\mathbb{R}^d)$ needed in order to use the concept of $\hat{w}$-convergence.) Let h be a positively homogeneous (i.e. $h(\lambda\mathbf{x}) = \lambda h(\mathbf{x})$ for $\lambda \geq 0$) measurable mapping from $D([0,1],\mathbb{R}^d)$ to $D([0,1],\mathbb{R}^d)$ (or to $\mathbb{R}^k$). Then, if (4.1) holds and if h satisfies some mild conditions, as $n \to \infty$,

$$n\,\mathbb{P}(a_n^{-1}h(\mathbf{X}) \in \cdot\,) \xrightarrow{\hat{w}} m \circ h^{-1}(\,\cdot \cap D([0,1],\mathbb{R}^d)) \quad \text{on } \mathcal{B}(\overline{D}([0,1],\mathbb{R}^d)) \tag{4.2}$$

(or on $\mathcal{B}(\overline{\mathbb{R}}^k\backslash\{\mathbf{0}\})$, $\overline{\mathbb{R}} = [-\infty,\infty]$), i.e. we have a version of the Continuous Mapping Theorem. Hence, under mild conditions on h, regular variation of $\mathbf{X}$ implies regular variation of $h(\mathbf{X})$ and we can express its limit measure in terms of m and h as in (4.2).

In Section 4.2 we state the two definitions of regular variation on $D([0,1],\mathbb{R}^d)$ and show that they are equivalent. Moreover, we give necessary and sufficient conditions for regular variation for a general stochastic process with sample paths in $D([0,1],\mathbb{R}^d)$. Finally, we give a continuous mapping theorem which provides a powerful tool in the subsequent analysis. In Section 4.3 we focus on strong Markov processes with asymptotically independent increments (see Section 4.3 for the precise meaning of asymptotically independent increments). We obtain sufficient conditions for regular variation for such processes which are easier to verify since they involve only the marginals $\mathbf{X}_t$ of the process $\mathbf{X}$. Moreover, we show that the limit measure m of such regularly varying Markov processes vanishes on $\mathcal{V}^c$ where

$$\mathcal{V} = \{\mathbf{x} \in D([0,1],\mathbb{R}^d) : \mathbf{x} = \mathbf{y}1_{[v,1]}, v \in [0,1], \mathbf{y} \in \mathbb{R}^d\backslash\{0\}\}.$$

This means that, asymptotically, the process reaches a set far away from the origin either by starting there or by making exactly one big jump to this set and, in comparison to the size of the jump, it stays essentially constant before and after the jump. On one hand this means that we are able to quantify the idea of one big jump in terms of the support of the regular variation limit measure. On the other hand, and equally important, this in combination with the Continuous Mapping

Theorem (4.2) allow us to explicitly compute tail probabilities of $h(\mathbf{X})$ for many interesting choices of h. See e.g. Example 4.14 and 4.15 with

$$h(\mathbf{x}) = \Big(\sup_{t\in[0,1]} x_t^{(1)}, \dots, \sup_{t\in[0,1]} x_t^{(d)} \Big) \quad \text{and} \quad h(\mathbf{x}) = \Big(\int_0^1 x_t^{(1)} \mathrm{d}t, \dots, \int_0^1 x_t^{(d)} \mathrm{d}t \Big),$$

respectively. In Section 4.4 we study filtered stochastic processes of the form

$$\mathbf{Y}_t = \int_0^t f(t,s) \mathrm{d}\mathbf{X}_s, \quad t \in [0,1], \tag{4.3}$$

where $\mathbf{X}$ is a regularly varying Markov process of the type studied in Section 4.3, with sample paths of finite variation. Under the assumption that the kernel f is continuous we show that $\mathbf{Y}$ can be viewed as a mapping of the process $\mathbf{X}$, which is sufficiently regular to satisfy the conditions of the Continuous Mapping Theorem. We show that $\mathbf{Y}$ is regularly varying and determine the limit measure. The proofs are given at the end of each section.

4.2 Regular variation on D

Let us introduce regular variation on $D = D([0,1], \mathbb{R}^d)$; the space of functions $\mathbf{x} : [0,1] \to \mathbb{R}^d$ which are right continuous with left limits. This space is equipped with the so-called J_1-metric (referred to as d_0 in Billingsley [2]) which makes it complete and separable. The formulation of regular variation we will use has recently been used in de Haan and Lin [9] in connection with max-infinitely divisible distributions on D. See also Giné, Hahn and Vatan [8].

We denote by $D_1 = D_1([0,1], \mathbb{R}^d)$ the subspace $\{\mathbf{x} \in D : \sup_{t\in[0,1]} |\mathbf{x}_t| = 1\}$ equipped with the subspace topology. Define $\overline{D} = (0,\infty] \times D_1$, where $(0,\infty]$ is equipped with the metric $\rho(x,y) = |1/x - 1/y|$ making it complete and separable. Then $\overline{D}$ is a complete separable metric space. Note that a nonzero function $\mathbf{x} \in D$ is associated with an element $(x^*, \tilde{\mathbf{x}}) \in \overline{D}$ where $x^* = \sup_{t\in[0,1]} |\mathbf{x}_t|$ and $\tilde{\mathbf{x}} = \mathbf{x}/x^*$. For $\mathbf{x} \in D$ we write $|\mathbf{x}|_\infty = \sup_{t\in[0,1]} |\mathbf{x}_t|$ and for $\mathbf{x} = (x^*, \tilde{\mathbf{x}}) \in \overline{D}$ we write $|\mathbf{x}|_\infty = x^*$. A consequence of the above construction is that

$$\mathcal{B}(\overline{D}) \cap (D \backslash \{\mathbf{0}\}) = \mathcal{B}(D) \cap (D \backslash \{\mathbf{0}\}),$$

i.e. the Borel sets we are interested in are the usual Borel sets on D which do not contain the zero function.

We will see that regular variation on D is naturally expressed in terms of so-called $\hat{w}$-convergence of boundedly finite measures on $\overline{D}$. A boundedly finite measure assigns finite measure to bounded sets. A sequence of boundedly finite measures $\{m_n : n \in \mathbb{N}\}$ on a complete and separable metric space $\mathbb{E}$ converges to m in the $\hat{w}$-topology, $m_n \xrightarrow{\hat{w}} m$, if $m_n(B) \to m(B)$ for every bounded Borel set B with $m(\partial B) = 0$. If the state space $\mathbb{E}$ is locally compact, which $\overline{D}$ is not but $\overline{\mathbb{R}}^d \backslash \{\mathbf{0}\}$

($\overline{\mathbb{R}} = [-\infty, \infty]$) is, then a boundedly finite measure is called a Radon measure, and $\hat{w}$-convergence coincides with vague convergence and we write $m_n \xrightarrow{v} m$. Finally we note that if $m_n \xrightarrow{\hat{w}} m$ and $m_n(\mathbb{E}) \to m(\mathbb{E}) < \infty$, then $m_n \xrightarrow{w} m$. For details on $\hat{w}$-, vague- and weak convergence we refer to Appendix 2 in Daley and Vere-Jones [6]. See also Kallenberg [11] for details on vague convergence.

Definition 4.1. A stochastic process $\mathbf{X} = \{\mathbf{X}_t : t \in [0,1]\}$ with sample paths in D is said to be regularly varying if there exist a sequence $\{a_n\}$, $0 < a_n \uparrow \infty$, and a nonzero boundedly finite measure m on $\mathcal{B}(\overline{D})$ with $m(\overline{D} \backslash D) = 0$ such that, as $n \to \infty$,

$$n\,\mathbb{P}(a_n^{-1}\mathbf{X} \in \cdot\,) \xrightarrow{\hat{w}} m(\cdot) \quad \text{on } \mathcal{B}(\overline{D}). \tag{4.4}$$

Remark 4.2. The limit measure m has a scaling property; there exists an $\alpha > 0$ such that $m(uB) = u^{-\alpha} m(B)$ for every $u > 0$ and $B \in \mathcal{B}(\overline{D})$. This follows from a combination of more or less standard regular variation arguments. We include a proof in the appendix.

An equivalent and perhaps more intuitive formulation of regular variation on D is given in the next result.

Theorem 4.3. *A stochastic process $\mathbf{X} = \{\mathbf{X}_t : t \in [0,1]\}$ with sample paths in D is regularly varying if and only if there exist an $\alpha > 0$ and a probability measure σ on $\mathcal{B}(D_1)$ such that, for every $x > 0$, as $u \to \infty$,*

$$\frac{\mathbb{P}(|\mathbf{X}|_\infty > ux, \mathbf{X}/|\mathbf{X}|_\infty \in \cdot\,)}{\mathbb{P}(|\mathbf{X}|_\infty > u)} \xrightarrow{w} x^{-\alpha}\sigma(\cdot) \quad \text{on } \mathcal{B}(D_1). \tag{4.5}$$

The probability measure σ is referred to as the spectral measure of $\mathbf{X}$ and α is referred to as the tail index.

A proof of Theorem 4.3 is given in the appendix. The two formulations of regular variation on D given here are much inspired by the formulations of regular variation for random vectors. Many of them are documented in e.g. Basrak [1].

Remark 4.4. For $S \in \mathcal{B}(D_1)$, let $V_{1,S} = \{\mathbf{x} \in D : |\mathbf{x}|_\infty > 1, \mathbf{x}/|\mathbf{x}|_\infty \in S\}$. It follows from the proof of Theorem 4.3 that the probability measure σ and the boundedly finite measure m are linked through

$$\sigma(S) = \frac{m(V_{1,S})}{m(V_{1,D_1})}, \quad S \in \mathcal{B}(D_1).$$

Let $h : D \to D$ or $h : D \to \mathbb{R}^k$ be a measurable, positively homogeneous mapping, i.e. $h(\lambda\mathbf{x}) = \lambda h(\mathbf{x})$ for $\lambda \geq 0$ and $\mathbf{x} \in D$. If $\mathbf{X}$ is a regularly varying stochastic process with sample paths in D we may be interested in the tail behavior of $h(\mathbf{X})$. This is achieved using an analogue of the Continuous Mapping Theorem for weak convergence. Let $D_h = \{\mathbf{x} \in D : h \text{ is discontinuous at } \mathbf{x}\}$. Note that $D_h \in \mathcal{B}(D)$ (see Billingsley [2] p. 225) and hence also $D_h \cap \overline{D} \in \mathcal{B}(\overline{D})$.

Theorem 4.5 (the Continuous Mapping Theorem). *Let $\mathbf{X} = \{\mathbf{X}_t : t \in [0,1]\}$ be a stochastic process with sample paths in D. Suppose that there exist a sequence $\{a_n\}$, $0 < a_n \uparrow \infty$, and a nonzero boundedly finite measure m on $\mathcal{B}(\overline{D})$ with $m(\overline{D} \backslash D) = 0$ such that, as $n \to \infty$,*

$$n\,\mathbb{P}(a_n^{-1}\mathbf{X} \in \cdot\,) \xrightarrow{\hat{w}} m(\cdot) \quad \text{on } \mathcal{B}(\overline{D}).$$

Let $h : D \to D$ be a positively homogeneous measurable mapping such that $h^{-1}(B)$ is bounded in $\overline{D}$ for every bounded $B \in \mathcal{B}(\overline{D}) \cap D$ and suppose $m(D_h \cap \overline{D}) = 0$. Then,

$$n\,\mathbb{P}(a_n^{-1}h(\mathbf{X}) \in \cdot\,) \xrightarrow{\hat{w}} m \circ h^{-1}(\cdot \cap D) \quad \text{on } \mathcal{B}(\overline{D}).$$

Moreover, the result holds for mappings $h : D \to \mathbb{R}^k$ with the obvious notational changes.

The formulation of regular variation on D in combination with Theorem 4.5 allow us to derive the tail behavior of a large class of continuous mappings of stochastic processes. This will be illustrated in the following sections.

The next theorem gives necessary and sufficient conditions for a stochastic process with sample paths in D to be regularly varying. Before stating these conditions we introduce some notation. For $\mathbf{x} \in D$, $T_0 \subset [0,1]$ and $\delta \in [0,1]$ let

$$\begin{aligned} w(\mathbf{x}, T_0) &= \sup\{|\mathbf{x}_s - \mathbf{x}_t| : s,t \in T_0\}, \\ w''(\mathbf{x}, \delta) &= \sup_{t_1 \le t \le t_2, t_2 - t_1 \le \delta} \min\{|\mathbf{x}_t - \mathbf{x}_{t_1}|, |\mathbf{x}_{t_2} - \mathbf{x}_t|\}. \end{aligned}$$

Theorem 4.6. *Let $\mathbf{X} = \{\mathbf{X}_t : t \in [0,1]\}$ be a stochastic process with sample paths in D. Then the following statements are equivalent.*

(i) There exist a set $T \subset [0,1]$ containing 0 and 1 and all but at most countably many points of $[0,1]$, a sequence $\{a_n\}$, $0 < a_n \uparrow \infty$, and a collection $\{m_{t_1 \dots t_k} : k \in \mathbb{N}, t_i \in T\}$ of Radon measures with $m(\overline{\mathbb{R}}^{dk} \backslash \mathbb{R}^{dk}) = 0$, and m_t nonzero for some $t \in T$, such that

$$n\,\mathbb{P}(a_n^{-1}(\mathbf{X}_{t_1}, \dots, \mathbf{X}_{t_k}) \in \cdot\,) \xrightarrow{v} m_{t_1 \dots t_k}(\cdot) \quad \text{on } \mathcal{B}(\overline{\mathbb{R}}^{dk} \backslash \{\mathbf{0}\}) \tag{4.6}$$

holds whenever $t_1, \dots, t_k \in T$. Moreover, for any $\epsilon > 0$ and $\eta > 0$, there exist a $\delta \in (0,1)$ and an integer n_0 such that

$$n\,\mathbb{P}(w''(\mathbf{X}, \delta) \ge a_n \epsilon) \le \eta, \qquad n \ge n_0, \tag{4.7}$$

$$n\,\mathbb{P}(w(\mathbf{X}, [0, \delta)) \ge a_n \epsilon) \le \eta, \qquad n \ge n_0, \tag{4.8}$$

and

$$n\,\mathbb{P}(w(\mathbf{X}, [1-\delta, 1)) \ge a_n \epsilon) \le \eta, \quad n \ge n_0. \tag{4.9}$$

(ii) There exist a sequence $\{a_n\}$, $0 < a_n \uparrow \infty$, *and a nonzero boundedly finite measure* m *on* $\mathcal{B}(\overline{D})$ *with* $m(\overline{D}\backslash D) = 0$, *such that*

$$n\,\mathbb{P}(a_n^{-1}\mathbf{X} \in \cdot\,) \xrightarrow{\hat{w}} m(\cdot) \quad \text{on } \mathcal{B}(\overline{D}). \tag{4.10}$$

The sequences $\{a_n\}$ *in (i) and (ii) can be taken to be equal. Moreover, the measure* m *in (ii) is uniquely determined by* $\{m_{t_1\ldots t_k} : k \in \mathbb{N}, t_i \in T\}$.

4.2.1 Proofs

Proof of Theorem 4.5. Let $N_h = \{\mathbf{x} \in D : h(\mathbf{x}) = \mathbf{0}\}$ and define $\overline{h} : \overline{D}\backslash N_h \to \overline{D}$ by

$$\overline{h}(\mathbf{x}) = \begin{cases} h(\mathbf{x}) & \text{if } \mathbf{x} \in D\backslash N_h, \\ \mathbf{x} & \text{if } \mathbf{x} \in \overline{D}\backslash D. \end{cases}$$

Then $D_{\overline{h}} \subset (D_h \cap \overline{D}) \cup (\overline{D}\backslash D)$, where $D_{\overline{h}}$ denotes the set of points of $\overline{D}\backslash N_h$ where $\overline{h}$ is discontinuous. Take arbitrary bounded $B \in \mathcal{B}(\overline{D})$ with $m(\overline{h}^{-1}(\partial B)) = 0$. Since $\partial\overline{h}^{-1}(B) \subset \overline{h}^{-1}(\partial B) \cup D_{\overline{h}}$, we have $m(\partial\overline{h}^{-1}(B)) \le m(\overline{h}^{-1}(\partial B)) + m(D_{\overline{h}}) = 0$. Hence

$$\begin{aligned} n\,\mathbb{P}(a_n^{-1}h(\mathbf{X}) \in B) &= n\,\mathbb{P}(a_n^{-1}h(\mathbf{X}) \in B, h(\mathbf{X}) \neq \mathbf{0}) \\ &= n\,\mathbb{P}(a_n^{-1}\overline{h}(\mathbf{X}) \in B, h(\mathbf{X}) \neq \mathbf{0}) \\ &= n\,\mathbb{P}(a_n^{-1}\mathbf{X} \in \overline{h}^{-1}(B) \cap (\overline{D}\backslash N_h)) \\ &= n\,\mathbb{P}(a_n^{-1}\mathbf{X} \in \overline{h}^{-1}(B)) \\ &\to m(\overline{h}^{-1}(B)). \end{aligned}$$

Hence, by Proposition A2.6.II p. 628 in Daley and Vere-Jones [6],

$$n\,\mathbb{P}(a_n^{-1}h(\mathbf{X}) \in \cdot\,) \xrightarrow{\hat{w}} m \circ \overline{h}^{-1}(\cdot) \quad \text{on } \mathcal{B}(\overline{D}).$$

However, for every $B \in \mathcal{B}(\overline{D})$, $m(\overline{h}^{-1}(B)) = m(h^{-1}(B \cap D))$. Hence

$$n\,\mathbb{P}(a_n^{-1}h(\mathbf{X}) \in \cdot\,) \xrightarrow{\hat{w}} m \circ h^{-1}(\cdot \cap D) \quad \text{on } \mathcal{B}(\overline{D}).$$

The proof for mappings $h : D \to \mathbb{R}^k$ is similar. □

Proof of Theorem 4.6. (i) ⇒ (ii) Let $m_n(\cdot) = n\,\mathbb{P}(a_n^{-1}\mathbf{X} \in \cdot\,)$. First we will show that the sequence $\{m_n\}$ is relatively compact in the $\hat{w}$-topology. To prove this we will apply Proposition A2.6.IV p. 630 in Daley and Vere-Jones [6], which says that it is sufficient that the restrictions $\{m_{n,\gamma}\}$ to a sequence of closed spheres $S_\gamma \uparrow \overline{D}$ are relatively compact in the weak topology. For $\gamma > 0$, let $S_\gamma = \{\mathbf{x} \in \overline{D} : |\mathbf{x}|_\infty \ge \gamma\}$, and for $n \ge 1$, let $m_{n,\gamma}(\cdot) = n\,\mathbb{P}(a_n^{-1}\mathbf{X} \in \cdot \cap S_\gamma)$. We will show that, for every $\gamma > 0$, the family $\{m_{n,\gamma}\}$ is uniformly bounded and that it is relatively compact in the weak topology.

Take $\gamma > 0$ and $t_1, \ldots, t_k \in T$ with $0 = t_1 < \cdots < t_k = 1$ and $t_i - t_{i-1} < \delta$, where $\delta > 0$ is such that $n\,\mathbb{P}(w''(\mathbf{X}, \delta) \geq a_n\gamma/2) \leq \eta$ for $n \geq n_0$. Then

$$\begin{aligned} m_{n,\gamma}(\overline{D}) &= n\,\mathbb{P}(|\mathbf{X}|_\infty \geq a_n\gamma) \\ &\leq n\,\mathbb{P}(\max_{1\leq i\leq k} |\mathbf{X}_{t_i}| \geq a_n\gamma/2 \text{ or } w''(\mathbf{X}, \delta) \geq a_n\gamma/2) \\ &\leq n\,\mathbb{P}(\max_{1\leq i\leq k} |\mathbf{X}_{t_i}| \geq a_n\gamma/2) + n\,\mathbb{P}(w''(\mathbf{X}, \delta) \geq a_n\gamma/2) \\ &=: f_n(\gamma) + g_n(\gamma). \end{aligned}$$

By (4.6), $\{f_n(\gamma)\}$ converges to some finite limit as $n \to \infty$ and hence the sequence $\{f_n(\gamma)\}$ is bounded. Moreover, $g_n(\gamma) \leq \eta$ for $n \geq n_0$, and clearly $g_n(\gamma) < n_0$ for $n < n_0$. Hence, $\sup_{n\geq 1} m_{n,\gamma}(\overline{D}) < \infty$, i.e. $\{m_{n,\gamma}\}$ is uniformly bounded.

Since $m_n(\cdot) = n\,\mathbb{P}(a_n^{-1}\mathbf{X} \in \cdot\,) < n_0$ for $n < n_0$ and since a probability measure $\mathbb{P}(a_n^{-1}\mathbf{X} \in \cdot\,)$ on $\mathcal{B}(D)$ is tight it follows by Theorem 15.3 p. 125 in Billingsley [2] that (4.7), (4.8) and (4.9) hold for the finitely many n preceding n_0 by taking δ small enough. Hence, we may assume that $n_0 = 1$. Note that $[\gamma, \infty] \times K_1 \in \mathcal{B}(\overline{D})$ is compact in $\overline{D}$ if and only if K_1 is compact in D_1. For any $\eta > 0$, by (4.7), (4.8) and (4.9), we can choose δ_k such that, if

$$\begin{aligned} A_{k,1} &= \{\mathbf{x} \in D_1 : w''(\mathbf{x}, \delta_k) < 1/k\}, \\ A_{k,2} &= \{\mathbf{x} \in D_1 : w(\mathbf{x}, [0, \delta_k)) < 1/k\}, \\ A_{k,3} &= \{\mathbf{x} \in D_1 : w(\mathbf{x}, [1 - \delta_k, 1)) < 1/k\}, \end{aligned}$$

then $m_{n,\gamma}([\gamma, \infty] \times (D_1 \backslash A_{k,j})) \leq (1/3)\eta/2^k$ for every j and n. Let $B = \cap_{k=1}^{\infty} \cap_{j=1}^{3} A_{k,j}$. If K_1 is the closure of B, then by Theorem 14.4 p. 119 in Billingsley [2], K_1 is compact in D_1. Moreover, for every n,

$$\begin{aligned} m_{n,\gamma}(\overline{D} \backslash ([\gamma, \infty] \times K_1)) &\leq m_{n,\gamma}([\gamma, \infty] \times (D_1 \backslash B)) \\ &\leq \sum_{k=1}^{\infty} \sum_{j=1}^{3} m_{n,\gamma}([\gamma, \infty] \times (D_1 \backslash A_{k,j})) \\ &\leq \eta \sum_{k=1}^{\infty} 2^{-k} = \eta. \end{aligned}$$

Hence, we have shown that $\{m_{n,\gamma}\}$ is uniformly bounded and tight. It follows from Prohorov's Theorem (Theorem A2.4.I p. 619 in Daley and Vere Jones [6]) that $\{m_{n,\gamma}\}$ is relatively compact in the weak topology. Thus, by Proposition A2.6.IV p. 630 in Daley and Vere-Jones [6], $\{n\,\mathbb{P}(a_n^{-1}\mathbf{X} \in \cdot\,)\}$ is relatively compact in the $\hat{w}$-topology. We will now show that any subsequential $\hat{w}$-limit m satisfies $m(\overline{D} \backslash D) = 0$. By (4.7) and the above argument we can choose u_1 and δ such that $n\,\mathbb{P}(w''(\mathbf{X}, \delta) \geq a_n u_1/2) \leq \eta/2$ for every $n \geq 1$ (i.e. we may take $n_0 = 1$ in (4.7)).

By (4.6) and Theorem 4.5 (for mappings $h : \mathbb{R}^k \to [0,\infty)$) there exist a Radon measure ν on $\mathcal{B}((0,\infty])$ with $\nu(\{\infty\}) = 0$ such that

$$\nu_n(\cdot) := n\,\mathbb{P}(a_n^{-1} \max_{1\le i\le k} |\mathbf{X}_{t_k}| \in \cdot\ ) \xrightarrow{v} \nu(\cdot) \quad \text{on } \mathcal{B}((0,\infty]).$$

It follows that ν has the scaling property described in Remark 4.2 (the same proof applies with the obvious notational changes). Hence there exists an $\alpha > 0$ such that $\nu([x,\infty]) = x^{-\alpha}\nu([1,\infty])$ for every $x > 0$. Choose x such that $\nu([x/2,\infty]) \le \eta/4$. Then there exists n' such that $\nu_n([x/2,\infty]) \le \eta/2$ for $n \ge n'$. Clearly there exists x' such that $\nu_n([x'/2,\infty]) \le \eta/2$ for $n < n'$. Hence, with $u_2 = \max(x, x')$, $\nu_n([u_2/2,\infty]) \le \eta/2$ for every $n \ge 1$. Hence, with $u = \max(u_1, u_2)$, for every $n \ge 1$,

$$\begin{aligned} n\,\mathbb{P}(|\mathbf{X}|_\infty \ge a_n u) &\le n\,\mathbb{P}(\max_{1\le i\le k} |\mathbf{X}_{t_i}| \ge a_n u/2) + n\,\mathbb{P}(w''(\mathbf{X},\delta) \ge a_n u/2) \\ &\le \eta/2 + \eta/2 = \eta. \end{aligned}$$

Suppose $m_{n'} \xrightarrow{\hat{w}} m$. We have just shown that for any $\eta > 0$ there exists $u > 0$ such that $m_{n'}(\{\mathbf{x} \in \overline{D} : |\mathbf{x}|_\infty > u\}) \le \eta$ for $n' \ge 1$. In particular, this implies that $m_{n'}(\{\mathbf{x} \in \overline{D} : |\mathbf{x}|_\infty > u\}) \to 0$ uniformly in n' as $u \to \infty$. Since $G_u = \{\mathbf{x} \in \overline{D} : |\mathbf{x}|_\infty > u\}$ is open and bounded we have $m(G_u) \le \liminf_{n'\to\infty} m_{n'}(G_u)$ and because of uniform convergence

$$m(\overline{D}\backslash D) = \lim_{u\to\infty} m(G_u) \le \lim_{u\to\infty} \liminf_{n'\to\infty} m_{n'}(G_u) = \liminf_{n'\to\infty} \lim_{u\to\infty} m_{n'}(G_u) = 0.$$

Let m and $\widetilde{m}$ be two subsequential $\hat{w}$-limits. We will show that $m = \widetilde{m}$ and that m is uniquely determined by $\{m_{t_1\dots t_k} : k \in \mathbb{N}, t_i \in T\}$. Let T_m and $T_{\widetilde{m}}$ consist of those $t \in [0,1]$ for which the projection π_t is continuous except at points forming a set of m-measure 0 and $\widetilde{m}$-measure 0, respectively. Then, by Theorem 4.5, for $t_1, \dots, t_k \in T_m \cap T_{\widetilde{m}} \cap T$,

$$m \circ \pi_{t_1\dots t_k}^{-1}(\cdot \cap \mathbb{R}^{dk}) = \widetilde{m} \circ \pi_{t_1\dots t_k}^{-1}(\cdot \cap \mathbb{R}^{dk}) = m_{t_1\dots t_k}(\cdot) \quad \text{on } \mathcal{B}(\overline{\mathbb{R}}^{dk}\backslash\{\mathbf{0}\}).$$

Since T_m, $T_{\widetilde{m}}$ and T each contain all but countably many points of $[0,1]$, the same is true for $T_m \cap T_{\widetilde{m}} \cap T$, in particular $T_m \cap T_{\widetilde{m}} \cap T$ is dense in $[0,1]$. Moreover, $0, 1 \in T_m \cap T_{\widetilde{m}} \cap T$. With some minor modifications of Theorem 14.5 p. 121 in Billingsley [2] one can show that

$$\{\pi_{t_1\dots t_k}^{-1}(H) : k \in \mathbb{N}, H \in \mathcal{B}(\overline{\mathbb{R}}^{dk}\backslash\{\mathbf{0}\}) \cap \mathbb{R}^{dk}, t_1, \dots, t_k \in T_m \cap T_{\widetilde{m}} \cap T\}$$

generates $\mathcal{B}(\overline{D}) \cap D$. Hence $\widetilde{m}$ and m coincides on $\mathcal{B}(\overline{D}) \cap D$ and since $m(\overline{D}\backslash D) = \widetilde{m}(\overline{D}\backslash D) = 0$ we have $m = \widetilde{m}$.

(ii) $\Rightarrow$ (i) Let T_m consist of those t in $[0,1]$ for which the projection π_t from D to $\mathbb{R}^d$ is continuous except at points forming a set of m-measure 0. The projections π_0 and π_1 are continuous and hence $0, 1 \in T_m$. For $t \in (0,1)$, π_t is continuous if and

only if $m(\{\mathbf{x} : \mathbf{x}_t \neq \mathbf{x}_{t-}\}) = 0$. By the same arguments as in Billingsley [2] p. 124 there are at most countably many $t \in (0,1]$ such that $m(\{\mathbf{x} : \mathbf{x}_t \neq \mathbf{x}_{t-}\}) > 0$. Then, since m is nonzero and T_m is dense in $[0,1]$, there exists $t \in T_m$ such that m_t is nonzero. Moreover, $\pi_{t_1 \ldots t_k}$ is continuous except at points forming a set of m-measure 0 if $t_1, \ldots, t_k \in T_m$. Hence, by Theorem 4.5, for $t_1, \ldots, t_k \in T_m$,

$$\begin{aligned} n\,\mathbb{P}(a_n^{-1}(\mathbf{X}_{t_1}, \ldots, \mathbf{X}_{t_k}) \in \cdot\ ) &= n\,\mathbb{P}(a_n^{-1}\mathbf{X} \in \pi_{t_1 \ldots t_k}^{-1}(\,\cdot \cap \mathbb{R}^{dk})) \\ &\xrightarrow{v} m \circ \pi_{t_1 \ldots t_k}^{-1}(\,\cdot \cap \mathbb{R}^{dk}) \quad \text{on } \mathcal{B}(\overline{\mathbb{R}}^{dk} \backslash \{\mathbf{0}\}). \end{aligned}$$

For $t_1, \ldots, t_k \in T_m$, let $m_{t_1 \ldots t_k}(\cdot) = m \circ \pi_{t_1 \ldots t_k}^{-1}(\,\cdot \cap \mathbb{R}^{dk})$.

For $n \geq 1$, let $m_n(\cdot) = n\,\mathbb{P}(a_n^{-1}\mathbf{X} \in \cdot\ )$. By the scaling property of m, the set $S_u = \{\mathbf{x} \in \overline{D} : |\mathbf{x}|_\infty \geq u\}$ is an m-continuity set for every $u > 0$. Hence, $m_n(S_u) \to m(S_u) = u^{-\alpha} m(S_1)$ for every $u > 0$. Choose u such that $u^{-\alpha} m(S_1) \leq \eta/4$. Then there exists n_1 such that $m_n(S_u) \leq \eta/2$ for $n \geq n_1$. By Proposition A2.6.IV p. 630 in Daley and Vere-Jones [6], for every $0 < \gamma < u < \infty$, $\{m_n(\,\cdot \cap \{\mathbf{x} \in D : |\mathbf{x}|_\infty \in [\gamma, u]\})\}$ is relatively compact in the weak topology on $\overline{D}$. Since $\{\mathbf{x} \in D : |\mathbf{x}|_\infty \in [\gamma, u]\} \subset D \backslash \{\mathbf{0}\}$ and on this subspace the subspace topologies (of $\overline{D}$ and D) coincide it follows that $\{m_n(\,\cdot \cap \{\mathbf{x} \in D : |\mathbf{x}|_\infty \in [\gamma, u]\})\}$ is relatively compact in the weak topology on D. Hence, by Theorem 15.3 p. 125 in Billingsley [2], for any $\epsilon > 0$ and $\eta > 0$ there exist $\delta \in (0,1)$ and integer n_2 such that

$$\begin{aligned} &n\,\mathbb{P}(w''(\mathbf{X}, \delta) \geq a_n \epsilon, |\mathbf{X}|_\infty \in a_n[\gamma, u]) \leq \eta/2, \quad n \geq n_2, \\ &n\,\mathbb{P}(w(\mathbf{X}, [0, \delta)) \geq a_n \epsilon, |\mathbf{X}|_\infty \in a_n[\gamma, u]) \leq \eta/2, \quad n \geq n_2, \\ &\text{and} \\ &n\,\mathbb{P}(w(\mathbf{X}, [1-\delta, 1)) \geq a_n \epsilon, |\mathbf{X}|_\infty \in a_n[\gamma, u]) \leq \eta/2, \quad n \geq n_2. \end{aligned}$$

In particular the three inequalities above hold, with $\eta/2$ replaced by η and n_2 replaced by $n_0 = \max(n_1, n_2)$, for $u = \infty$ and $\gamma \leq \epsilon/2$ and for such γ they coincide with (4.7), (4.8) and (4.9).

□

4.3 Markov processes with asymptotically independent increments

In this section we will study Markov processes with increments that are not too strongly dependent in the sense that an extreme jump does not trigger further jumps or oscillations of the same magnitude with a nonnegligible probability. We will derive surprisingly concrete results for such Markov processes (see Theorem 4.8 and 4.12) which will prove very useful when used in combination with Theorem 4.5 (see e.g. Example 4.14 and 4.15).

Let $\{\mathbf{X}_t : t \in [0,1]\}$ be a Markov process on $\mathbb{R}^d$ with transition function $P_{s,t}(x,B)$. For $r \geq 0$ and $0 \leq u \leq T \leq 1$ define

$$\alpha_{r,T}(u) = \sup\{P_{s,t}(\mathbf{x}, B^c_{\mathbf{x},r}) : \mathbf{x} \in \mathbb{R}^d \text{ and } s,t \in [0,T], t-s \in [0,u]\}.$$

Note that if the random vectors $\mathbf{Y}$ and $\widetilde{\mathbf{Y}}$ are independent and, for some sequence $\{a_n\}$, $0 < a_n \uparrow \infty$, and nonzero Radon measures m and $\widetilde{m}$ with $m(\overline{\mathbb{R}}^d \backslash \mathbb{R}^d) = \widetilde{m}(\overline{\mathbb{R}}^d \backslash \mathbb{R}^d) = 0$ we have

$$n\,\mathbb{P}(a_n^{-1}\mathbf{Y} \in \cdot\,) \xrightarrow{v} m(\cdot) \quad \text{and} \quad n\,\mathbb{P}(a_n^{-1}\widetilde{\mathbf{Y}} \in \cdot\,) \xrightarrow{v} \widetilde{m}(\cdot) \quad \text{on } \mathcal{B}(\overline{\mathbb{R}}^d \backslash \{\mathbf{0}\}),$$

then, by Proposition 3.11 in Chapter 3,

$$n\,\mathbb{P}(a_n^{-1}(\mathbf{Y} + \widetilde{\mathbf{Y}}) \in \cdot\,) \xrightarrow{v} m(\cdot) + \widetilde{m}(\cdot) \quad \text{on } \mathcal{B}(\overline{\mathbb{R}}^d \backslash \{\mathbf{0}\}),$$

i.e. the limit measure of the sum is the sum of the limit measures. If a regularly varying Markov process satisfies $\alpha_{r,1}(1) \to 0$ as $r \to \infty$, then it has asymptotically independent increments in the above sense (with $\mathbf{Y}$ and $\widetilde{\mathbf{Y}}$ representing two nonoverlapping increments).

Lemma 4.7. *Let $\{\mathbf{X}_t : t \in [0,1]\}$ be a Markov process on $\mathbb{R}^d$ such that $\alpha_{r,1}(1) \to 0$ as $r \to \infty$. Fix arbitrary $s,t \in [0,1]$ with $s < t$. Let $\{a_n\}$ be a sequence with $0 < a_n \uparrow \infty$, and let m_s, m_t and μ be Radon measures on $\mathcal{B}(\overline{\mathbb{R}}^d \backslash \{\mathbf{0}\})$ with $m_s(\overline{\mathbb{R}}^d \backslash \mathbb{R}^d) = m_t(\overline{\mathbb{R}}^d \backslash \mathbb{R}^d) = \mu(\overline{\mathbb{R}}^d \backslash \mathbb{R}^d) = 0$. Consider the following statements.*

$$n\,\mathbb{P}(a_n^{-1}\mathbf{X}_s \in \cdot\,) \xrightarrow{v} m_s(\cdot) \quad \text{on } \mathcal{B}(\overline{\mathbb{R}}^d \backslash \{\mathbf{0}\}), \tag{4.11}$$

$$n\,\mathbb{P}(a_n^{-1}\mathbf{X}_t \in \cdot\,) \xrightarrow{v} m_t(\cdot) \quad \text{on } \mathcal{B}(\overline{\mathbb{R}}^d \backslash \{\mathbf{0}\}), \tag{4.12}$$

$$n\,\mathbb{P}(a_n^{-1}(\mathbf{X}_t - \mathbf{X}_s) \in \cdot\,) \xrightarrow{v} \mu(\cdot) \quad \text{on } \mathcal{B}(\overline{\mathbb{R}}^d \backslash \{\mathbf{0}\}). \tag{4.13}$$

If any two of the above three statements hold, then the third also holds and the limit measures are related through $m_t = m_s + \mu$.

It turns out that for a strong Markov process with sample paths in D satisfying $\alpha_{r,1}(1) \to 0$ as $r \to \infty$ we can obtain sufficient conditions for regular variation on D, which are easier to verify than the general conditions of Theorem 4.6.

Theorem 4.8. *Let $\mathbf{X} = \{\mathbf{X}_t : t \in [0,1]\}$ be a strong Markov process with sample paths in D such that $\alpha_{r,1}(1) \to 0$ as $r \to \infty$. Suppose there exist a set $T \subset [0,1]$ containing 0 and 1 and all but at most countably many points of $[0,1]$, a sequence*

$\{a_n\}$, $0 < a_n \uparrow \infty$, and a collection $\{m_t : t \in T\}$ of Radon measures on $\mathcal{B}(\overline{\mathbb{R}}^d \backslash \{\mathbf{0}\})$, with $m_t(\overline{\mathbb{R}}^d \backslash \mathbb{R}^d) = 0$ and with m_1 nonzero, such that

$$n\,\mathbb{P}(a_n^{-1}\mathbf{X}_t \in \cdot\,) \xrightarrow{v} m_t(\cdot) \quad \text{on } \mathcal{B}(\overline{\mathbb{R}}^d \backslash \{\mathbf{0}\}) \text{ for every } t \in T, \tag{4.14}$$

and such that, for any $\epsilon > 0$ and $\eta > 0$ there exists a $\delta > 0$, $\delta \in T$, $1 - \delta \in T$ such that

$$m_\delta(B^c_{\mathbf{0},\epsilon}) - m_0(B^c_{\mathbf{0},\epsilon}) \le \eta \quad \text{and} \quad m_1(B^c_{\mathbf{0},\epsilon}) - m_{1-\delta}(B^c_{\mathbf{0},\epsilon}) \le \eta. \tag{4.15}$$

Then there exists a nonzero boundedly finite measure m on $\mathcal{B}(\overline{D})$ with $m(\overline{D}\backslash D) = 0$, such that, as $n \to \infty$,

$$n\,\mathbb{P}(a_n^{-1}\mathbf{X} \in \cdot\,) \xrightarrow{\hat{w}} m(\cdot) \quad \text{on } \mathcal{B}(\overline{D}).$$

Moreover, the measure m is uniquely determined by $\{m_t : t \in T\}$.

Example 4.9. If $\{\mathbf{X}_t : t \ge 0\}$ is a Lévy process, then it is strong Markov, it has sample paths in D, $\alpha_{r,t}(t) \to 0$ as $r \to \infty$ for every $t > 0$ and, if (4.14) holds, then it is not difficult to show that (4.15) holds and Theorem 4.8 applies.

Remark 4.10. By Theorem 11.1 p. 59 in Sato [14], a sufficient condition for a Markov process $\{\mathbf{X}_t : t \ge 0\}$ on $\mathbb{R}^d$ to have a version in D is that

$$\lim_{u \downarrow 0} \alpha_{\epsilon,T}(u) = 0 \quad \text{for any } \epsilon > 0 \text{ and } T > 0.$$

Theorem 11.1 considers Markov processes with fixed starting points, however this requirement can be dropped as seen from the proof. Note that the above condition implies stochastic continuity.

Remark 4.11. We have chosen to formulate the above theorem and results below for strong Markov processes, since it is convenient to use the strong Markov property in some of the proofs. We could however, instead of the strong Markov property, have assumed that we have the Markov property and that, for every $\epsilon > 0$, $n\alpha^2_{a_n\epsilon,1}(1) \to 0$ as $n \to \infty$.

It turns out that a regularly varying Markov process with asymptotically independent increments, in the sense of Lemma 4.7, has a very simple extremal behavior. In this case the process reaches a set far away from the origin by making at most one jump to that set (it might start there at time 0 since we allow for a regularly varying starting point) and the process essentially stays constant before and after the jump. This is formalized in the next theorem. Let

$$\mathcal{V} = \{\mathbf{x} \in D : \mathbf{x} = \mathbf{y}1_{[v,1]}, v \in [0,1], \mathbf{y} \in \mathbb{R}^d \backslash \{\mathbf{0}\}\}.$$

Theorem 4.12. *Let $\mathbf{X} = \{\mathbf{X}_t : t \in [0,1]\}$ be a strong Markov process with sample paths in D such that $\alpha_{r,1}(1) \to 0$ as $r \to \infty$. Suppose there exist a sequence $\{a_n\}$,*

$0 < a_n \uparrow \infty$, and a nonzero boundedly finite measure m on $\mathcal{B}(\overline{D})$ with $m(\overline{D}\backslash D) = 0$ such that, as $n \to \infty$,

$$n\,\mathbb{P}(a_n^{-1}\mathbf{X} \in \cdot\,) \xrightarrow{\hat{w}} m(\cdot) \quad \text{on } \mathcal{B}(\overline{D}).$$

Then $m(\mathcal{V}^c) = 0$. Moreover, there exist an $\alpha > 0$ and a probability measure σ on $\mathcal{B}(D_1)$ such that, for every $x > 0$, as $u \to \infty$,

$$\frac{\mathbb{P}(|\mathbf{X}|_\infty > ux, \mathbf{X}/|\mathbf{X}|_\infty \in \cdot\,)}{\mathbb{P}(|\mathbf{X}|_\infty > u)} \xrightarrow{w} x^{-\alpha}\sigma(\cdot) \quad \text{on } \mathcal{B}(D_1),$$

with $\sigma(\{\mathbf{x} \in D_1 : \mathbf{x} = \mathbf{y}1_{[v,1]}, v \in [0,1], \mathbf{y} \in \mathbb{S}^{d-1}\}) = 1$ and

$$\sigma(\{\mathbf{x} \in D_1 : \mathbf{x} = \mathbf{y}1_{[v,1]}, v \in [0,1], \mathbf{y} \in \cdot\,\})$$

coincides with the spectral measure of $\mathbf{X}_1$ on $\mathcal{B}(\mathbb{S}^{d-1})$.

For Lévy processes we can be even more explicit.

Example 4.13. Let $\mathbf{X} = \{\mathbf{X}_t : t \in [0,1]\}$ be a Lévy process on $\mathbb{R}^d$. Suppose there exist a sequence $\{a_n\}$, $0 < a_n \uparrow \infty$, and a nonzero Radon measure m_1 with $m_1(\overline{\mathbb{R}}^d\backslash\mathbb{R}^d) = 0$ such that

$$n\,\mathbb{P}(a_n^{-1}\mathbf{X}_1 \in \cdot\,) \xrightarrow{v} m_1(\cdot) \quad \text{on } \mathcal{B}(\overline{\mathbb{R}}^d\backslash\{\mathbf{0}\}).$$

Then, since $\mathbf{X}$ has stationary and independent increments, $n\,\mathbb{P}(a_n^{-1}\mathbf{X}_t \in \cdot) \xrightarrow{v} tm_1(\cdot)$ on $\mathcal{B}(\overline{\mathbb{R}}^d\backslash\{\mathbf{0}\})$ for every $t \in [0,1]$. Hence, combining Theorems 4.3 and 4.8 gives, for every $x > 0$, as $u \to \infty$,

$$\frac{\mathbb{P}(|\mathbf{X}|_\infty > ux, \mathbf{X}/|\mathbf{X}|_\infty \in \cdot\,)}{\mathbb{P}(|\mathbf{X}|_\infty > u)} \xrightarrow{w} x^{-\alpha}\sigma(\cdot) \quad \text{on } \mathcal{B}(D_1),$$

where $\alpha > 0$ is the tail index of $\mathbf{X}_1$ and it follows that

$$\sigma(\cdot) = \mathbb{P}(\{\mathbf{Z}1_{[V,1]}(t), t \in [0,1]\} \in \cdot\,),$$

where $\mathbf{Z}$ and V are independent, the distribution of $\mathbf{Z}$ is the spectral measure of $\mathbf{X}_1$ and V is uniformly distributed on $[0,1]$. The random vector $\mathbf{Z}$ is the direction of the big jump and V is the time of the big jump.

The following two examples illustrate the usefulness of Theorem 4.12 in combination with the Continuous Mapping Theorem.

Example 4.14. Let $\mathbf{X}$ be a strong Markov process with $\mathbf{X}_0 = \mathbf{0}$ satisfying the conditions in Theorem 4.12 and let $h : D \to \mathbb{R}^d$ be defined by

$$h(\mathbf{x}) = \Big(\sup_{t\in[0,1]} x_t^{(1)}, \ldots, \sup_{t\in[0,1]} x_t^{(d)} \Big).$$

Then it is straightforward to show that h satisfies the conditions of Theorem 4.5. Hence,

$$n\,\mathbb{P}(a_n^{-1}h(\mathbf{X}) \in \cdot\,) \xrightarrow{v} m \circ h^{-1}(\cdot \cap \mathbb{R}^d) \quad \text{on } \mathcal{B}(\overline{\mathbb{R}}^d\backslash\{\mathbf{0}\})$$

and for $B \in \mathcal{B}(\overline{\mathbb{R}}^d\backslash\{\mathbf{0}\})$, with $\mathbb{R}^d_+ = [0,\infty)$,

$$\begin{aligned} m \circ h^{-1}(B \cap \mathbb{R}^d) &= m(\{\mathbf{x} \in D : h(\mathbf{x}) \in B \cap \mathbb{R}^d\} \cap \mathcal{V}) \\ &= m(\{\mathbf{x} \in D : \mathbf{x} = \mathbf{y}1_{[v,1]}, v \in [0,1], \mathbf{y} \in B \cap \mathbb{R}^d_+\}) \\ &= m(\pi_1^{-1}(B \cap \mathbb{R}^d_+) \cap \mathcal{V}) \\ &= m(\pi_1^{-1}(B \cap \mathbb{R}^d_+)) \\ &= m_1(B \cap \mathbb{R}^d_+), \end{aligned}$$

where m_1 is vague limit of $\{n\,\mathbb{P}(a_n^{-1}\mathbf{X}_1 \in \cdot\,)\}$.

Example 4.15. Let $\mathbf{X}$ be a strong Markov process with $\mathbf{X}_0 = \mathbf{0}$ satisfying the conditions in Theorem 4.12 and let $h : D \to \mathbb{R}^d$ be defined by

$$h(\mathbf{x}) = \Big(\int_0^1 x_t^{(1)}\mathrm{d}t, \ldots, \int_0^1 x_t^{(d)}\mathrm{d}t \Big).$$

Then it is straightforward to show that h satisfies the conditions of Theorem 4.5. Hence,

$$n\,\mathbb{P}(a_n^{-1}h(\mathbf{X}) \in \cdot\,) \xrightarrow{v} m \circ h^{-1}(\cdot \cap \mathbb{R}^d) \quad \text{on } \mathcal{B}(\overline{\mathbb{R}}^d\backslash\{\mathbf{0}\})$$

and for $B \in \mathcal{B}(\overline{\mathbb{R}}^d\backslash\{\mathbf{0}\})$

$$\begin{aligned} m \circ h^{-1}(B \cap \mathbb{R}^d) &= m(\{\mathbf{x} \in D : h(\mathbf{x}) \in B \cap \mathbb{R}^d\} \cap \mathcal{V}) \\ &= m(\{\mathbf{x} \in D : \mathbf{x} = \mathbf{y}1_{[v,1]}, v \in [0,1], \mathbf{y}(1-v) \in B \cap \mathbb{R}^d\}) \\ &= m(\{\mathbf{x} \in D : \mathbf{x}_t \in \frac{1}{1-t}B \cap \mathbb{R}^d \text{ some } t \in [0,1]\} \cap \mathcal{V}) \\ &= m(\{\mathbf{x} \in D : \mathbf{x}_t \in \frac{1}{1-t}B \cap \mathbb{R}^d \text{ some } t \in [0,1]\}). \end{aligned}$$

In particular, if $\{\mathbf{X}_t : t \in [0,1]\}$ is a Lévy process, then the last expression reduces to (see also Chapter 3)

$$\int_0^1 m_1(\frac{1}{1-s}B \cap \mathbb{R}^d)\mathrm{d}s = m_1(B \cap \mathbb{R}^d)\int_0^1 (1-s)^\alpha \mathrm{d}s = \frac{1}{\alpha+1}m_1(B),$$

where m_1 is vague limit of $\{n\,\mathbb{P}(a_n^{-1}\mathbf{X}_1 \in \cdot\,)\}$

4.3.1 Proofs

Proof of Lemma 4.7. Fix a relatively compact $B \in \mathcal{B}(\overline{\mathbb{R}}^d \backslash \{\mathbf{0}\})$. Then there exist $r > 0$ such that $B \subset B^c_{\mathbf{0},r}$. Note that the scaling property implies that sets of the form $B^c_{\mathbf{0},r}$, $r > 0$, are always m_s-, m_t- and μ-continuity sets.

Suppose that (4.11) and (4.12) hold. We first show that $\{n\,\mathbb{P}(a_n^{-1}(\mathbf{X}_t - \mathbf{X}_s) \in \cdot\,)\}$ is vaguely relatively compact. We have

$$\begin{aligned}\sup_{n\geq 1} n\,\mathbb{P}(a_n^{-1}(\mathbf{X}_t - \mathbf{X}_s) \in B) &\leq \sup_{n\geq 1} n\,\mathbb{P}(a_n^{-1}(\mathbf{X}_t - \mathbf{X}_s) \in B^c_{\mathbf{0},r}) \\ &\leq \sup_{n\geq 1} n\,\mathbb{P}(a_n^{-1}\mathbf{X}_s \in B^c_{\mathbf{0},r/2}) + \sup_{n\geq 1} n\,\mathbb{P}(a_n^{-1}\mathbf{X}_t \in B^c_{\mathbf{0},r/2}) < \infty,\end{aligned}$$

since $\{n\,\mathbb{P}(a_n^{-1}\mathbf{X}_s \in \cdot\,)\}$ and $\{n\,\mathbb{P}(a_n^{-1}\mathbf{X}_t \in \cdot\,)\}$ are vaguely relatively compact. Hence $\{n\,\mathbb{P}(a_n^{-1}(\mathbf{X}_t - \mathbf{X}_s) \in \cdot\,)\}$ is vaguely relatively compact.

By essentially the same argument it follows that if (4.11) and (4.13) hold, then $\{n\,\mathbb{P}(a_n^{-1}\mathbf{X}_t \in \cdot\,)\}$ is vaguely relatively compact, and if (4.12) and (4.13) hold, then $\{n\,\mathbb{P}(a_n^{-1}\mathbf{X}_s \in \cdot\,)\}$ is vaguely relatively compact.

Suppose that (4.11) and (4.12) hold. Let μ be a subsequential vague limit such that

$$n'\,\mathbb{P}(a_{n'}^{-1}(\mathbf{X}_t - \mathbf{X}_s) \in \cdot\,) \xrightarrow{v} \mu(\cdot).$$

Fix $\epsilon_1 > 0$, $\epsilon_2 > 0$ and a relatively compact $B \in \mathcal{B}(\overline{\mathbb{R}}^d \backslash \{\mathbf{0}\})$ with $m_s(\partial B) = \mu(\partial B) = 0$. We have

$$\begin{aligned}&n'\,\mathbb{P}(a_{n'}^{-1}(\mathbf{X}_s, \mathbf{X}_t - \mathbf{X}_s) \in B^c_{\mathbf{0},\epsilon_1} \times B^c_{\mathbf{0},\epsilon_2}) \\ &\quad= \underbrace{n'\,\mathbb{P}(a_{n'}^{-1}\mathbf{X}_s \in B^c_{\mathbf{0},\epsilon_1})}_{\to m_s(B^c_{\mathbf{0},\epsilon_1})} \underbrace{\mathbb{P}(a_{n'}^{-1}(\mathbf{X}_t - \mathbf{X}_s) \in B^c_{\mathbf{0},\epsilon_2} \mid a_{n'}^{-1}\mathbf{X}_s \in B^c_{\mathbf{0},\epsilon_1})}_{\leq \alpha_{a_{n'\epsilon_2,1}}(1) \to 0} \to 0.\end{aligned}$$

Since $m_s(\overline{\mathbb{R}}^d \backslash \mathbb{R}^d) = \mu(\overline{\mathbb{R}}^d \backslash \mathbb{R}^d) = 0$ we may without loss of generality assume that $B \cap \mathbb{R}^d \neq \emptyset$. Then,

$$\begin{aligned}&n'\,\mathbb{P}(a_{n'}^{-1}(\mathbf{X}_s, \mathbf{X}_t - \mathbf{X}_s) \in B \times B_{\mathbf{0},\epsilon_2}) \\ &\quad= \underbrace{n'\,\mathbb{P}(a_{n'}^{-1}\mathbf{X}_s \in B)}_{\to m_s(B)} (1 - \underbrace{\mathbb{P}(a_{n'}^{-1}(\mathbf{X}_t - \mathbf{X}_s) \in B^c_{\mathbf{0},\epsilon_2} \mid a_{n'}^{-1}\mathbf{X}_s \in B)}_{\leq \alpha_{a_{n'\epsilon_2,1}}(1) \to 0}) \to m_s(B).\end{aligned}$$

Clearly,

$$n'\,\mathbb{P}(a_{n'}^{-1}(\mathbf{X}_s, \mathbf{X}_t - \mathbf{X}_s) \in B_{\mathbf{0},\epsilon_1} \times B) \leq n'\,\mathbb{P}(a_{n'}^{-1}(\mathbf{X}_t - \mathbf{X}_s) \in B) \to \mu(B).$$

Set $\gamma = \inf_{\mathbf{x}\in B\cap\mathbb{R}^d} |\mathbf{x}|$. Then

$$\begin{aligned}
&n'\,\mathbb{P}(a_{n'}^{-1}(\mathbf{X}_s, \mathbf{X}_t - \mathbf{X}_s) \in B_{\mathbf{0},\epsilon_1} \times B) \\
&\quad = n'\,\mathbb{P}(a_{n'}^{-1}(\mathbf{X}_t - \mathbf{X}_s) \in B) \\
&\qquad - n'\,\mathbb{P}(a_{n'}^{-1}\mathbf{X}_s \in B_{\mathbf{0},\epsilon_1}^c)\,\mathbb{P}(a_{n'}^{-1}(\mathbf{X}_t - \mathbf{X}_s) \in B \mid a_{n'}^{-1}\mathbf{X}_s \in B_{\mathbf{0},\epsilon_1}^c) \\
&\quad \geq n'\,\mathbb{P}(a_{n'}^{-1}(\mathbf{X}_t - \mathbf{X}_s) \in B) \\
&\qquad - \underbrace{n'\,\mathbb{P}(a_{n'}^{-1}\mathbf{X}_s \in B_{\mathbf{0},\epsilon_1}^c)}_{\to m_s(B_{\mathbf{0},\epsilon_1}^c)}\underbrace{\mathbb{P}(a_{n'}^{-1}(\mathbf{X}_t - \mathbf{X}_s) \in B_{\mathbf{0},\gamma}^c \mid a_{n'}^{-1}\mathbf{X}_s \in B_{\mathbf{0},\epsilon_1}^c)}_{\leq \alpha_{a_{n'}\gamma,1}(1)\to 0} \to \mu(B).
\end{aligned}$$

It follows that $n'\,\mathbb{P}(a_{n'}^{-1}(\mathbf{X}_s, \mathbf{X}_t - \mathbf{X}_s) \in \cdot\,) \xrightarrow{v} \widehat{\mu}(\cdot)$ on $\mathcal{B}(\overline{\mathbb{R}}^{2d}\backslash\{\mathbf{0}\})$, where $\widehat{\mu}$ is a Radon measure which concentrates on $(\{\mathbf{0}\}\times\mathbb{R}^d)\cup(\mathbb{R}^d\times\{\mathbf{0}\})$. Hence

$$n'\,\mathbb{P}(a_{n'}^{-1}(\mathbf{X}_s + \mathbf{X}_t - \mathbf{X}_s) \in \cdot\,) \xrightarrow{v} \widehat{\mu}((\mathbf{x},\widetilde{\mathbf{x}}) : \mathbf{x}+\widetilde{\mathbf{x}} \in \cdot\,),$$

where

$$\begin{aligned}
\widehat{\mu}((\mathbf{x},\widetilde{\mathbf{x}}) : \mathbf{x}+\widetilde{\mathbf{x}} \in \cdot\,) &= \widehat{\mu}((\mathbf{x},\mathbf{0}) : \mathbf{x}+\mathbf{0} \in \cdot\,) + \widehat{\mu}((\mathbf{0},\widetilde{\mathbf{x}}) : \mathbf{0}+\widetilde{\mathbf{x}} \in \cdot\,) \\
&= m_s(\cdot) + \mu(\cdot).
\end{aligned}$$

However, $n'\,\mathbb{P}(a_{n'}^{-1}(\mathbf{X}_s + \mathbf{X}_t - \mathbf{X}_s) \in \cdot\,) \xrightarrow{v} m_t(\cdot)$ and hence $\mu = m_t - m_s$. Since this is true for any subsequential vague limit of $\{n\,\mathbb{P}(a_n^{-1}(\mathbf{X}_t - \mathbf{X}_s) \in \cdot\,)\}$ it follows that

$$n\,\mathbb{P}(a_n^{-1}(\mathbf{X}_t - \mathbf{X}_s) \in \cdot\,) \xrightarrow{v} m_t(\cdot) - m_s(\cdot).$$

Suppose now that (4.11) and (4.13) hold. By the same arguments as above, replacing n' by n, it follows that

$$\begin{aligned}
n\,\mathbb{P}(a_n^{-1}\mathbf{X}_t \in \cdot\,) &= n\,\mathbb{P}(a_n^{-1}(\mathbf{X}_s + \mathbf{X}_t - \mathbf{X}_s) \in \cdot\,) \\
&\xrightarrow{v} \widehat{\mu}((\mathbf{x},\widetilde{\mathbf{x}}) : \mathbf{x}+\widetilde{\mathbf{x}} \in \cdot\,) = m_s(\cdot) + \mu(\cdot).
\end{aligned}$$

Suppose now that (4.12) and (4.13) hold, and let m_s be a subsequential limit such that $n'\,\mathbb{P}(a_{n'}^{-1}\mathbf{X}_s \in \cdot\,) \xrightarrow{v} m_s(\cdot)$. By the same arguments as above it follows that

$$n'\,\mathbb{P}(a_{n'}^{-1}(\mathbf{X}_s + \mathbf{X}_t - \mathbf{X}_s) \in \cdot\,) \xrightarrow{v} \widehat{\mu}((\mathbf{x},\widetilde{\mathbf{x}}) : \mathbf{x}+\widetilde{\mathbf{x}} \in \cdot\,) = m_s(\cdot) + \mu(\cdot).$$

However, $n'\,\mathbb{P}(a_{n'}^{-1}(\mathbf{X}_s + \mathbf{X}_t - \mathbf{X}_s) \in \cdot\,) \xrightarrow{v} m_t(\cdot)$ along every subsequence $\{n'\}$ so we must have $m_s = m_t - \mu$. □

To prove Theorems 4.8 and 4.12 we need a couple of technical lemmas. For $\epsilon > 0$, positive integer p and $M \subset [0,1]$ we say that an element $\mathbf{x} \in D$ has ϵ-oscillation p times in M if there exist $t_0, \ldots, t_p \in M$ with $t_0 < \cdots < t_p$ such that $|\mathbf{x}_{t_i} - \mathbf{x}_{t_{i-1}}| > \epsilon$ for $i = 1, \ldots, p$. Let

$$B(p,\epsilon,M) = \{\mathbf{x} \in D : \mathbf{x} \text{ has } \epsilon\text{-oscillation } p \text{ times in } M\}.$$

The following lemma is an immediate consequence of Lemma 2 p. 420 in Gihman and Skorohod [7].

Lemma 4.16. *Let $\mathbf{X} = \{\mathbf{X} : t \in [0,1]\}$ be a Markov process with sample paths in D. If for $\epsilon > 0$ and $0 \le T_1 < T_2 \le 1$ the quantity $\alpha_{\epsilon/4,T_2}(T_2 - T_1)$ is less than 1, then*

$$\mathbb{P}(\mathbf{X} \in B(1,\epsilon,[T_1,T_2])) \le \frac{\alpha_{\epsilon/4,T_2}(T_2-T_1)}{1-\alpha_{\epsilon/4,T_2}(T_2-T_1)}.$$

Proof. First,

$$\begin{aligned}\mathbb{P}(\mathbf{X} \in B(1,\epsilon,[T_1,T_2])) &= \mathbb{P}(\sup_{s,t\in[T_1,T_2]} |\mathbf{X}_t - \mathbf{X}_s| > \epsilon) \\ &\le \mathbb{P}(\sup_{s\in[T_1,T_2]} |\mathbf{X}_s - \mathbf{X}_{T_1}| \ge \epsilon/2),\end{aligned}$$

and, by Lemma 2 p. 420 in Gihman and Skorohod [7],

$$\mathbb{P}(\sup_{s\in[T_1,T_2]} |\mathbf{X}_s - \mathbf{X}_{T_1}| \ge \epsilon/2) \le \frac{\mathbb{P}(|\mathbf{X}_{T_2} - \mathbf{X}_{T_1}| \ge \epsilon/4)}{1-\alpha_{\epsilon/4,T_2}(T_2-T_1)}.$$

Since $\mathbb{P}(|\mathbf{X}_{T_2} - \mathbf{X}_{T_1}| \ge \epsilon/4) \le \alpha_{\epsilon/4,T_2}(T_2 - T_1)$ the conclusion follows. □

Lemma 4.17. *Let $\mathbf{X} = \{\mathbf{X}_t : t \in [0,1]\}$ be a strong Markov process with sample paths in D such that $\alpha_{r,1}(1) \to 0$ as $r \to \infty$. Suppose there exist a sequence $\{a_n\}$, $0 < a_n \uparrow \infty$, and Radon measures m_0 and m_1 on $\mathcal{B}(\overline{\mathbb{R}}^d \backslash \{\mathbf{0}\}$ with $m_0(\overline{\mathbb{R}}^d \backslash \mathbb{R}^d) = m_1(\overline{\mathbb{R}}^d \backslash \mathbb{R}^d) = 0$ such that*

$$n\,\mathbb{P}(a_n^{-1}\mathbf{X}_0 \in \cdot\,) \xrightarrow{v} m_0(\cdot) \quad \text{and} \quad n\,\mathbb{P}(a_n^{-1}\mathbf{X}_1 \in \cdot\,) \xrightarrow{v} m_1(\cdot) \quad \text{on } \mathcal{B}(\overline{\mathbb{R}}^d \backslash \{\mathbf{0}\}).$$

Then, for every $\epsilon > 0$, $n\,\mathbb{P}(\mathbf{X} \in a_n B(2,\epsilon,[0,1])) \to 0$ as $n \to \infty$.

Proof. Fix an arbitrary $\epsilon > 0$ and let $\tau_n = \inf\{t : |\mathbf{X}_t - \mathbf{X}_0| \ge a_n\epsilon/2\}$ with the convention $\inf \emptyset = \infty$. Then

$$\begin{aligned}n\,\mathbb{P}(\mathbf{X} \in a_n B(2,\epsilon,[0,1])) &\le n\mathbb{E}(1_{\{\tau_n \le 1\}} \mathbb{E}^{\tau_n,\mathbf{X}_{\tau_n}}(1_{B(1,a_n\epsilon,[\tau_n,1])}(\mathbf{X}))) \\ &\le n\mathbb{E}(1_{\{\tau_n \le 1\}} \alpha_{a_n\epsilon/4,1}(1)/(1-\alpha_{a_n\epsilon/4,1}(1))) \\ &= n\,\mathbb{P}(\sup_{t\in[0,1]} |\mathbf{X}_t - \mathbf{X}_0| \ge a_n\epsilon/2) \frac{\alpha_{a_n\epsilon/4,1}(1)}{1-\alpha_{a_n\epsilon/4,1}(1)}\end{aligned}$$

by combining Lemma 4.16 and the strong Markov property. Moreover, by combining Lemma 2 p. 420 in Gihman and Skorohod [7] and Lemma 4.7,

$$\begin{aligned}n\,\mathbb{P}(\sup_{t\in[0,1]} |\mathbf{X}_t - \mathbf{X}_0| \ge a_n\epsilon/2) &\le \frac{n\,\mathbb{P}(|\mathbf{X}_1 - \mathbf{X}_0| \ge a_n\epsilon/4)}{1-\alpha_{a_n\epsilon/4,1}(1)} \\ &\to m_1(B^c_{\mathbf{0},\epsilon/4}) - m_0(B^c_{\mathbf{0},\epsilon/4}),\end{aligned}$$

as $n \to \infty$, from which the conclusion follows. □

Proof of Theorem 4.8. Fix $s, t \in T$ with $s < t$. We will show that there is a unique vague limit $m_{s,t}$ such that $n\,\mathbb{P}(a_n^{-1}(\mathbf{X}_s, \mathbf{X}_t) \in \cdot\ ) \xrightarrow{v} m_{s,t}(\cdot)$. By repeating the procedure one can then show that, for any $k \in \mathbb{N}$, there is a unique vague limit $m_{t_1,\ldots,t_k}$, with $m_{t_1,\ldots,t_k}(\overline{\mathbb{R}}^{dk} \backslash \mathbb{R}^{dk}) = 0$, such that $n\,\mathbb{P}(a_n^{-1}(\mathbf{X}_{t_1}, \ldots, \mathbf{X}_{t_k}) \in \cdot\ ) \xrightarrow{v} m_{t_1,\ldots,t_k}(\cdot)$ if $t_1, \ldots, t_k \in T$. By Lemma 4.7,

$$n\,\mathbb{P}(a_n^{-1}(\mathbf{X}_t - \mathbf{X}_s) \in \cdot\ ) \xrightarrow{v} m_t(\cdot) - m_s(\cdot).$$

Clearly, there are unique vague limits (Radon measures) $m_{s,s}$ and $\widetilde{m}$ on $\mathcal{B}(\overline{\mathbb{R}}^{2d} \backslash \{\mathbf{0}\})$ with $m_{s,s}(\overline{\mathbb{R}}^{2d} \backslash \mathbb{R}^{2d}) = \widetilde{m}(\overline{\mathbb{R}}^{2d} \backslash \mathbb{R}^{2d}) = 0$ such that

$$n\,\mathbb{P}(a_n^{-1}(\mathbf{X}_s, \mathbf{X}_s) \in \cdot\ ) \xrightarrow{v} m_{s,s}(\cdot) \quad \text{and} \quad n\,\mathbb{P}(a_n^{-1}(0, \mathbf{X}_t - \mathbf{X}_s) \in \cdot\ ) \xrightarrow{v} \widetilde{m}(\cdot).$$

on $\mathcal{B}(\overline{R}^{2d} \backslash \{\mathbf{0}\})$. By arguments similar to those in the proof of Lemma 4.7,

$$\begin{aligned} n\,\mathbb{P}(a_n^{-1}(\mathbf{X}_s, \mathbf{X}_t) \in \cdot\ ) &= n\,\mathbb{P}(a_n^{-1}((\mathbf{X}_s, \mathbf{X}_s) + (\mathbf{0}, \mathbf{X}_t - \mathbf{X}_s)) \in \cdot\ ) \\ &\xrightarrow{v} m_{s,s}(\cdot) + \widetilde{m}(\cdot) =: m_{s,t}(\cdot) \quad \text{on } \mathcal{B}(\overline{\mathbb{R}}^{2d} \backslash \{\mathbf{0}\}). \end{aligned}$$

Note that, by Lemma 4.17,

$$n\,\mathbb{P}(w''(\mathbf{X}, \delta) \geq a_n \epsilon) \leq n\,\mathbb{P}(\mathbf{X} \in a_n B(2, \epsilon, [0,1])) \to 0$$

as $n \to \infty$. Hence, for any positive ϵ and η there is an n_0 such that $n\,\mathbb{P}(w''(\mathbf{X}, \delta) \geq a_n \epsilon) \leq \eta$ for any $\delta \in (0,1)$ if $n \geq n_0$. Hence condition (4.7) of Theorem 4.6 holds.

It remains to show that conditions (4.8) and (4.9) also hold. Fix arbitrary $\epsilon > 0$ and $\eta > 0$. By Lemma 2 p. 420 in Gihman and Skorohod [7],

$$\begin{aligned} n\,\mathbb{P}(w(\mathbf{X}, [1-\delta, 1)) \geq a_n \epsilon) &\leq n\,\mathbb{P}(\sup_{t \in [1-\delta, 1]} |\mathbf{X}_t - \mathbf{X}_{1-\delta}| \geq a_n \epsilon / 2) \\ &\leq \frac{n\,\mathbb{P}(|\mathbf{X}_1 - \mathbf{X}_{1-\delta}| \geq a_n \epsilon / 4)}{1 - \alpha_{a_n \epsilon/4, 1}(\delta)}. \end{aligned}$$

By Lemma 4.7, for $1 - \delta \in T$

$$\lim_{n \to \infty} n\,\mathbb{P}(|\mathbf{X}_1 - \mathbf{X}_{1-\delta}| \geq a_n \epsilon / 4) = m_1(B^c_{\mathbf{0}, \epsilon/4}) - m_{1-\delta}(B^c_{\mathbf{0}, \epsilon/4}).$$

Hence, by (4.15) there exists a $\delta > 0$, $1 - \delta \in T$ such that

$$\begin{aligned} \limsup_{n \to \infty} n\,\mathbb{P}(w(\mathbf{X}, [1-\delta, 1)) \geq a_n \epsilon) &\leq \limsup_{n \to \infty} \frac{n\,\mathbb{P}(|\mathbf{X}_1 - \mathbf{X}_{1-\delta}| \geq a_n \epsilon / 4)}{1 - \alpha_{a_n \epsilon/4, 1}(\delta)} \\ &= m_1(B^c_{\mathbf{0}, \epsilon/4}) - m_{1-\delta}(B^c_{\mathbf{0}, \epsilon/4}) \leq \eta, \end{aligned}$$

and it follows that (4.9) holds. That (4.8) holds is shown by an almost identical argument. The conclusion now follows by Theorem 4.6. □

Proof of Theorem 4.12. First note that $B(2,\epsilon,[0,1])$ is open and, by Lemma 4.17, $\liminf_{n\to\infty} n\,\mathbb{P}(\mathbf{X} \in a_n B(2,\epsilon,[0,1])) = 0$. By assumption, $\liminf_{n\to\infty} n\,\mathbb{P}(\mathbf{X} \in a_n G) \geq m(G)$ for every open bounded $G \in \mathcal{B}(\overline{D})$. Hence $m(B(2,\epsilon,[0,1])) = 0$. Since $\epsilon > 0$ was arbitrary it follows that $m(B(2,\epsilon,[0,1])) = 0$ for every $\epsilon > 0$ and hence also $m(\cup_{\epsilon>0,\epsilon\in\mathbb{Q}} B(2,\epsilon,[0,1])) = 0$. Since

$$\cup_{\epsilon>0,\epsilon\in\mathbb{Q}} B(2,\epsilon,[0,1]) = (\overline{D}\backslash D) \cup \mathcal{V}^c,$$

it follows that $m(\mathcal{V}^c) \leq m(\cup_{\epsilon>0,\epsilon\in\mathbb{Q}} B(2,\epsilon,[0,1])) = 0$.

Moreover, by Theorem 4.3, there exist $\alpha > 0$ and a probability measure σ such that (4.5) holds. Furthermore, by Theorem 4.5,

$$n\,\mathbb{P}(a_n^{-1}\mathbf{X}_1 \in \cdot\,) \xrightarrow{v} m \circ \pi_1^{-1}(\cdot) \quad \text{on } \mathcal{B}(\overline{\mathbb{R}}^d\backslash\{\mathbf{0}\}),$$

which holds if and only if there exists a probability measure σ_1 on $\mathcal{B}(\mathbb{S}^{d-1})$ such that, for every $x > 0$, as $u \to \infty$,

$$\frac{\mathbb{P}(|\mathbf{X}_1| > ux, \mathbf{X}_1/|\mathbf{X}_1| \in \cdot\,)}{\mathbb{P}(|\mathbf{X}_1| > u)} \xrightarrow{w} x^{-\alpha}\sigma_1(\cdot) \quad \text{on } \mathcal{B}(\mathbb{S}^{d-1})$$

holds, and σ_1 is given by

$$\sigma_1(\cdot) = \frac{m \circ \pi_1^{-1}(\{\mathbf{x} \in \overline{\mathbb{R}}^d\backslash\{\mathbf{0}\} : |\mathbf{x}| > 1, \mathbf{x}/|\mathbf{x}| \in \cdot\,\})}{m \circ \pi_1^{-1}(\{\mathbf{x} \in \overline{\mathbb{R}}^d\backslash\{\mathbf{0}\} : |\mathbf{x}| > 1\})}.$$

We have

$$\begin{aligned}\frac{\mathbb{P}(|\mathbf{X}|_\infty > a_n, \mathbf{X}/|\mathbf{X}|_\infty \in \cdot\,)}{\mathbb{P}(|\mathbf{X}|_\infty > a_n)} &= \frac{n\,\mathbb{P}(a_n^{-1}\mathbf{X} \in \{\mathbf{x} \in D : |\mathbf{x}|_\infty > 1, \mathbf{x}/|\mathbf{x}|_\infty \in \cdot\,\})}{n\,\mathbb{P}(a_n^{-1}\mathbf{X} \in \{\mathbf{x} \in D : |\mathbf{x}|_\infty > 1\})} \\ &\xrightarrow{\hat{w}} \frac{m(\{\mathbf{x} \in D : |\mathbf{x}|_\infty > 1, \mathbf{x}/|\mathbf{x}|_\infty \in \cdot\,\})}{m(\{\mathbf{x} \in D : |\mathbf{x}|_\infty > 1\})},\end{aligned}$$

which necessarily is equal to $\sigma(\cdot)$. Moreover,

$$\begin{aligned}\sigma(\{\mathbf{x} \in D_1 : \mathbf{x} = \mathbf{y}1_{[v,1]}, v \in [0,1], \mathbf{y} \in \mathbb{S}^{d-1}\}) &= \frac{m(\{\mathbf{x} \in D : |\mathbf{x}|_\infty > 1\} \cap \mathcal{V})}{m(\{\mathbf{x} \in D : |\mathbf{x}|_\infty > 1\})} \\ &= \frac{m(\{\mathbf{x} \in D : |\mathbf{x}|_\infty > 1\})}{m(\{\mathbf{x} \in D : |\mathbf{x}|_\infty > 1\})} = 1\end{aligned}$$

and

$$\begin{aligned}
\sigma_1(\cdot) &= \frac{m(\pi_1^{-1}(\{\mathbf{x}\in\overline{\mathbb{R}}^d\backslash\{\mathbf{0}\}:|\mathbf{x}|>1,\mathbf{x}/|\mathbf{x}|\in\cdot\}))}{m(\pi_1^{-1}(\{\mathbf{x}\in\overline{\mathbb{R}}^d\backslash\{\mathbf{0}\}:|\mathbf{x}|>1\}))}\\
&= \frac{m(\pi_1^{-1}(\{\mathbf{x}\in\overline{\mathbb{R}}^d\backslash\{\mathbf{0}\}:|\mathbf{x}|>1,\mathbf{x}/|\mathbf{x}|\in\cdot\})\cap\mathcal{V})}{m(\pi_1^{-1}(\{\mathbf{x}\in\overline{\mathbb{R}}^d\backslash\{\mathbf{0}\}:|\mathbf{x}|>1\})\cap\mathcal{V})}\\
&= \frac{m(\{\mathbf{x}\in D:\mathbf{x}=\mathbf{y}1_{[v,1]},v\in[0,1],|\mathbf{y}|>1,\mathbf{y}/|\mathbf{y}|\in\cdot\})}{m(\{\mathbf{x}\in D:\mathbf{x}=\mathbf{y}1_{[v,1]},v\in[0,1],|\mathbf{y}|>1\})}\\
&= \frac{m(\{\mathbf{x}\in D:|\mathbf{x}|_\infty>1,\mathbf{x}/|\mathbf{x}|_\infty=\mathbf{y}1_{[v,1]},v\in[0,1],\mathbf{y}\in\cdot\})}{m(\{\mathbf{x}\in D:|\mathbf{x}|_\infty>1\}\cap\mathcal{V})}\\
&= \sigma(\{\mathbf{x}\in D_1:\mathbf{x}=\mathbf{y}1_{[v,1]},v\in[0,1],\mathbf{y}\in\cdot\}).
\end{aligned}$$

The conclusion follows. □

4.4 Filtered Markov processes

In this section we will give another application of regular variation on D by studying asymptotics of stochastic processes $\mathbf{Y}$ of the type

$$\mathbf{Y}_t=\int_0^t f(t,s)\mathrm{d}\mathbf{X}_s,\quad t\in[0,1], \tag{4.16}$$

where $\mathbf{X}$ is a regularly varying stochastic process with sample paths in D of finite variation. The idea here is that if $\mathbf{X}$ is a strong Markov process satisfying $\alpha_{r,1}(1)\to 0$ as $r\to 0$, then the process $\mathbf{Y}$ can be viewed as a mapping of $\mathbf{X}$. This mapping is sufficiently regular to apply Theorem 4.5. In this way we can show that the process $\mathbf{Y}$ is regularly varying. Furthermore, and equally important, we are able to explicitly compute the spectral measure of such processes. In doing so we provide a natural way to understanding the extremal behavior of such filtered regularly varying Markov processes. As a concrete example we will compute the spectral measure of an Ornstein-Uhlenbeck type process driven by a regularly varying Lévy process (Example 4.19). Note that finite variation of the sample paths of $\mathbf{X}$ allows us to define the integral in (4.16) in a pathwise sense. As in the previous section, let

$$\mathcal{V}=\{\mathbf{x}\in D:\mathbf{x}=\mathbf{y}1_{[v,1]},v\in[0,1],\mathbf{y}\in\mathbb{R}^d\backslash\{\mathbf{0}\}\}.$$

For a continuous $f:[0,1]^2\to\mathbb{R}$ define $h_f:D\to D$ by

$$h_f(\mathbf{x})(t)=\begin{cases}\int_0^t f(t,s)\mathrm{d}\mathbf{x}_s & \text{if } \mathbf{x} \text{ has finite variation,}\\ 0 & \text{otherwise.}\end{cases}$$

Note that h_f is in general not continuous, not even on the finite variation part of D. However, it is continuous on $\mathcal{V}$, and this is sufficient when considering integrators whose regular variation limit measure concentrates on $\mathcal{V}$.

Theorem 4.18. *Let $\mathbf{X} = \{\mathbf{X}_t : t \in [0,1]\}$ be a strong Markov process with sample paths in D of finite variation such that $\alpha_{r,1}(1) \to 0$ as $r \to 0$. Suppose that there exist a sequence $\{a_n\}$, $0 < a_n \uparrow \infty$, and a nonzero boundedly finite measure m on $\mathcal{B}(\overline{D})$ with $m(\overline{D}\backslash D) = 0$ such that, as $n \to \infty$,*

$$n\,\mathbb{P}(a_n^{-1}\mathbf{X} \in \cdot\,) \xrightarrow{\hat{w}} m(\cdot) \quad \text{on } \mathcal{B}(\overline{D}).$$

For $f : [0,1]^2 \mapsto \mathbb{R}$ continuous, define $\mathbf{Y} = \{\mathbf{Y}_t : t \in [0,1]\}$ by $\mathbf{Y}_t = \int_0^t f(t,s)\mathrm{d}\mathbf{X}_s$. Then $\mathbf{Y}$ has sample paths in D and, as $n \to \infty$,

$$n\,\mathbb{P}(a_n^{-1}\mathbf{Y} \in \cdot\,) \xrightarrow{\hat{w}} \widetilde{m}(\cdot) \quad \text{on } \mathcal{B}(\overline{D}),$$

where $\widetilde{m}(\cdot) = m \circ h_f^{-1}(\cdot \cap D)$ is a nonzero boundedly finite measure on $\mathcal{B}(\overline{D})$. Moreover, $\widetilde{m}(h_f(\mathcal{V})^c) = 0$.

To illustrate Theorem 4.18 we will now compute the the spectral measure of an Ornstein-Uhlenbeck type process driven by a regularly varying Lévy process.

Example 4.19. Let $\mathbf{X} = \{\mathbf{X}_t : t \in [0,1]\}$ be a Lévy process on $\mathbb{R}^d$ with sample paths of finite variation. Necessary and sufficient conditions for having sample paths of finite variation are that the generating triplet (A, ν, γ) satisfies $A = 0$ and either (i) $\nu(\mathbb{R}^d\backslash\{\mathbf{0}\}) < \infty$ or (ii) $\nu(\mathbb{R}^d\backslash\{\mathbf{0}\}) = \infty$ and $\int_{|\mathbf{x}|\le 1, \mathbf{x}\neq\mathbf{0}} \nu(\mathrm{d}\mathbf{x}) < \infty$ (see Sato [14]p. 140). Suppose there exist a sequence $\{a_n\}$, $0 < a_n \uparrow \infty$, and a nonzero Radon measure m_1 with $m_1(\overline{\mathbb{R}}^d\backslash\mathbb{R}^d) = 0$ such that

$$n\,\mathbb{P}(a_n^{-1}\mathbf{X}_1 \in \cdot\,) \xrightarrow{v} m_1(\cdot) \quad \text{on } \mathcal{B}(\overline{\mathbb{R}}^d\backslash\{\mathbf{0}\}).$$

Then, since $\mathbf{X}$ has stationary and independent increments, $n\,\mathbb{P}(a_n^{-1}\mathbf{X}_t \in \cdot\,) \xrightarrow{v} tm_1(\cdot)$ on $\mathcal{B}(\overline{\mathbb{R}}^d\backslash\{\mathbf{0}\})$ for every $t \in [0,1]$. Let $\mathbf{Y} = \{\mathbf{Y}_t : t \in [0,1]\}$ be an Ornstein-Uhlenbeck type process driven by $\mathbf{X}$, given by,

$$\mathbf{Y}_t = \int_0^t e^{-\theta(t-s)}\mathrm{d}\mathbf{X}_s, \quad \theta > 0, \quad t \in [0,1].$$

Hence, by combining Theorems 4.3, 4.8 and 4.18, for every $x > 0$, as $u \to \infty$,

$$\frac{\mathbb{P}(|\mathbf{Y}|_\infty > ux, \mathbf{Y}/|\mathbf{Y}|_\infty \in \cdot\,)}{\mathbb{P}(|\mathbf{Y}|_\infty > u)} \xrightarrow{w} x^{-\alpha}\sigma(\cdot) \quad \text{on } \mathcal{B}(D_1),$$

where $\alpha > 0$ is the tail index of $\mathbf{X}_1$ and it follows that

$$\sigma(\cdot) = \mathbb{P}(\{\mathbf{Z}e^{-\theta(t-V)}1_{[V,1]}(t), t \in [0,1]\} \in \cdot\,),$$

where $\mathbf{Z}$ and V are independent, the distribution of $\mathbf{Z}$ is the spectral measure of $\mathbf{X}_1$ and V is uniformly distributed on $[0,1]$.

Proof of Theorem 4.18. We first show that $\mathbf{Y}$ has sample paths in D. By assumption there is $\Omega' \subset \Omega$ with $\mathbb{P}(\Omega') = 1$ such that for each $\omega \in \Omega'$, $\mathbf{X}(\omega) \in D$ and has finite variation. For such ω we also have, since f is continuous on $[0,1]^2$ and hence also uniformly continuous on $[0,1]^2$,

$$\lim_{v \uparrow t} \Big| \int_{[0,v]} (f(t,s) - f(v,s)) \mathrm{d}\mathbf{X}_s(\omega) \Big| \leq \lim_{v \uparrow t} \sup_{s \in [0,t]} |f(t,s) - f(v,s)| \mathrm{FV}(\mathbf{X}(\omega); [0,t]) = 0,$$

where $\mathrm{FV}(g; T)$ denotes the total variation of g on $T \subset [0,1]$. Hence, for $\omega \in \Omega'$,

$$\begin{aligned}
\lim_{v \uparrow t} (\mathbf{Y}_t(\omega) - \mathbf{Y}_v(\omega)) &= \lim_{v \uparrow t} \Big(\int_{[0,t]} f(t,s) \mathrm{d}\mathbf{X}_s(\omega) - \int_{[0,v]} f(v,s) \mathrm{d}\mathbf{X}_s(\omega) \Big) \\
&= \lim_{v \uparrow t} \int_{[0,v]} (f(t,s) - f(v,s)) \mathrm{d}\mathbf{X}_s(\omega) + \lim_{v \uparrow t} \int_{(v,t]} f(t,s) \mathrm{d}\mathbf{X}_s(\omega) \\
&= 0 + f(t,t)(\mathbf{X}_t(\omega) - \mathbf{X}_{t-}(\omega)),
\end{aligned}$$

since $\mathbf{X}(\omega)$ is right-continuous with left limits. Similarly,

$$\begin{aligned}
\lim_{v \downarrow t} (\mathbf{Y}_v(\omega) - \mathbf{Y}_t(\omega)) &= \lim_{v \downarrow t} \Big(\int_{[0,v]} f(v,s) \mathrm{d}\mathbf{X}_s(\omega) - \int_{[0,t]} f(t,s) \mathrm{d}\mathbf{X}_s(\omega) \Big) \\
&= \lim_{v \downarrow t} \int_{[0,t]} (f(v,s) - f(t,s)) \mathrm{d}\mathbf{X}_s(\omega) + \lim_{v \downarrow t} \int_{(t,v]} f(v,s) \mathrm{d}\mathbf{X}_s(\omega) \\
&= 0 + f(t,t)(\mathbf{X}_{t+}(\omega) - \mathbf{X}_t(\omega)) = 0.
\end{aligned}$$

Hence $\mathbf{Y}(\omega)$ is right-continuous with left limits. Since m vanishes on $\mathcal{V}^c$ it is sufficient to show that h_f is continuous on $\mathcal{V}$. Moreover, it is sufficient to show that h_f is continuous on $\mathcal{V} \subset D$ equipped with the Skorohod metric since this metric and the J_1-metric are equivalent (see Billingsley [2] p. 114). Let $\mathbf{x}_n, \mathbf{x} \in \mathcal{V}$. Then there is $\mathbf{y}, \mathbf{y}_n, v, v_n$ such that $\mathbf{x}_n = \mathbf{y}_n 1_{[v_n,1]}$ and $\mathbf{x} = \mathbf{y} 1_{[v,1]}$. Suppose that $\mathbf{x}_n \to \mathbf{x}$. Then there exists a sequence λ_n of strictly increasing continuous mappings of $[0,1]$ onto itself satisfying $\sup_{t \in [0,1]} |\lambda_n(t) - t| \to 0$ and

$$\sup_{t \in [0,1]} |\mathbf{y}_n 1_{[v_n,1]}(\lambda_n(t)) - \mathbf{y} 1_{[v,1]}(t)| \to 0 \quad \text{as } n \to \infty.$$

First we show that $\mathbf{x}_n \to \mathbf{x}$ implies that $\mathbf{y}_n \to \mathbf{y}$ and $v_n \to v$. Since

$$\sup_{t \in [0,1]} |\mathbf{y}_n 1_{[v_n,1]}(\lambda_n(t)) - \mathbf{y} 1_{[v,1]}(t)| \geq |\mathbf{y}_n - \mathbf{y}|,$$

it follows that $\mathbf{y}_n \to \mathbf{y}$. Suppose that $v_n \not\to v$. Then there exists $\epsilon > 0$ such that $\limsup_{n \to \infty} |v_n - v| > \epsilon$. Since $\sup_{t \in [0,1]} |\lambda_n(t) - t| \to 0$ and $\mathbf{y}_n \to \mathbf{y}$, this implies that

$$\limsup_{n \to \infty} \sup_{t \in [0,1]} |\mathbf{y}_n 1_{[v_n,1]}(\lambda_n(t)) - \mathbf{y} 1_{[v,1]}(t)| \geq |\mathbf{y}|/2,$$

which is a contradiction. Hence $v_n \to v$. We may now proceed to show that $\mathbf{x}_n \to \mathbf{x}$ implies $h_f(\mathbf{x}_n) \to h_f(\mathbf{x})$. Indeed

$$\begin{aligned}
&\sup_{t\in[0,1]} \Big| \int_0^{\lambda_n(t)} f(\lambda_n(t), s)\mathrm{d}\mathbf{x}_n(s) - \int_0^t f(t,s)\mathrm{d}\mathbf{x}(s)\Big| \\
&\quad \le \sup_{t\in[0,1]} |f(\lambda_n(t), v_n)(\mathbf{y}_n 1_{[v_n,1]}(\lambda_n(t)) - \mathbf{y}1_{[v,1]}(t))| \\
&\qquad + \sup_{t\in[0,1]} |(f(\lambda_n(t), v_n) - f(t,v))\mathbf{y}1_{[v,1]}(t)| \\
&\quad \le \sup_{t\in[0,1]} |f(\lambda_n(t), v_n)| \sup_{t\in[0,1]} |\mathbf{y}_n 1_{[v_n,1]}(\lambda_n(t)) - \mathbf{y}1_{[v,1]}(t)| \\
&\qquad + |\mathbf{y}| \sup_{t\in[0,1]} |f(\lambda_n(t), v_n) - f(t,v)|.
\end{aligned}$$

Since f is bounded and $\mathbf{x}_n \to \mathbf{x}$,

$$\sup_{t\in[0,1]} |f(\lambda_n(t), v_n)| \sup_{t\in[0,1]} |\mathbf{y}_n 1_{[v_n,1]}(\lambda_n(t)) - \mathbf{y}1_{[v,1]}(t)| \to 0.$$

Since f is uniformly continuous on $[0,1]^2$, $\sup_{t\in[0,1]} |\lambda_n(t) - t| \to 0$ and $v_n \to v$ it follows that

$$\sup_{t\in[0,1]} |f(\lambda_n(t), v_n) - f(t,v)| \to 0.$$

Hence h_f is continuous on $\mathcal{V}$. Note that $n\,\mathbb{P}(a_n^{-1}\mathbf{Y} \in B) = n\,\mathbb{P}(a_n^{-1}h_f(\mathbf{X}) \in B)$ for every bounded $B \in \mathcal{B}(\overline{D})$ and, by Theorem 4.5,

$$n\,\mathbb{P}(a_n^{-1}h_f(\mathbf{X}) \in \cdot\,) \xrightarrow{\hat{w}} m \circ h_f^{-1}(\cdot \cap D) \quad \text{on } \mathcal{B}(\overline{D}).$$

Hence $n\,\mathbb{P}(a_n^{-1}\mathbf{Y} \in \cdot\,) \xrightarrow{\hat{w}} m \circ h_f^{-1}(\,\cdot \cap D)$ on $\mathcal{B}(\overline{D})$. Note that

$$h_f^{-1}(h_f(\mathcal{V})^c \cap D) = h_f^{-1}(h_f(\mathcal{V}))^c \cap D \subset \mathcal{V}^c \cap D \subset \mathcal{V}^c$$

and, by Theorem 4.12, $m(\mathcal{V}^c) = 0$. Hence $\widetilde{m}(h_f(\mathcal{V})^c) = m(h_f^{-1}(h_f(\mathcal{V})^c \cap D)) = 0$. □

References

[1] Basrak, B. (2000) *The Sample Autocorrelation Function of Non-Linear Time Series*, Ph.D. Thesis, University of Groningen, Department of Mathematics.

[2] Billingsley, P. (1968) *Convergence of Probability Measures*, first edition, Wiley, New York.

[3] Billingsley, P. (1995) *Probability and Measure*, 3rd edition, Wiley, New York.

[4] Bingham, N.H., Goldie, C.M. and Teugels, J.L. (1987) *Regular variation*, Encyclopedia of mathematics and its applications 27, Cambridge University Press, Cambridge.

[5] Braverman, M., Mikosch, T. and Samorodnitsky, G. (2002) Tail probabilities of subadditive functionals acting on Lévy processes, *Ann. Appl. Probab.*, Vol. 12, 69–100.

[6] Daley, D.J. and Vere-Jones, D. (1988) *An Introduction to the Theory of Point Processes*, Springer Verlag, New York.

[9] de Haan, L. and Lin, T. (2002) On convergence toward an extreme value limit in $C[0,1]$, *Ann. Probab.*, Vol. 29, 467–483.

[10] Hult, H. and Lindskog, F. (2002) Multivariate regular variation for additive processes, submitted.

[7] Gihman, I.I. and Skorohod, A.V. (1974) *The Theory of Stochastic Processes I*, Springer Verlag, Berlin.

[8] Giné, E., Hahn, M.G. and Vatan, P. (1990) Max-infinitely divisible and max-stable sample continuous processes, *Probab. Theory Related Fields*, Vol. 87, 139–165.

[11] Kallenberg, O. (1983) *Random Measures*, 3rd edition, Akademie Verlag, Berlin.

[12] Resnick, S.I. (1987) *Extreme Values, Regular Variation, and Point Processes*, Springer Verlag, New York.

[13] Rosiński, J. and Samorodnitsky, G. (1993) Distributions of subadditive functionals of sample paths of infinitely divisible processes, *Ann. Probab.*, Vol. 21, 996–1014.

[14] Sato, K.-I. (1999) *Lévy Processes and Infinitely Divisible Distributions*, Cambridge University Press, Cambridge.

Appendix: Results on regular variation

Definition 4.20. A stochastic process $\mathbf{X} = \{\mathbf{X}_t : t \in [0,1]\}$ with sample paths in D is said to be regularly varying if there exist a sequence $\{a_n\}$, $0 < a_n \uparrow \infty$ and a nonzero boundedly finite measure m on $\mathcal{B}(\overline{D})$ with $m(\overline{D}\backslash D) = 0$, such that, as $n \to \infty$,

$$n\,\mathbb{P}(a_n^{-1}\mathbf{X} \in \cdot\,) \xrightarrow{\hat{w}} m(\cdot) \quad \text{on } \mathcal{B}(\overline{D}). \tag{4.17}$$

The limit measure m has a scaling property; there exists an $\alpha > 0$ such that $m(uB) = u^{-\alpha} m(B)$ for every $u > 0$ and $B \in \mathcal{B}(\overline{D})$. This follows from a combination of standard regular variation arguments. For $r > 0$ and $S \in \mathcal{B}(D_1)$, let $V_{r,S} = \{\mathbf{x} \in \overline{D} \cap D : |\mathbf{x}|_\infty > r, \mathbf{x}/|\mathbf{x}|_\infty \in S\}$. Fix $S \in \mathcal{B}(D_1)$ such that for some $r > 0$ we have $m(\partial V_{r,S}) = 0$. Since m is boundedly finite such a set $S \in \mathcal{B}(D_1)$ exists. Then $m(\partial u V_{r,S}) = 0$ for all but at most countably many $u \in [1, \infty)$. Denote this set by U, i.e. $U = \{u \in [1, \infty) : m(\partial u V_{r,S}) = 0\}$.

Suppose that $m(V_{r,S}) > 0$. Define f and g on $(0, \infty)$ by $f(x) = \mathbb{P}(\mathbf{X} \in x V_{r,S})$ and $g(x) = m(x V_{r,S})$. Then

$$n f(u a_n) = n\, \mathbb{P}(\mathbf{X} \in a_n u V_{r,S}) \to m(u V_{r,S}) = g(u), \quad \text{as } n \to \infty, \quad u \in U.$$

For $x \geq a_1$, let $t = t(x)$ be the largest integer with $a_t \leq x$. Since f is nonincreasing, i.e. $x \leq y$ implies $f(x) \geq f(y)$,

$$f(u a_{t+1})/f(a_t) \leq f(ux)/f(x) \leq f(u a_t)/f(a_{t+1}).$$

However, the lower bound is $(t/(t+1))(t+1) f(u a_{t+1})/(t f(a_t))$ which tends to $g(u)/g(1)$, for each $u \in U$. Similarly for the upper bound. Hence

$$f(ux)/f(x) \to g(u)/g(1), \quad \text{as } x \to \infty, \quad u \in U.$$

For arbitrary $u > 0$, let $g^*(u) = \limsup_{x \to \infty} f(ux)/f(x)$. Then, $g^*(u) = g(u)/g(1)$ for $u \in U$. Moreover, g^* is nonincreasing since for any $x > 0$, $f(ux)/f(x)$ is nonincreasing in u. Hence, for $u > 1$,

$$g^*(u) \leq \inf_{v \in U \cap [1,u)} g(v)/g(1).$$

However, since $m(\partial V_{r,S}) = 0$, g is continuous at $u = 1$ and hence

$$\limsup_{u \downarrow 1} g^*(u) \leq \limsup_{u \downarrow 1} \inf_{v \in U \cap [1,u)} g(v)/g(1) = 1.$$

It now follows by Theorem 1.4.3 (ii) p. 18 in Bingham, Goldie and Teugels [4] that there exists $\alpha \in \mathbb{R}$ such that

$$f(ux)/f(x) \to u^{-\alpha}, \quad \text{as } x \to \infty, \quad \text{for each } u > 0,$$

i.e. $m(u V_{r,S}) = u^{-\alpha} m(V_{r,S})$ for each $u > 0$.

Suppose that $m(V_{r,S}) = 0$. We will show that this implies that $m(V_{u,S}) = 0$ for every $u > 0$. Suppose that there exists $r_0 \in (0, r)$ such that $m(V_{r_0,S}) > 0$. Then, by the above arguments, there exists $r_1 \in (0, r)$ such that $m(V_{r_1,S}) > 0$ and $m(\partial V_{r_1,S}) = 0$. However, by the above arguments, we must then have

$$m(u V_{r_1,S}) = u^{-\alpha} m(V_{r_1,S}) \quad \text{for every } u > 0.$$

In particular, $0 = m(V_{r,S}) = (r/r_1)^{-\alpha} m(V_{r_1,S}) > 0$, which is a contradiction.

Hence, for each m-continuity set $V_{r,S}$ there is an $\alpha \in \mathbb{R}$ such that

$$m(uV_{r,S}) = u^{-\alpha} m(V_{r,S}) \quad \text{for every } u > 0.$$

It remains to show that α does not depend on $V_{r,S}$. This can be shown by the same arguments as in Resnick [12] p. 277. Since the m-continuity sets of the form $V_{r,S}$ for $r > 0$ and $S \in \mathcal{B}(D_1)$ form a π-system which generates $\mathcal{B}(\overline{D}) \cap D$ the scaling property holds for every $B \in \mathcal{B}(\overline{D}) \cap D$. However, since $m(\overline{D} \backslash D) = 0$ the scaling property holds for every $B \in \mathcal{B}(\overline{D})$. Since m is a boundedly finite measure we must have $\alpha \geq 0$ and since $m(\overline{D} \backslash D) = 0$ we must have $\alpha > 0$.

Theorem 4.21. *A stochastic process $\mathbf{X} = \{\mathbf{X}_t : t \in [0,1]\}$ with sample paths in D is regularly varying if and only if there exist an $\alpha > 0$ and a probability measure σ on $\mathcal{B}(D_1)$ such that, for every $x > 0$, as $u \to \infty$,*

$$\frac{\mathbb{P}(|\mathbf{X}|_\infty > ux, \mathbf{X}/|\mathbf{X}|_\infty \in \cdot\,)}{\mathbb{P}(|\mathbf{X}|_\infty > u)} \xrightarrow{w} x^{-\alpha}\sigma(\cdot) \quad \text{on } \mathcal{B}(D_1). \tag{4.18}$$

The probability measure σ is referred to as the spectral measure and α is referred to as the tail index.

Proof. Suppose that (4.18) holds. For $x > 0$ and $S \in \mathcal{B}(D_1)$, let

$$\begin{aligned} V_{x,S} &= \{\mathbf{x} \in \overline{D} \cap D : |\mathbf{x}|_\infty > x, \mathbf{x}/|\mathbf{x}|_\infty \in S\}, \\ \mathcal{P} &= \{V_{x,S} : x > 0, S \in \mathcal{B}(D_1)\}. \end{aligned}$$

For $u, x > 0$ define the measures

$$\begin{aligned} m_u(V_{x,\cdot}) &= \mathbb{P}(|\mathbf{X}|_\infty > ux, \mathbf{X}/|\mathbf{X}|_\infty \in \cdot\,)/\,\mathbb{P}(|\mathbf{X}|_\infty > u), \\ m(V_{x,\cdot}) &= x^{-\alpha}\sigma(\cdot) \end{aligned}$$

on $\mathcal{B}(D_1)$. This also defines, uniquely, set functions m_u and m on the semiring $\mathcal{P}$. By Theorem 11.3 p. 166 in Billingsley [3], m_u and m can be uniquely extended to measures on $\sigma(\mathcal{P}) = \mathcal{B}(\overline{D}) \cap D$. By requiring that $m_u(\overline{D} \backslash D) = m(\overline{D} \backslash D) = 0$, m_u and m can be uniquely extended to (boundedly finite) measures on $\mathcal{B}(\overline{D})$. By definition of m, $m(V_{x,S}) = x^{-\alpha} m(V_{1,S})$ for $x > 0$ and $S \in \mathcal{B}(D_1)$. Suppose that there exist $x, c > 0$ such that $m(\partial V_{x,D_1}) = c$. Then, for $y > x$,

$$\begin{aligned} m(V_{x,D_1} \backslash V_{y,D_1}) &\geq m(\cup_{q \in \mathbb{Q} \cap (x,y]} \partial V_{q,D_1}) = \sum_{q \in \mathbb{Q} \cap (x,y]} m(\partial V_{q,D_1}) \\ &\geq cy^{-\alpha} \sum_{q \in \mathbb{Q} \cap (x,y]} 1 = \infty. \end{aligned}$$

Since $V_{x,D_1} \backslash V_{y,D_1}$ is bounded in $\overline{D}$ this is a contradiction and we conclude that $m(\partial V_{x,D_1}) = 0$ for every $x > 0$. This implies in particular that $m(\partial V_{x,S}) =$

$m(V_{x,\partial S})$ for every $x > 0$ and $S \in \mathcal{B}(D_1)$. Since $m_u(V_{x,\cdot}) \xrightarrow{\hat{w}} m(V_{x,\cdot})$ for every $x > 0$ this implies $m_u(V_{x,S}) \to m(V_{x,S})$ as $u \to \infty$ for every m-continuity set $V_{x,S}$. Since $m_u(\overline{D}\backslash D) = m(\overline{D}\backslash D) = 0$, the m-continuity sets of $\mathcal{P}$ form a convergence determining class, i.e. $m_u \xrightarrow{\hat{w}} m$ on $\mathcal{B}(\overline{D})$. Finally, choose a sequence $\{a_n\}$ such that $n\,\mathbb{P}(|\mathbf{X}|_\infty > a_n) \to 1$. Then

$$\frac{n\,\mathbb{P}(a_n^{-1}\mathbf{X} \in \cdot\,)}{n\,\mathbb{P}(|\mathbf{X}|_\infty > a_n)} = m_{a_n}(\cdot) \xrightarrow{\hat{w}} m(\cdot) \quad \text{on } \mathcal{B}(\overline{D}),$$

i.e. $n\,\mathbb{P}(a_n^{-1}\mathbf{X} \in \cdot\,) \xrightarrow{\hat{w}} m(\cdot)$ on $\mathcal{B}(\overline{D})$.

Suppose that (4.17) holds. Then, for $x > 0$ and $S \in \mathcal{B}(D_1)$,

$$n\,\mathbb{P}(|\mathbf{X}|_\infty > a_n x, \mathbf{X}/|\mathbf{X}|_\infty \in S) = n\,\mathbb{P}(a_n^{-1}\mathbf{X} \in V_{x,S}) \to m(V_{x,S}) = x^{-\alpha} m(V_{1,S}),$$

if $m(\partial V_{x,S}) = 0$. Since m has the scaling property $m(xB) = x^{-\alpha} m(B)$ it follows by the same arguments as above that $m(\partial V_{x,S}) = m(V_{x,\partial S})$ for every $x > 0$ and $S \in \mathcal{B}(D_1)$. Hence, with $\widehat{a}_n = (m(V_{1,D_1}))^{1/\alpha} a_n$,

$$n\,\mathbb{P}(|\mathbf{X}|_\infty > \widehat{a}_n x, \mathbf{X}/|\mathbf{X}|_\infty \in \cdot\,) \xrightarrow{w} x^{-\alpha}\sigma(\cdot) \quad \text{on } \mathcal{B}(D_1),$$

where $\sigma(\cdot) = m(V_{1,\cdot})/m(V_{1,D_1})$ is a probability measure on $\mathcal{B}(D_1)$. For $u \geq a_1$, let $n = n(u)$ be the largest integer with $\widehat{a}_n \leq u$. For any $x > 0$ and $S \in \mathcal{B}(D_1)$ with $\sigma(\partial S) = 0$ we have

$$\begin{aligned}
\frac{n\,\mathbb{P}(|\mathbf{X}|_\infty > \widehat{a}_{n+1} x, \mathbf{X}/|\mathbf{X}|_\infty \in S)}{n\,\mathbb{P}(|\mathbf{X}|_\infty > \widehat{a}_n)} &\leq \frac{\mathbb{P}(|\mathbf{X}|_\infty > ux, \mathbf{X}/|\mathbf{X}|_\infty \in S)}{\mathbb{P}(|\mathbf{X}|_\infty > u)} \\
&\leq \frac{n\,\mathbb{P}(|\mathbf{X}|_\infty > \widehat{a}_n x, \mathbf{X}/|\mathbf{X}|_\infty \in S)}{n\,\mathbb{P}(|\mathbf{X}|_\infty > \widehat{a}_{n+1})}.
\end{aligned}$$

However,

$$\begin{aligned}
\frac{n\,\mathbb{P}(|\mathbf{X}|_\infty > \widehat{a}_{n+1} x, \mathbf{X}/|\mathbf{X}|_\infty \in S)}{n\,\mathbb{P}(|\mathbf{X}|_\infty > \widehat{a}_n)} &= \frac{n}{n+1} \frac{(n+1)\,\mathbb{P}(\widehat{a}_{n+1}^{-1}\mathbf{X} \in V_{x,S})}{n\,\mathbb{P}(\widehat{a}_n^{-1}\mathbf{X} \in V_{1,D_1})} \\
&\to \frac{x^{-\alpha}\sigma(S)}{\sigma(D_1)} = x^{-\alpha}\sigma(S)
\end{aligned}$$

as $u \to \infty$. Similarly for the upper bound. Hence (4.18) holds.

□

Part II

Fractional Brownian motion and parameter estimation

Chapter 5

Approximating some Volterra type stochastic integrals with applications to parameter estimation

Henrik Hult (2002) Approximating some Volterra type stochastic integrals with applications to parameter estimation, *Stochastic Processes and their Applications*, Vol. 105 No. 1, 1–32.

Abstract We consider Volterra type processes which are Gaussian processes admitting representation as a Volterra type stochastic integral with respect to the standard Brownian motion, for instance the fractional Brownian motion. Gaussian processes can be represented as a limit of a sequence of processes in the associated reproducing kernel Hilbert space and as a special case of this representation we derive Karhunen-Loéve expansions for Volterra type processes. In particular a wavelet decomposition for the fractional Brownian motion is obtained. We also consider a Skorohod type stochastic integral with respect to a Volterra type process and using the Karhunen-Loéve expansions we show how it can be approximated. Finally we apply the results to estimation of drift parameters in stochastic models driven by Volterra type processes using a Girsanov transformation and we prove consistency, the rate of convergence and asymptotic normality of the derived maximum likelihood estimators.

2000 *Mathematics Subject Classification.* 60H07, 28C20 (primary); 60H10, 60H20 (secondary).
Keywords and phrases. Fractional Brownian motion; Reproducing kernel Hilbert space; Gaussian process; Likelihood function.
Acknowledgments. The author want to thank Boualem Djehiche for suggesting the topic of this paper and for his comments on the manuscript. I am grateful for the careful reading and comments by Svante Janson and an anonymous referee.

5.1 Introduction

Gaussian processes admitting representation as a Volterra type stochastic integral with respect to the standard Brownian motion such as the fractional Brownian motion are used in several fields including telecommunication networks, subsurface hydrology and mathematical finance. An important problem in such areas is to estimate the parameters within a given family of models, for example in a stochastic differential equation driven by the fractional Brownian motion. In this paper we will consider the fractional Brownian motion and related processes. Since the fractional Brownian motion is quite complicated, one generally tries to find an approximation which is easier to handle. Different approximations have been studied in the literature. In Comte [6], and Comte and Renault [7], Gaussian processes of the form,

$$X_t = \int_0^t a(t-s)\mathrm{d}B_s, \tag{5.1}$$

where $\{B_t : t \geq 0\}$ is standard Brownian motion are studied. In particular the authors show how to estimate parameters in some special cases when $a(\cdot)$ depend on an unknown parameter. A natural tool when studying approximations is to use wavelets. We can represent the process in a wavelet basis, $\{\psi_{j,k}\}$, as a Mercer-type representation,

$$X_t = \sum_{j,k} \psi_{j,k}(t)\xi_{j,k}, \tag{5.2}$$

where the wavelet coefficients $\{\xi_{j,k}\}$ is a sequence of Gaussian variables (see Abry, Flandrin, Taqqu and Veitch [1] and the references therein). The correlations between the wavelet coefficients depend on the number of vanishing moments of the wavelet used. Ideal would be to have independent coefficients. In Meyer, Sellan and Taqqu [21], the authors obtain a wavelet decomposition for the fractional Brownian motion using a generalization of the midpoint displacement technique. They show that it has the representation

$$B^H(t) = \sum_k S_k \Phi_k^H(t) + \sum_{j,k} \xi_{j,k}\Psi_{j,k}^H(t), \tag{5.3}$$

where convergence holds almost surely uniformly on compact intervals, $\{S_k\}$ is a fractional ARIMA process and $\{\xi_{j,k}\}$ is a sequence of iid $N(0,1)$ random variables and independent of $\{S_k\}$. The sequence $\{\Phi_k^H, \Psi_{j,k}^H\}$ is constructed from a suitable wavelet basis, for instance the Meyer wavelets. As we shall see below, in Section 5.3.1, the representation (5.3) can be derived as a special case of a more general representation. Let us also mention the work by Carmona, Coutin and Montseny [5], Decreusefond and Üstünel [9] and Norros, Mannersalo and Wang [22], on further approximations and problems concerning simulation of the fractional Brownian motion.

In this paper we study centered Gaussian processes $X = \{X_t : t \in T\}$, where T is an index set, admitting a representation of the form:

$$X_t = \int_T V(t,r)\mathrm{d}B_r, \quad t \in T, \tag{5.4}$$

where $\{B_t : t \in T\}$ denotes standard Brownian motion and V is a kernel satisfying some conditions (see Section 5.2). The conditions are such that X has a continuous modification and includes the fractional Brownian motion. We will throughout the paper use the representation of Gaussian processes

$$X_t = \sum_j \Psi_j(t)\xi_j, \quad t \in T, \tag{5.5}$$

where $\{\xi_j\}$ is a sequence of iid $N(0,1)$ random variables and $\{\Psi_j\}$ is an orthonormal basis in the reproducing kernel Hilbert space associated with X. If T is compact and X is a.s. continuous then the sum converges a.s. uniformly. To get an explicit representation (5.5) we need to find an orthonormal basis in the reproducing kernel Hilbert space. For processes of the form (5.4) this is not difficult since, as we will see, the reproducing kernel Hilbert space is the image of $L^2(T)$ under the integral transform

$$(Vf)(t) = \int_T V(t,s)f(s)\mathrm{d}s, \quad f \in L^2(T),$$

equipped with the inner product,

$$\langle Vf, Vg\rangle_{V(L^2(T))} = \langle f, g\rangle_{L^2(T)}.$$

An orthonormal basis for the reproducing kernel Hilbert space is then obtained from an orthonormal basis in $L^2(T)$ by applying the integral transform V on each basis function. That is, if $\{\psi_j\}$ is an orthonormal basis in $L^2(T)$ then $\{\Psi_j\}$, where $\Psi_j = V\psi_j$, is an orthonormal basis in $V(L^2(T))$. It should be noted that the representation (5.5) can also be useful for simulations.

Furthermore, we introduce a Skorohod type stochastic integral with respect to a Volterra type process by extending the stochastic integral for fractional Brownian motion introduced by Decreusefond and Üstünel in [9]. The integral is defined as $\delta^X(Vu)$ where $u = \{u(t,\omega) : t \in T, \omega \in \Omega\}$ satisfies suitable conditions and δ^X is the divergence operator associated with the Gaussian process X. Using the expansion (5.5) we find an approximation for this stochastic integral.

Using a Girsanov transformation we derive a maximum likelihood estimator of the drift parameter θ in models of the type

$$Y(t,\omega) = \theta A(t,\omega) + X(t,\omega),$$

where $A(\cdot)$ is a sufficiently smooth drift function and $\{X_t : t \in T\}$ is a Volterra type process. We prove strong consistency, the rate of convergence and asymptotic

normality of the estimator. Our results extends previous work by Norros, Valkeila and Virtamo [23] who studied the case where $A(t) = t$ for the fractional Brownian motion. We can also derive an estimator for the mean reverting parameter in an Ornstein-Uhlenbeck type process

$$Y(t) = \theta \int_0^t Y_s \mathrm{d}s + X(t).$$

This problem is also studied in Kleptsyna and Le Breton [15] for the fractional Brownian motion. There the authors also obtain the asymptotic bias and the asymptotic mean square error of the estimator using Laplace transforms. We show how their estimator can be obtained via the approach taken in this paper.

The paper is organized as follows. In Section 5.2 we give some preliminaries on fractional calculus, define Volterra type processes and introduce the reproducing kernel Hilbert space associated with a Gaussian process. In Section 5.3 we state the general representation (5.5) for Volterra type processes (processes of the form (5.4)) and obtain as special cases, the Karhunen-Loéve expansion for the standard Brownian motion and a wavelet representation (5.3) for the fractional Brownian motion. In Section 5.4 we introduce a Skorohod type stochastic integral with respect to a Volterra type process. We also give approximations of these integrals for deterministic as well as stochastic integrands. In Section 5.5 we apply the results to parameter estimation using a Girsanov transformation. As examples we consider the estimation of a fractional Brownian motion with linear drift and estimation of the mean-reverting parameter in an Ornstein-Uhlenbeck type process driven by a fractional Brownian motion. We also prove consistency, the law of the iterated logarithm and asymptotic normality of the derived estimators.

5.2 Preliminaries and definitions

We begin by reviewing some results on fractional calculus to be used later in the paper. The main reference on fractional calculus is the book by Samko, Kilbas and Marichev [26]. We will use the Liouville fractional integral.

Definition 5.1. For $f \in L^1([a,b])$ and $\alpha > 0$, the integrals

$$(I_{a^+}^{\alpha} f)(t) \triangleq \frac{1}{\Gamma(\alpha)} \int_a^t (t-s)^{\alpha-1} f(s) \mathrm{d}s, \quad t \geq a, \tag{5.6}$$

$$(I_{b^-}^{\alpha} f)(t) \triangleq \frac{1}{\Gamma(\alpha)} \int_t^b (s-t)^{\alpha-1} f(s) \mathrm{d}s, \quad t \leq b, \tag{5.7}$$

is called the *right-* and *left fractional integral* of order α, respectively.

One can also define fractional differentiation as follows.

Definition 5.2. For functions f given in the interval $[a, b]$, the *left-handed* and *right-handed fractional derivative* of order $\alpha > 0$, is defined by,

$$(D^{\alpha}_{a^+} f)(t) \triangleq \left(\frac{\mathrm{d}}{\mathrm{d}t}\right)^{[\alpha]+1} I^{1-\{\alpha\}}_{a^+} f(t),$$
$$(D^{\alpha}_{b^-} f)(t) \triangleq \left(-\frac{\mathrm{d}}{\mathrm{d}t}\right)^{[\alpha]+1} I^{1-\{\alpha\}}_{b^-} f(t),$$

respectively where $[\alpha]$ denotes the integer part of α and $\{\alpha\} = \alpha - [\alpha]$.

The connection between fractional integration and differentiation is given by the following theorem (see Samko et. al. [26] Theorem 2.4, p. 44).

Theorem 5.3. *For $\alpha > 0$ we have,*

$$D^{\alpha}_{a^+} I^{\alpha}_{a^+} f = f, \quad \text{for } f \in L^1([a, b]), \tag{5.8}$$
$$I^{\alpha}_{a^+} D^{\alpha}_{a^+} f = f, \quad \text{for } f \in I^{\alpha}_{a^+}(L^1([a, b])). \tag{5.9}$$

Because of this theorem we will sometimes write, $I^{-\alpha}_{a^+}$ for $D^{\alpha}_{a^+}$. By Corollary 2 on p. 46 in Samko et. al. [26], we have the integration by parts formula

$$\int_a^b f(t)(D^{\alpha}_{a^+} g)(t)dt = \int_a^b g(t)(D^{\alpha}_{b^-} f)(t)dt, \tag{5.10}$$

for $0 < \alpha < 1$ and $f \in I^{\alpha}_{b^-}(L^p), g \in I^{\alpha}_{a^+}(L^q), 1/p + 1/q \leq 1 + \alpha$. Similarly, we can define fractional integration and differentiation on the real line by similar expressions

$$(I^{\alpha}_{+} f)(t) \triangleq \frac{1}{\Gamma(\alpha)} \int_{-\infty}^{t} (t - s)^{\alpha-1} f(s)\mathrm{d}s, \quad t \in \mathbb{R}, \tag{5.11}$$
$$(I^{\alpha}_{-} f)(t) \triangleq \frac{1}{\Gamma(\alpha)} \int_{t}^{\infty} (s - t)^{\alpha-1} f(s)\mathrm{d}s, \quad t \in \mathbb{R}. \tag{5.12}$$

We refer to Samko et. al. [26] Chapter 2 for further details.

5.2.1 Gaussian processes of Volterra type

Let us denote by T an index set which will be a compact interval or the whole of $\mathbb{R}$. We will often use the unit interval and therefore we introduce the notation $I = [0, 1] \subset \mathbb{R}$. We may think of $t \in T$ as time. Consider a deterministic function $V : T \times T \to [0, \infty)$ satisfying the following hypothesis.

Hypothesis 1.

(H1) $V(0,s) = 0$ for all $s \in T$ and $V(t,s) = 0$ for $s > t$.

(H2) There are constants $C, \alpha > 0$ such that for all $s, t \in T$

$$\int_T (V(t,r) - V(s,r))^2 \, \mathrm{d}r \leq C|t-s|^\alpha.$$

(H3) V is injective as a transformation of functions in $L^2(T)$

$$(Vf)(t) = \int_T V(t,s) f(s) \mathrm{d}s, \qquad f \in L^2(T).$$

Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space and $B = \{B_t : t \in T\}$ denote the standard Brownian motion. If $T = \mathbb{R}$ the standard Brownian motion is constructed as $B_t = B'_t$ if $t \geq 0$ and $B_t = B''_{-t}$ if $t < 0$ where $\{B'_t : t \geq 0\}$ and $\{B''_t : t \geq 0\}$ are independent Brownian motions starting at $t = 0$. Let $X = \{X_t : t \in T\}$ be the process defined by

$$X_t = \int_T V(t,s) \mathrm{d}B_s, \quad t \in T. \tag{5.13}$$

Since (H2) implies that $V(t,\cdot) \in L^2(T)$ for each $t \in T$, the process X is well defined. Clearly X is Gaussian and has covariance function

$$\rho(t,s) = \mathbb{E}\,(X_t X_s) = \int_T V(t,r) V(s,r) \mathrm{d}r. \tag{5.14}$$

The assumption (H2) guarantees the existence of a Hölder continuous modification of the process X. Indeed, since the process is Gaussian, for even $p \geq 2$ there is a constant C_p such that for each $t, s \in T$

$$\begin{aligned}
\mathbb{E}\,(|X_t - X_s|^p) &= \mathbb{E}\left(\left| \int_T V(t,r) \mathrm{d}B_r - \int_T V(s,r) \mathrm{d}B_r \right|^p \right) \\
&\leq C_p \left(\mathbb{E} \left| \int_T V(t,r) \mathrm{d}B_r - \int_T V(s,r) \mathrm{d}B_r \right|^2 \right)^{p/2} \\
&= C_p \left(\int_T (V(t,r) - V(s,r))^2 \, \mathrm{d}r \right)^{p/2} \\
&\leq C'_p |t-s|^{\alpha p/2} \leq C'_p |t-s|^{1+\delta},
\end{aligned}$$

for some $\delta > 0$ and $p > 2/\alpha$. Hence, Kolmogorov's condition is satisfied so X has a modification with sample paths which are Hölder continuous of index β for every $\beta < \alpha/2$. Note also that the image of $L^2(T)$ under the integral transform in (H3) consists of continuous functions. Clearly, by (H2) and the Cauchy-Schwartz inequality the function $(Vf)(t)$, $f \in L^2(T)$, is continuous. We also remark that by (H1) X is adapted to the natural filtration generated by B. Finally, note that a

sufficient condition for the operator V in (H3) to be injective is that $V(t,s) > 0$ for Lebesgue a.e. $t > s$. Indeed, put $\phi(t) = f(t) - g(t)$, suppose that $(V\phi)(t) = 0$ for every $t \in T$ and let $t_0 < t_1 \in T$. Then it follows that,

$$\int_{t_0}^{t} V(t_1, s)\phi(s)\mathrm{d}s = -\int_{t}^{t_1} V(t_1, s)\phi(s)\mathrm{d}s, \quad \text{for every } t_0 \leq t \leq t_1.$$

If we differentiate the left- and the right-hand side with respect to t then it holds that $V(t_1, t)\phi(t) = -V(t_1, t)\phi(t)$ for every $t_0 \leq t \leq t_1$ which implies $\phi(t) = 0$ for every $t_0 \leq t \leq t_1$ where $V(t_1, t) > 0$. Since t_0, t_1 was arbitrary $\phi(t) = 0$ for Lebesgue a.e. $t \in T$.

Hypothesis 1 enables the following setup which we will use throughout the paper. Let $V : T \times T \to [0, \infty)$ satisfy Hypothesis 1. We take $\Omega = C(T)$, the space of continuous functions, equipped with the Borel σ-field $\mathcal{A}$. The canonical process $X = \{X_t : t \in T\}$ is defined by $X_t(\omega) = \omega(t)$ and the probability measure $\mathbb{P}$ on $(\Omega, \mathcal{A})$ is the unique measure such that X is centered Gaussian with covariance given by (5.14). Then the representation (5.13) holds in law. We denote by $\{\mathcal{F}_t^X : t \in T\}$ the natural filtration generated by X.

Definition 5.4. A centered Gaussian process $X = \{X_t : t \in T\}$ with covariance (5.14) and kernel V satisfying Hypothesis 1 is called a *Volterra type process.*

The following are our primary examples.

Example 5.5 *(Brownian motion).* Let $T = I$ and

$$V(t, s) = 1_{[0,t]}(s).$$

Then V satisfies Hypothesis 1 and X is the standard Brownian motion with covariance function $\rho(t, s) = t \wedge s$. We will denote standard Brownian motion by $\{B_t : t \in T\}$. Note that, as an integral operator acting on functions in $L^2(T)$, V is simply the integration operator

$$(Vf)(t) = \int_T V(t, s)f(s)\mathrm{d}s = \int_0^t f(s)\mathrm{d}s = (I_{0+}^1 f)(t).$$

Example 5.6 *(Ornstein-Uhlenbeck process).* Let $T = I$ and

$$V(t, s) = e^{\theta(t-s)}1_{[0,t]}(s).$$

Then V satisfies Hypothesis 1 and X is the Ornstein-Uhlenbeck process. That is, the solution to the stochastic differential equation

$$\mathrm{d}X_t = \theta X_t \mathrm{d}t + \mathrm{d}B_t, \quad X_0 = 0.$$

This can easily be verified using Itô's formula.

Example 5.7 *(Fractional Brownian motion).* For $H \in (0,1)$, let $T = I$ and $V(t,s) = K_H(t,s)$ where

$$K_H(t,s) \triangleq \frac{1}{\sqrt{V_H}\Gamma(H+\frac{1}{2})}(t-s)^{H-\frac{1}{2}}{}_1F_2\Big(H-\frac{1}{2}, \frac{1}{2}-H, H+\frac{1}{2}, 1-\frac{t}{s}\Big)1_{[0,t)}(s),$$

with ${}_1F_2$ the Gauss hypergeometric function and

$$V_H \triangleq \frac{\Gamma(2-2H)\cos(\pi H)}{\pi H(1-2H)} \tag{5.15}$$

a normalizing constant which makes $\mathrm{Var}[X(1)] = 1$. Then V satisfies Hypothesis 1 and X is called the fractional Brownian motion with index H (see Decreusefond and Üstünel [9]). Note that, as an integral operator acting on functions in $L^2(I)$, V is the operator

$$(Vf)(t) = \int_I V(t,s)f(s)\mathrm{d}s = \int_0^t K_H(t,s)f(s)\mathrm{d}s = (K_H f)(t),$$

which is an isomorphism from $L^2(I)$ onto $I_{0^+}^{H+1/2}(L^2(I))$ and

$$\sqrt{V_H}K_H\, f = \begin{cases} I_{0^+}^{2H}t^{1/2-H}I_{0^+}^{1/2-H}t^{H-1/2}\, f, & \text{for } H \le 1/2, \\ I_{0^+}^{1}t^{H-1/2}I_{0^+}^{H-1/2}t^{1/2-H}\, f, & \text{for } H \ge 1/2, \end{cases}$$

see Samko et. al. [26] p. 187. The fBm has stationary increments and covariance function

$$\rho(t,s) = \mathbb{E}\left(X_t X_s\right) = \frac{1}{2}\left(t^{2H} + s^{2H} - |t-s|^{2H}\right).$$

We will denote fBm by $\{B_t^H : t \in I\}$. One of the most important properties of the fBm is self-similarity in the sense that for any $c > 0$,

$$\left\{B^H(ct) : t \in I\right\} \overset{\mathrm{d}}{=} \left\{c^H B^H(t) : t \in I\right\}.$$

This property allows us to give a representation of fBm on any interval. In fact, the representation

$$B_t^H = \int_0^t K_H(t,u)\mathrm{d}B_u,$$

holds for any $t \geq 0$. Indeed, since $K_H(t,u) = T^{H-1/2}K_H(t/T, u/T)$ for any $0 \leq u \leq t \leq T$ we get for any $0 \leq s \leq t$

$$\begin{aligned}
\mathbb{E}\Big(\int_0^t K_H(t,u)\mathrm{d}B_u \int_0^s K_H(s,v)\mathrm{d}B_v \Big) &= \int_0^s K_H(t,u)K_H(s,u)\mathrm{d}u \\
&= \int_0^{s/t} t^{2H-1} K_H(1,w)K_H(s/t,w)t\mathrm{d}w \\
&= t^{2H}\frac{1}{2}(1^{2H} + (s/t)^{2H} - |1-s/t|^{2H}) \\
&= \frac{1}{2}(t^{2H} + s^{2H} - |t-s|^{2H})
\end{aligned}$$

which is the covariance function of the fBm.

One can also define the fractional Brownian on the whole real line in another way. This moving average representation is often used in the literature. For $H \in (0,1)$ let

$$M_H(t,s) = \frac{1}{C_1(H)}\left((t-s)_+^{H-1/2} - (-s)_+^{H-1/2} \right), \quad s,t \in \mathbb{R},$$

where $u_+ = \max(0,u)$ and

$$C_1(H) = \left\{ \int_0^\infty \left((1+s)^{H-1/2} - s^{H-1/2} \right)^2 \mathrm{d}s + \frac{1}{2H} \right\}^{1/2}.$$

Then X given by $X_t = \int_{-\infty}^\infty M_H(t,s)\mathrm{d}B_s$ is the fractional Brownian motion (see e.g. Samorodnitsky and Taqqu [27] p. 321). This function can be expressed in terms of the Marchaud fractional derivative $\mathbb{D}$ (see Pipiras and Taqqu [25] and Samko et. al. [26]) for $H < 1/2$

$$M_H(t,s) = \frac{\Gamma(H+1/2)}{C_1(H)}(\mathbb{D}_-^{-(H-1/2)}1_{[0,t)})(s) \tag{5.16}$$

and for $H > 1/2$ as

$$M_H(t,s) = \frac{\Gamma(H+1/2)}{C_1(H)}(\mathbb{I}_-^{H+1/2}1_{[0,t)})(s) \tag{5.17}$$

Note that, as an integral operator acting on functions in $L^2(\mathbb{R})$, M_H can be written as

$$(M_H f)(t) = \frac{1}{C_1(H)} \int_{-\infty}^\infty \left((t-s)_+^{H-1/2} - (-s)_+^{H-1/2} \right) f(s)ds,$$

or in terms of the fractional integration operator in (5.11)

$$(M_H f)(t) = \frac{\Gamma(H+1/2)}{C_1(H)}(I_+^{H+1/2}f)(t) - \frac{\Gamma(H+1/2)}{C_1(H)}(I_+^{H+1/2}f)(0).$$

It should be noted that for $t < 0$, M_H is not a Volterra type operator since (H1) is not satisfied.

Since the kernel of the fractional Brownian motion is fairly complicated one sometimes modifies it to simplify computations. One such modification is given in the next example where we simply remove the Gauss hypergeometric function (and its singularity at zero). This process is sometimes referred to as fractional Brownian motion of type II, fBm(II).

Example 5.8 *(Fractional Brownian motion of type II).* For $H \in (0,1)$, let $T = I$ and $V(t,s) = J_H(t,s)$ where

$$J_H(t,s) \triangleq \frac{1}{\Gamma(H+\frac{1}{2})}(t-s)^{H-\frac{1}{2}}1_{[0,t)}(s).$$

Then V satisfies Hypothesis 1 and X is fractional Brownian motion of type II (see Feyel and la Pradelle [11]). It is convenient to use the parameterization $\alpha = H + 1/2$ when working with fBm(II). The fBm(II) has non-stationary increments and covariance function

$$\rho(t,s) = \mathbb{E}\left(X_t X_s\right) = \frac{1}{\Gamma(\alpha)^2}\int_0^{t\wedge s}(t-r)^{\alpha-1}(s-r)^{\alpha-1}\mathrm{d}r.$$

We will denote fBm(II) by $\{W_t^H : t \in I\}$. Note that, as an integral operator acting on functions in $L^2(I)$, V is the Liouville fractional integration operator

$$(Vf)(t) = \int_I V(t,s)f(s)\mathrm{d}s = \int_0^t J_H(t,s)f(s)\mathrm{d}s = (I_{0^+}^{\alpha}f)(t).$$

Example 5.9 *(Multifractal Brownian motion).* Let $T = \mathbb{R}$, $H \in C^r(\mathbb{R};(0,1))$ with $\sup_{t\in T} H(t) < r$ and put

$$V(t,s) = M_{H(t)}(t,s) = \frac{1}{C_1(H(t))}\left((t-s)_+^{H(t)-1/2} - (-s)_+^{H(t)-1/2}\right), \quad s,t \in \mathbb{R}.$$

Then V satisfies Hypothesis 1 for $t > 0$. For $t < 0$ it does not satisfy (H1). The process X is then the multifractal Brownian motion of Benassi, Jaffard and Roux [3]. In [3] the multifractal Brownian motion is introduced through a spectral representation as

$$X^H(t) = \int_{-\infty}^{\infty}\frac{e^{it\xi}-1}{|\xi|^{H(t)+1/2}}\widehat{M}(\mathrm{d}\xi)$$

where $\widehat{M}$ is a complex Gaussian random measure. Let us show that the two definitions gives the same process in law. Writing $M_{H(t)}(t,s)$ as in (5.16) or (5.17) if

$H(t) > 1/2$ or $H(t) < 1/2$ (or $1_{[0,t)}(s)$ if $H = 1/2$) we see that by Proposition 3.3 (3) in Pipiras and Taqqu [25] the Fourier transform of $M_{H(t)}(t,\cdot)$ equals

$$\widehat{M}_{H(t)}(t,\xi) = \mathcal{F}\{M_{H(t)}(t,\cdot)\}(\xi) = \frac{e^{it\xi}-1}{(i\xi)|\xi|^{H(t)-1/2}}.$$

By Proposition 7.2.7 in Samorodnitsky and Taqqu [27] it follows that the process $Y = \{Y(t) : t \in \mathbb{R}\}$ defined by

$$Y(t) = \int_{-\infty}^{\infty} \frac{e^{it\xi}-1}{(i\xi)|\xi|^{H(t)-1/2}} \widehat{M}(\mathrm{d}\xi)$$

has the same distribution as X and since the covariance functions of Y and X^H coincides we also have $Y \stackrel{\mathrm{d}}{=} X^H$. Thus $X \stackrel{\mathrm{d}}{=} X^H$ which is the multifractal Brownian motion of Benassi et. al. [3].

5.2.2 The reproducing kernel Hilbert space

Let us introduce the general definition of a reproducing kernel.

Definition 5.10. A function ρ defined on $T \times T$ is said to be a *reproducing kernel* for the Hilbert space $\mathcal{H}$ of functions on T if

(1) $\rho(t,\cdot) \in \mathcal{H}$ for each $t \in T$,
(2) $\langle G, \rho(t,\cdot)\rangle_{\mathcal{H}} = G(t)$, for each $G \in \mathcal{H}$.

For Gaussian processes we can construct the reproducing kernel Hilbert space with reproducing kernel $\rho(t,s) = \mathbb{E}[X_t X_s]$ as follows. Let H be the closure in $L^2(\Omega)$ of the space spanned by $\{X_t : t \in T\}$ equipped with the inner product $\langle \xi, \zeta\rangle_H = \mathbb{E}(\xi\zeta)$, for $\xi, \zeta \in H$. The reproducing kernel Hilbert space $\mathcal{H}$ associated with X is the space $R(H) = \{R(\xi) : \xi \in H\}$ where for any $\xi \in H$, $R(\xi)$ is the function $R(\xi)(t) = \langle \xi, X_t\rangle_H = \mathbb{E}(\xi X_t)$. $\mathcal{H}$ has inner product $\langle F, G\rangle_{\mathcal{H}} = \langle R^{-1}F, R^{-1}G\rangle_H$. Since $\rho(s,\cdot) = R(X_s)$, we clearly have $\rho(s,\cdot) \in \mathcal{H}$. Furthermore, (2) is satisfied since for $\xi \in H$, $\langle R(\xi), \rho(s,\cdot)\rangle_{\mathcal{H}} = \mathbb{E}(\xi X_s) = R(\xi)(s)$. For more details on reproducing kernel Hilbert spaces we refer to Grenander [12] or Janson [13].

For processes of the form (5.13) where the kernel V satisfies (H2) and (H3) we have the expression (5.14) of the covariance function. Then the reproducing kernel Hilbert space can also be represented as the image of $L^2(T)$ under the integral transform V, $\mathcal{H} = V(L^2(T))$ equipped with the inner product, $\langle F, G\rangle_{\mathcal{H}} = \langle V^{-1}F, V^{-1}G\rangle_{L^2(T)}$. Indeed, it is clear that $V(L^2(T))$ is a Hilbert space with $\rho(t,\cdot) \in V(L^2(T))$, proving (1) and for any $F = Vf$ we have

$$\langle F, \rho(t,\cdot)\rangle_{\mathcal{H}} = \langle f, V(t,\cdot)\rangle_{L^2(T)} = \int_T V(t,s)f(s)\mathrm{d}s = F(t).$$

which proves (2).

We also note that if $K \subseteq T$ is compact then the embedding $\mathcal{H} \hookrightarrow \{F|_K : F \in C(T)\}$ given by $F \mapsto F|_K$ is continuous, where $F|_K$ denotes the restriction of F to K. Indeed, since K is compact we have by (H2) that there is a constant C such that for each $t \in K$, $\int_T V(t,s)^2 \mathrm{d}s \leq C$. It follows that for each $F(t) = \int_T V(t,s) f(s) \mathrm{d}s$,

$$\begin{aligned}\sup_{t\in K} |F(t)| &= \sup_{t\in K} \Big| \int_T V(t,s) f(s) \mathrm{d}s \Big| \\ &\leq \sup_{t\in K} \Big(\int_T |V(t,s)|^2 ds \Big)^{1/2} \Big(\int_T |f(s)|^2 \mathrm{d}s \Big)^{1/2} \\ &\leq C \|f\|_{L^2(T)} = C\|F\|_{\mathcal{H}}.\end{aligned}$$

Finally, by Theorem 3 in Kallianpur [14] we have that, for compact T, the closure of $\mathcal{H}$ in $C(T)$ is equal to the support of $\mathbb{P}$.

Example 5.11 *(Brownian motion).* For the standard Brownian motion on $T = I$ we saw that V is just the integration operator. Therefore the reproducing kernel Hilbert space is simply the Sobolev space $W_0^{1,2}$ of differentiable functions, vanishing at zero, with first derivative in $L^2(T)$.

Example 5.12 *(Fractional Brownian motion on I).* For the fractional Brownian motion on $T = I$ we have $V = K_H$. Since K_H is an isomorphism from $L^2(I)$ onto $I_{0^+}^{H+1/2}(L^2(I))$ the reproducing kernel Hilbert space is $I_{0^+}^{H+1/2}(L^2(I))$ with the inner product $\langle F, G\rangle_{\mathcal{H}} = \langle K_H^{-1} F, K_H^{-1} G\rangle_{L^2(I)}$.

Example 5.13 *(Fractional Brownian motion on $\mathbb{R}$).* For the fractional Brownian motion on $T = \mathbb{R}$ we have the representation $X_t = \int_{-\infty}^{\infty} M_H(t,s) \mathrm{d}B_s$ which is not a Volterra type representation for $t < 0$ but satisfies (H2) and (H3). We conclude that the reproducing kernel Hilbert space is the space $M_H(L^2(\mathbb{R}))$ (see also Pipiras and Taqqu [25]).

Example 5.14 *(Fractional Brownian motion of type II).* For $V = I_{0^+}^{\alpha}$, i.e. in the case of fractional Brownian motion of type II on $T = I$, the reproducing kernel Hilbert space is the space $I_{0^+}^{\alpha}(L^2(T))$ with the inner product $\langle F, G\rangle_{\mathcal{H}} = \langle I_{0^+}^{-\alpha} F, I_{0^+}^{-\alpha} G\rangle_{L^2(T)}$. That is, the fractional Sobolev space $W_0^{\alpha,2}$.

Example 5.15 *(Multifractal Brownian motion).* As for the Brownian motion on $\mathbb{R}$ we can obtain the reproducing kernel Hilbert space for the mutifractal Brownian motion as $M_{H(\cdot)}(L^2(\mathbb{R}))$. An alternative representation is given in Benassi, Jaffard and Roux [3] as the set of functions of the form

$$f_\varphi(t) = \int_{-\infty}^{\infty} \frac{e^{it\xi} - 1}{|\xi|^{H(t)+1/2}} \widehat{\varphi}(\xi) \mathrm{d}\xi, \quad t \in \mathbb{R},$$

where $\widehat{\varphi}$ is the Fourier transform of $\varphi \in L^2(\mathbb{R})$.

5.3 Representation of Volterra type stochastic integrals

In this section we apply a general representation theorem for Gaussian processes to Volterra type processes. As examples we obtain the Karhunen-Loéve decomposition and the classical Lévy construction of standard Brownian motion and also a wavelet representation for the fractional Brownian motion studied in Meyer et. al. [21]. In the Section 5.5 we will apply this representation in the context of parameter estimation. The general representation theorem mentioned above is the following, which is proved in e.g. Janson [13] Theorem 8.22.

Theorem 5.16. *Suppose that* $\{X_t : t \in \Lambda\}$ *is a Gaussian process on some index set* Λ *and* $\{\Psi_j : j = 1, 2, \dots\}$ *is a countable orthonormal basis in the associated reproducing kernel Hilbert space* $\mathcal{H}$*. Then there exist independent standard normal random variables* $\{\xi_j : j = 1, 2, \dots\}$ *such that for each* $t \in \Lambda$*,*

$$X_t = \sum_{j=1}^{\infty} \Psi_j(t)\xi_j \tag{5.18}$$

with the sum converging in $L^2(\Omega)$ *and almost surely.*

Conversely, for any sequence $\{\xi_j : j = 1, 2, \dots\}$ *of independent standard normal random variables, the sum converges almost surely for each* $t \in \Lambda$ *and defines a centered Gaussian process with the same distribution as* $\{X_t : t \in \Lambda\}$*.*

As indicated in Remark 8.24 in Janson [13] we can prove the following corollary.

Corollary 5.17. *If* X *is a.s. continuous and* $T \subset \Lambda$ *is compact then the sum in (5.18) converges almost surely uniformly on* T*.*

Proof. Since X is a.s. continuous, $\mathcal{H}$ consists of continuous functions on Λ. We consider the compact subset $T \subset \Lambda$. We have in particular that each $\Psi_j(\cdot)$ restricted to T is in $C(T)$ so the partial sums,

$$S_n(\cdot) = \sum_{j=1}^{n} \Psi_j(\cdot)\xi_j$$

form a random sequence of elements in $C(T)$. Now, since T is compact, $C(T)^* \cong M(T)$, where $C(T)^*$ is the dual of $C(T)$ and $M(T)$ is the space of finite signed regular Borel measures on T. On Λ the function $\rho(t, \cdot)$ can be expanded in the basis $\{\Psi_j : j = 1, 2, \dots\}$ and we find that

$$\rho(t, \cdot) = \sum_{j=1}^{\infty} \Psi_j(\cdot)\langle \Psi_j(\cdot), \rho(t, \cdot)\rangle_{\mathcal{H}} = \sum_{j=1}^{\infty} \Psi_j(\cdot)\Psi_j(t).$$

Furthermore, $\mathbb{E}\left[S_n(t)^2\right] = \sum_{j=1}^n \Psi_j(t)^2 \leq \sum_{j=1}^\infty \Psi_j(t)^2 = \rho(t,t)$ and $\mathbb{E}\left[X(t)^2\right] = \rho(t,t)$. Since $S_n(t) \to X(t)$ in $L^2(\Omega)$ for each t, and $\rho(t,t)$ is bounded on T, the bounded convergence theorem implies that for $\mu \in M(T)$,

$$\int_T S_n(t)\mu(\mathrm{d}t) \to \int_T X(t)\mu(\mathrm{d}t), \quad \text{in } L^2(\Omega).$$

The statement now follows from the Lévy-Itô-Nisio Theorem (see Ledoux and Talagrand [18] Theorem 2.4). □

In our quest for finding an explicit representation we are left with the problem of finding an orthonormal basis in $\mathcal{H}$. This is however very simple because $\mathcal{H} = V(L^2(T))$. Simply take any orthonormal basis $\{\psi_j : j = 1, 2, \dots\}$ in $L^2(T)$ and apply the integral transform V on each function. Then $\{\Psi_j : j = 1, 2, \dots\}$ with $\Psi_j(\cdot) = (V\psi_j)(\cdot)$, is an orthonormal basis in $V(L^2(T))$. Indeed,

$$\langle \Psi_j, \Psi_k \rangle_{V(L^2(T))} = \langle V^{-1}\Psi_j, V^{-1}\Psi_k \rangle_{L^2(T)} = \langle \psi_j, \psi_k \rangle_{L^2(T)} = 0, \quad j \neq k$$

and

$$\|\Psi_j\|_{V(L^2(T))} = \|\psi_j\|_{L^2(T)} = 1.$$

Example 5.18 *(Brownian motion).* Let $\{\psi_j : j = 0, 1, \dots\}$ be the orthonormal basis of $L^2(I)$ defined by

$$\psi_j(t) = \sqrt{2}\cos(t/\lambda_j), \qquad j \geq 0,$$

where $\lambda_j = 1/\pi(j + 1/2)$, $j \geq 0$. Then the functions

$$\Psi_j(t) = (I^1_{0+}\psi_j)(t) = \int_0^t \sqrt{2}\cos(r/\lambda_j)\mathrm{d}r = \lambda_j\sqrt{2}\sin(t/\lambda_j), \qquad j \geq 0,$$

form an orthonormal basis in the reproducing kernel Hilbert space $W_0^{1,2}$ and

$$B_t = \sum_{j=0}^\infty \lambda_j\sqrt{2}\sin(t/\lambda_j)\xi_j,$$

where $\{\xi_j : j = 0, 1, \dots\}$ is a sequence of iid $N(0,1)$ random variables. This is the standard Karhunen-Loéve expansion for Brownian motion.

Example 5.19 *(Brownian motion).* Let $\{\psi_n : n = 0, 1, \dots\}$ be the Haar basis of $L^2(I)$ defined by,

$$\psi_0(t) = 1_{[0,1]}(t),$$
$$\psi_n(t) = 2^{j/2}\psi\left(2^j t - k\right), \quad n = 2^j + k,\ j \geq 0 \text{ and } 0 \leq k < 2^j.$$

where $\psi(t) = 1$ on $[0, 1/2)$ and $\psi(t) = -1$ on $[1/2, 1]$. Integrating gives us the basis $\{\Psi_n : n = 0, 1, \dots\}$ of the reproducing kernel Hilbert space,

$$\Psi_0(t) = t,$$
$$\Psi_n(t) = 2^{-j/2}\Psi\left(2^j t - k\right), \qquad n = 2^j + k,$$

where $\Psi(\cdot)$ is the primitive of $\psi(\cdot)$,

$$\Psi(t) = \frac{1}{2}\max(0, 1 - |2t - 1|).$$

We obtain the representation,

$$B_t = \sum_n \Psi_n(t)\xi_n,$$

where $\{\xi_n : n = 0, 1, \dots\}$ is an iid $N(0, 1)$ sequence. This is a classical construction of Brownian motion by Lévy (see e.g. Steele [28] for a nice exposition).

Remark 5.20. We see that any orthonormal $L^2(T)$-basis, $\{\psi_j : j = 1, 2, \dots\}$, yields a representation of Brownian motion as,

$$B_t = \sum_j \Psi_j(t)\xi_j,$$

where, $\Psi_j(t) = \int_0^t \psi(s)\mathrm{d}s$ and $\{\xi_j\}_1^\infty$ is a sequence of iid $N(0, 1)$ random variables. This is of course a very well known result.

5.3.1 A wavelet representation of fractional Brownian motion

Consider the fractional Brownian motion on $T = \mathbb{R}$ with the kernel $M_H(t, s)$. Even though M_H is not of Volterra type for $t < 0$ it is easy to verify (see Example 5.13) that the reproducing kernel Hilbert space is $\mathcal{H} = M_H(L^2(\mathbb{R}))$. Let $\{\phi, \psi\}$ be a wavelet pair such that $\{\phi_k, \psi_{j,k}\}$, where,

$$\phi_k(t) = \phi(t - k) \quad k = 0, \pm 1, \pm 2, \dots$$
$$\psi_{j,k}(t) = 2^{j/2}\psi(2^j t - k), \quad j \geq 0,\ k = 0, \pm 1, \pm 2, \dots$$

form a basis in $L^2(\mathbb{R})$. For instance we may take the Meyer wavelets which have vanishing moments of all orders (see Meyer et. al. [21] and references therein). Since

B^H has representation (5.18) there exists independent sequences $\{\eta_k\}$ and $\{\xi_{j,k}\}$ of iid $N(0,1)$ random variables such that (5.18) can be written as,

$$B^H(t) = \sum_{k=-\infty}^{\infty} \Phi(t-k)\eta_k + \sum_{j\geq 0}\sum_{k=-\infty}^{\infty} 2^{-jH}\Psi(2^j t - k)\xi_{j,k}, \tag{5.19}$$

where $\Phi(t) = (M_H\phi)(t)$, $\Psi(t) = (M_H\psi)(t)$ and M_H is the integral transform in Example 5.7. Indeed, a simple transformation of variables yields

$$\begin{aligned}(M_H\phi_k)(t) &= (M_H\phi)(t-k) = \Phi(t-k),\\ (M_H\psi_{j,k})(t) &= 2^{-jH}(M_H\psi)(2^j t - k) = 2^{-jH}\Psi(2^j t - k).\end{aligned}$$

By Corollary 5.17, convergence holds almost surely uniformly on compact intervals.

From this representation of fBm we can, as a special case, derive the wavelet representation

$$B^H(t) = c_H \sum_{k=-\infty}^{\infty} S_k^H \Phi_H(t-k) + c_H \sum_{j\geq 0}\sum_{k=-\infty}^{\infty} 2^{-jH}\Psi_H(2^j t - k)\xi_{j,k} - b_0, \tag{5.20}$$

given in a recent paper by Meyer et. al. [21]. Here convergence holds uniformly on compact intervals, $\{S_k\}$ is a fractional ARIMA process and $\{\xi_{j,k}\}$ is a sequence of iid $N(0,1)$ random variables, independent of $\{S_k\}$. The random variable b_0 is a correction such that $B^H(0) = 0$ and the functions Φ_H and Ψ_H are defined via their Fourier transform as

$$\begin{aligned}\widehat{\Phi}_H(\xi) &= \Big(\frac{1-e^{i\xi}}{i\xi}\Big)^{H+1/2}\widehat{\phi}(\xi),\\ \widehat{\Psi}_H(\xi) &= (i\xi)^{-1/2-H}\widehat{\psi}(\xi).\end{aligned}$$

In the representation (5.20) we have included the constants c_H to have a representation of the standard fBm ($\mathrm{Var}[B^H(1)] = 1$). Let us show how one obtains (5.20) from (5.19).

We start with the second sum in (5.19). For $f \in \mathcal{S}(\mathbb{R})$, the Schwartz space of rapidly decreasing functions on $\mathbb{R}$, with vanishing moments of all orders, the Fourier transform of $I_+^\alpha f$, $\alpha > 0$, is (see Samko et. al. [26] Section 8.2)

$$\mathcal{F}\{I_+^\alpha f\}(\xi) = (i\xi)^{-\alpha}\widehat{f}(\xi).$$

This implies that

$$\Psi_H(t) = (I_+^{H+1/2}\psi)(t).$$

Using Definition 5.11 of the fractional integral we see that

$$\begin{aligned}\Psi(t) = (M_H\psi)(t) &= \frac{1}{C_1(H)}\int_{-\infty}^{\infty}[(t-s)_+^{H-1/2} - (-s)_+^{H-1/2}]\psi(s)\mathrm{d}s \\ &= \frac{1}{C_1(H)}\Big(\int_{-\infty}^{t}(t-s)^{H-1/2}\psi(s)\mathrm{d}s - \int_{-\infty}^{0}(-s)^{H-1/2}\psi(s)\mathrm{d}s\Big) \\ &= \frac{\Gamma(H+1/2)}{C_1(H)}\big((I_+^{H+1/2}\psi)(t) - (I_+^{H+1/2}\psi)(0)\big)\end{aligned}$$

So we have

$$\Psi(t) = c_H\Psi_H(t) - c_H\Psi_H(0),$$

with $c_H = \Gamma(H+1/2)/C_1(H)$. Hence,

$$\begin{aligned}\sum_{j\geq 0}\sum_{k=-\infty}^{\infty} 2^{-jH}\Psi(2^jt-k)\xi_{j,k} = c_H\sum_{j\geq 0}\sum_{k=-\infty}^{\infty} 2^{-jH}\Psi_H(2^jt-k)\xi_{j,k} \\ - c_H\sum_{j\geq 0}\sum_{k=-\infty}^{\infty} 2^{-jH}\Psi_H(-k)\xi_{j,k}\end{aligned}$$

and we include the second sum in the last expression in b_0.

Let us now focus on the first sum in (5.19). We want to show that

$$\sum_{k=-\infty}^{\infty}\eta_k\Phi(t-k) = c_H\sum_{k=-\infty}^{\infty}S_k^{(H)}\Phi_H(t-k) - c_H\sum_{k=-\infty}^{\infty}S_k^{(H)}\Phi_H(-k)$$

and this requires a little more work. First note that

$$\Phi(t) = c_H\big((I_+^{H+1/2}\phi)(t) - (I_+^{H+1/2}\phi)(0)\big) = c_H\Phi^{(1)}(t) - c_H\Phi^{(1)}(0),$$

where $\Phi^{(1)}(t) = (I_+^{H+1/2}\phi)(t)$. If we include the sum $c_H\sum_{k=-\infty}^{\infty}\Phi^{(1)}(-k)\eta_k$ in b_0, then it is sufficient to show that

$$\sum_{k=-\infty}^{\infty}\Phi^{(1)}(t-k)\eta_k = \sum_{k=-\infty}^{\infty}\Phi_H(t-k)S_k^{(H)}. \tag{5.21}$$

The fractional ARIMA process $S_k^{(H)}$ can be defined as

$$S_k^{(H)} = \sum_{l=0}^{\infty}\gamma_l^{(-1/2-H)}(\eta_{k-l} - \eta_{-l}), \qquad k = 0, \pm 1, \pm 2, \ldots$$

where $\gamma_l^{(-d)}$ is the coefficients in the expansion $(1-x)^{-d} = \sum_{l=0}^{\infty}\gamma_l^{(-d)}x^l$. We also have (see Meyer et. al. [21] Section 4) that for a function $f \in \mathcal{S}(\mathbb{R})$

$$\mathcal{F}\Big\{\sum_{l=0}^{\infty}\gamma_l^{(-d)}f(t-l)\Big\}(\xi) = (1-e^{i\xi})^{-d}\widehat{f}(\xi).$$

Since $\Phi_H \in \mathcal{S}(\mathbb{R})$ (see Lemma 6 in Meyer et. al. [21]) it follows that

$$\begin{aligned}\mathcal{F}\{I_+^{H+1/2}\phi\}(\xi) &= (i\xi)^{-1/2-H}\widehat{\phi}(\xi) = (1-e^{i\xi})^{-1/2-H}\widehat{\Phi}_H(\xi) \\ &= \mathcal{F}\Big\{\sum_{l=0}^{\infty}\gamma_l^{(-1/2-H)}\Phi_H(t-l)\Big\}(\xi)\end{aligned}$$

and then (5.21) follows since

$$\begin{aligned}\sum_{k=-\infty}^{\infty}\Phi^{(1)}(t-k)\eta_k &= \sum_{k=-\infty}^{\infty}(I_+^{H+1/2}\phi)(t-k)\eta_k \\ &= \sum_{k=-\infty}^{\infty}\sum_{l=0}^{\infty}\gamma_l^{(-1/2-H)}\Phi_H(t-l-k)\eta_k \\ &= \sum_{j=-\infty}^{\infty}\sum_{l=0}^{\infty}\gamma_l^{(-1/2-H)}\Phi_H(t-j)\eta_{j-l} \\ &= \sum_{j=-\infty}^{\infty}\Phi_H(t-j)S_j^{(H)}.\end{aligned}$$

5.4 Stochastic integrals with respect to Volterra type processes

In this section introduce stochastic integrals w.r.t. a Volterra type process X on $T = [0,\tau]$ and show how they can be approximated. A stochastic integral of a deterministic integrand with respect to a Gaussian process can be defined as an isometry between $L^2(T)$ and a subspace of $L^2(\Omega)$. There are essentially two different ways to define a stochastic integral of deterministic functions with respect to the Volterra type process X. Either as the isometry which associates $1_{[0,t]} \mapsto X(t)$ or the one which associates $V(t,\cdot) \mapsto X(t)$ (see e.g. Decreusefond [8]). For our purposes it is convenient to define the stochastic integral as the isometry $\mathcal{I}_X$ of $L^2(T)$ onto H given by $\mathcal{I}_X = R^{-1}\circ V$. With this definition we see that $\mathcal{I}_X$ maps $V(t,\cdot) \mapsto X(t)$. One advantage with this approach is that orthogonality relations become straightforward

$$\mathbb{E}\left[X_t|\mathcal{F}_s^X\right] = \mathcal{I}_X(V(t,\cdot)1_{[0,s]}), \quad s<t$$

but contrary to many other stochastic integrals it is *not* the limit of Riemann sums. We have for instance that $\{W_t : t\in T\} \triangleq \{\mathcal{I}_X(1_{[0,t]}) : t\in T\}$ is a Brownian motion with respect to $\{\mathcal{F}_t^X : t\in T\}$. Indeed, W is Gaussian with covariance function

$$\mathbb{E}[W_tW_s] = \langle V(1_{[0,t]}), V(1_{[0,s]})\rangle_{\mathcal{H}} = \langle 1_{[0,t]}, 1_{[0,s]}\rangle_{L^2(T)} = t\wedge s.$$

Moreover, the filtration $\{\mathcal{F}_t^W : t\in T\}$ generated by W coincides with the filtration $\{\mathcal{F}_t^X : t\in T\}$ generated by X. To see this note that span$\{1_{[0,s]} : s\le t\}$ is dense

in $L^2([0,t])$ and hence $\text{span}\{W_s : s \leq t\} = \text{span}\{\mathcal{I}_X(1_{[0,s]}) : s \leq t\}$ is dense in H_t, the closure of $\text{span}\{X_s : s \leq t\}$ in $L^2(\Omega)$. The sigma algebra $\sigma(\xi : \xi \in H_t)$ is equal to $\mathcal{F}_t^X$ and does not change if we restrict the ξ's to a dense subset of H_t. Hence $\mathcal{F}_t^X = \mathcal{F}_t^W$.

It may seem difficult to approximate this stochastic integral since we can not approximate it by Riemann sums. However, we can approximate it using the representation (5.18) (see Proposition 5.29 below). We will also define a stochastic integral for stochastic integrands with respect to X. For our purposes it will be natural to consider a stochastic integral w.r.t. X which has been introduced for the fractional Brownian motion by Decreusefond and Üstünel [9]. We use similar techniques as in [9] with fractional Brownian motion replaced by a Volterra type process.

Another approach to stochastic integration is to consider pathwise stochastic integrals defined ω by ω. This can be done for very general stochastic processes and do not rely on the Gaussian Hilbert space structure used here. A detailed account on these constructions can be found in Dudley and Norvaisa [10].

In order to define the stochastic integral for stochastic integrands we introduce some concepts of Malliavin calculus. For further details we refer to Nualart [24] and Üstünel [29].

5.4.1 Malliavin calculus

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a complete probability space and $\{G(h) : h \in \mathcal{H}\}$ an isonormal Gaussian process on $(\Omega, \mathcal{F}, \mathbb{P})$ (see Definition 1.1.1 in Nualart [24]) indexed by a separable Hilbert space $\mathcal{H}$ that consists of real valued functions on T and equipped with the inner product $\langle \cdot, \cdot \rangle_{\mathcal{H}}$. Let $\mathcal{X}$ be a separable Hilbert space and $F : \Omega \to \mathcal{X}$ be an $\mathcal{X}$-valued functional of the form

$$F(\omega) = f(G(h_1), G(h_2), \ldots, G(h_n))x,$$

where $h_1, h_2, \ldots, h_n \in \mathcal{H}$ and $x \in \mathcal{X}$. If f belongs to the space $C_b^\infty(\mathbb{R}^n)$ of bounded infinitely differentiable functions from $\mathbb{R}^n$ to $\mathbb{R}$ with all partial derivatives bounded, then we call F an *$\mathcal{X}$-valued smooth cylindrical functional* and denote by $\mathcal{S}(\mathcal{X})$ the space of all $\mathcal{X}$-valued smooth cylindrical functionals. If $\mathcal{X} = \mathbb{R}$ we write $\mathcal{S}$ for the space of all real-valued smooth cylindrical functionals. We introduce the derivative of $F \in \mathcal{S}(\mathcal{X})$ as

$$\nabla . F = \sum_{j=1}^{n} \partial_j f(G(h_1), \ldots, G(h_n)) h_j(\cdot) \otimes x.$$

The directional derivative of $F \in \mathcal{S}$ in the direction $h \in \mathcal{H}$ is then $\langle \nabla F, h \rangle_{\mathcal{H}}$ which we sometimes denote by $\nabla_h F$. By Lemma 1.2.1 in Nualart [24] we have the following result.

Lemma 5.21. *If $F \in \mathcal{S}$ and $h \in \mathcal{H}$, then*

$$\mathbb{E}[\langle \nabla F, h\rangle_{\mathcal{H}}] = \mathbb{E}[FG(h)].$$

Applying Lemma 5.21 with F_1F_2 gives us:

Lemma 5.22. *If $F_1, F_2 \in \mathcal{S}$ and $h \in \mathcal{H}$, then*

$$\mathbb{E}[F_2\langle \nabla F_1, h\rangle_{\mathcal{H}}] = \mathbb{E}[-F_1\langle \nabla F_2, h\rangle_{\mathcal{H}}] + \mathbb{E}[F_1F_2G(h)]. \tag{5.22}$$

Furthermore, the derivative operator ∇ is a closable operator (see page 27 in Nualart [24] or Proposition I.1 in Üstünel [29]).

Lemma 5.23. *∇ is a closable operator from $L^p(\Omega)$ to $L^p(\Omega;\mathcal{H})$, $p > 1$.*

This result implies that we can define the L^p-domain of ∇.

Definition 5.24. We say that F is in $\mathrm{Dom}_p(\nabla)$ if there exists a sequence $\{F_n : n \geq 1\}$ of smooth cylindrical functionals such that $F_n \to F$ in $L^p(\Omega)$ and $\{\nabla F_n : n \geq 1\}$ is a Cauchy sequence in $L^p(\Omega;\mathcal{H})$. Then we define

$$\nabla F = \lim_{n\to\infty} \nabla F_n.$$

This extended operator ∇ is called the *Gross-Sobolev derivative.*

We denote by $\mathbb{D}_{p,1}$ the linear space $\mathrm{Dom}_p(\nabla)$ equipped with the norm

$$\|F\|^p_{p,1} = \|F\|^p_{L^p(\Omega)} + \|\nabla F\|^p_{L^p(\Omega;\mathcal{H})}.$$

Starting with $\mathcal{X} = \mathbb{R}$ we can define the derivative of a real valued smooth cylindrical functional F. Then, for $k \geq 2$, taking $\mathcal{X} = \mathcal{H}^{\otimes k-1}$ we define the k-th derivative as $\nabla^{(k)}F = \nabla^{(k-1)}(\nabla F)$ and we can define the spaces $\mathbb{D}_{p,k}$ as the closure of $\mathcal{S}$ with respect to the norm

$$\|F\|^p_{p,k} = \|F\|^p_{L^p(\Omega)} + \sum_{j=1}^{k} \|\nabla^{(j)}F\|^p_{L^p(\Omega;\mathcal{H}^{\otimes j})}.$$

Using the Meyer inequalities (see Üstünel [29] or Decreusefond and Üstünel [9]) one can extend ∇ to a continuous linear operator from $\mathbb{D}_{p,k}$ to $\mathbb{D}_{p,k-1}(\mathcal{H})$ for any $p > 1$, but we will not enter the details here. From the closablility of ∇ we the conclude.

Lemma 5.25. *The relation (5.22) holds for any $F_1, F_2 \in \mathbb{D}_{p,1}$.*

Next we introduce the *divergence operator* δ which is the formal adjoint of ∇.

Definition 5.26. Let $U : \Omega \to \mathcal{H}$. We say that $U \in \mathrm{Dom}_p(\delta)$, if for any $F \in \mathbb{D}_{q,1}$ with $1/p + 1/q = 1$ there exists a constant $c_{p,q}(U)$ such that

$$\mathbb{E}[\langle \nabla F, U\rangle_{\mathcal{H}}] \leq c_{p,q}(U)\|F\|_q,$$

and in this case we define $\delta(U)$ by

$$\mathbb{E}[F\delta(U)] = \mathbb{E}[\langle \nabla F, U\rangle_{\mathcal{H}}] \qquad \text{for each } F \in \mathbb{D}_{q,1}.$$

The element $\delta(U)$ is called the *Skorohod integral* of U.

For $k \geq 1$ we denote by $\mathbb{D}_{p,-k}$ the dual of $\mathbb{D}_{p,k}$. Since ∇ has continuous extension δ also has continuous extension from $\mathbb{D}_{p,-k}(\mathcal{H})$ to $\mathbb{D}_{p,-k-1}$ for any $p > 1$. In this paper we will use the domain $\mathrm{Dom}_2(\delta)$. Next we turn to Skorohod integration of elementary processes. Let

$$U^n(t) = \sum_{j=1}^n F_j h_j(t),$$

where $F_j \in \mathcal{S}$ and $h_j \in \mathcal{H}$ for each $1 \leq j \leq n$. Any process of this type is in $\mathrm{Dom}_2(\delta)$. Indeed, using Lemma 5.25 we get for each $F \in \mathbb{D}_{2,1}$

$$\begin{aligned}
\mathbb{E}[\langle \nabla F, U^n\rangle_{\mathcal{H}}] &= \mathbb{E}[\langle \nabla F, \sum_{j=1}^n F_j h_j\rangle_{\mathcal{H}}] = \sum_{j=1}^n \mathbb{E}[F_j\langle \nabla F, h_j\rangle_{\mathcal{H}}] \\
&= \sum_{j=1}^n \mathbb{E}[-F\langle \nabla F_j, h_j\rangle_{\mathcal{H}} + FF_jG(h_j)] \\
&= \sum_{j=1}^n \mathbb{E}[-F\langle \nabla F_j, h_j\rangle_{\mathcal{H}}] + \sum_{j=1}^n \mathbb{E}[FF_jG(h_j)] \\
&\leq \mathbb{E}[F^2]^{1/2}\Big(\sum_{j=1}^n \mathbb{E}[(\langle \nabla F_j, h_j\rangle_{\mathcal{H}})^2]^{1/2} + \sum_{j=1}^n \mathbb{E}[(F_jG(h_j))^2]^{1/2}\Big) \\
&\leq c(U^n)\mathbb{E}[F^2]^{1/2}
\end{aligned}$$

since $\langle \nabla F_j, h_j\rangle_{\mathcal{H}}$ and $F_jG(h_j)$ are in $L^2(\Omega)$ for each $1 \leq j \leq n$. From this computation we also see that

$$\delta(U^n) = \sum_{j=1}^n F_jG(h_j) - \sum_{j=1}^n \langle \nabla F_j, h_j\rangle_{\mathcal{H}}.$$

5.4.2 A stochastic integral with respect to a Volterra type process

We use the setup described in Section 5.2.1. Let $V : T \times T \to [0, \infty)$ satisfy Hypothesis 1. We take $\Omega = C(T)$ equipped with the Borel σ-field $\mathcal{A}$. The canonical

process $X = \{X_t : t \in T\}$ is defined by $X_t(\omega) = \omega(t)$ and the probability measure $\mathbb{P}$ on $(\Omega, \mathcal{A})$ is the unique measure such that X is centered Gaussian with covariance

$$\rho(t,s) = \int_T V(t,r)V(s,r)\mathrm{d}r.$$

We denote by $\{\mathcal{F}_t^X : t \in T\}$ the natural filtration generated by X. To the stochastic process $\{X_t : t \in T\}$ we can associate the stochastic process $\{X(h) : h \in \mathcal{H}\}$ by $X(h) = R^{-1}(h)$ with R as in Section 5.2.2 and $\mathcal{H} = V(L^2(T))$. Note that for $h_1, h_2 \in \mathcal{H}$ we have $\mathbb{E}[X(h_1)X(h_2)] = \langle h_1, h_2\rangle_{\mathcal{H}}$ and $\{X(h) : h \in \mathcal{H}\}$ is an isonormal Gaussian process. Since $\{W_t : t \in T\} = \{\mathcal{I}_X(1_{[0,t]}) : t \in T\}$ is a Brownian motion we can also consider the process$\{W(f) : f \in L^2(T)\}$ where $W(f) = X(Vf)$. We denote by ∇^X, δ^X, ∇^W and δ^W the Gross-Sobolev derivative and the divergence associated with the Gaussian processes $\{X(h)\}_{h\in\mathcal{H}}$ and $\{W(f) : f \in L^2(T)\}$ respectively. We see that for any $f_1, \ldots, f_n \in L^2(T)$ and any $L^2(T)$-valued square integrable random variable u we have, with $F = f(X(Vf_1), \ldots, X(Vf_n))$, $f \in C_b^\infty(\mathbb{R}^n)$ that

$$\begin{aligned}\mathbb{E}[\langle Vu, \nabla^X F\rangle_{\mathcal{H}}] &= \mathbb{E}[\langle Vu, \sum_{j=1}^n \partial_j f(X(Vf_1), \ldots, X(Vf_n))Vf_j\rangle_{\mathcal{H}}] \\ &= E[\langle u, \sum_{j=1}^n \partial_j f(W(f_1), \ldots, W(f_n))f_j\rangle_{L^2(T)}] \\ &= \mathbb{E}[\langle u, \nabla^W F\rangle_{L^2(T)}].\end{aligned}$$

This implies that $\delta^X(Vu) = \delta^W(u)$ and in particular that $\mathrm{Dom}_2(\delta^X)$ is the image of $\mathrm{Dom}_2(\delta^W)$ under the integral transform V. Similarly $F \in \mathbb{D}_{2,1}^X$ if and only if $F \in \mathbb{D}_{2,1}^W$. Let us define a stochastic integral for stochastic integrands with respect to a Volterra type process X for integrands u such that $Vu \in \mathrm{Dom}_2(\delta^X)$.

Definition 5.27. For a process $\{u(t,\omega) : t \in T, \omega \in \Omega\}$ in $L^2(T \times \Omega)$ such that $Vu \in \mathrm{Dom}_2(\delta^X)$, the *stochastic integral* with respect to X is defined as,

$$\int_T u_s \delta X_s \triangleq \delta^X(Vu).$$

Note that for adapted processes $\{u(t,\omega) : t \in T, \omega \in \Omega\}$ this definition reduces to the Ito integral w.r.t. the standard Brownian motion $\{W_t : t \in T\}$ (see also Decreusefond and Üstünel [9] Theorem 4.8). This stochastic integral is analogous to the one introduced by Alós, Mazet and Nualart in [2] where the authors consider the Volterra process X indexed by $\mathcal{H} = (V^*)^{-1}(L^2(T))$ and the stochastic integral of $u \in (V^*)^{-1}(\mathrm{Dom}_2\,\delta^W)$ with respect to X is given by $\delta^W(V^*u)$. For further details on the construction of different stochastic integrals in the case of fractional Brownian motion we refer to Decreusefond [8]. Using the representation (5.18) we can approximate the stochastic integral $\delta^X(Vu)$. Let us first make a couple of remarks.

Remark 5.28. Recall that the Skorohod integral of an elementary process $U^n(t) = \sum_{j=1}^n F_j V(h_j)(t)$ given in the last section is

$$\delta^X(U^n) = \delta^W(u^n) = \sum_{j=1}^n F_j X(V(h_j)) - \sum_{j=1}^n \langle \nabla^W F_j, h_j \rangle_{L^2(T)}.$$

(i) Note that if the F_j's are deterministic, then $\nabla^W F_j = 0$ for each $1 \leq j \leq n$ and the second sum in the expression for $\delta^X(U^n)$ vanishes.
(ii) If U^n is an elementary and adapted process $U^n(t) = \sum_{j=1}^n F_j V(1_{(t_j, t_{j+1}]})(t)$ with $F_j \in \mathcal{S}$ and F_j is $\mathcal{F}_{t_j}^X$-measurable, then F_j is also $\mathcal{F}_{t_j}^W$-measurable and Proposition 1.2.4 in Nualart [24] implies that $\nabla_t^W F_j = 0$ on $(t_j, T]$. Thus, in this case the second sum in the expression for $\delta^X(U^n)$ vanishes as well.

Proposition 5.29. *Let $\{\psi_j : j = 1, 2, \dots\}$ be an orthonormal basis in $L^2(T)$ and X a Volterra type process with kernel V. Then there exist a sequence $\{\xi_j : j = 1, 2, \dots\}$ of iid $N(0,1)$ random variables such that the following statements hold.*

(1) If $f \in L^2(T)$ is a deterministic function and $f_j = \langle f, \psi_j \rangle_{L^2(T)}$, then

$$\delta^X(Vf) = \mathcal{I}_X(f) = \sum_{j=1}^\infty f_j \xi_j, \quad a.s.$$

(2) If $u \in L^2(T \times \Omega)$ is an adapted process with $Vu \in \mathrm{Dom}_2(\delta^X)$ and $u_j = \langle u, \psi_j \rangle_{L^2(T)}$, then

$$\delta^X(Vu) = \sum_{j=1}^\infty u_j \xi_j, \quad a.s.$$

(3) Let $u \in L^2(T \times \Omega)$ be a process with $Vu \in \mathbb{D}_{2,1}^X(\mathcal{H})$ and $u_j = \langle u, \psi_j \rangle_{L^2(T)}$. Then

$$\delta^X(Vu) = \sum_{j=1}^\infty \left[u_j \xi_j - \langle \nabla_{\Psi_j}^X Vu, \Psi_j \rangle_{\mathcal{H}} \right], \quad a.s.$$

Proof. (1) First we show that the integrals $\delta^X(Vf)$ and $\mathcal{I}_X(f)$ coincide. Since f is deterministic Vf is a deterministic function in $\mathcal{H}$ and it is in $\mathrm{Dom}_2(\delta^X)$ by the same arguments as for an elementary process. By Remark 5.28 (i) we have that

$$\delta^X(Vf) = X(Vf) = R^{-1} \circ V(f) = \mathcal{I}_X(f).$$

Hence the two integrals coincide. Since $\{\psi_j : j = 1, 2, \dots\}$ is an orthonormal basis in $L^2(T)$ it follows that $\Psi_j = V\psi_j$ is an orthonormal basis of the reproducing kernel Hilbert space $\mathcal{H}$ and therefore the sequence $\{\xi_j : j = 1, 2, \dots\}$ defined by

$\xi_j \triangleq R^{-1} \circ V\psi_j$, $j \geq 1$, is an orthonormal basis in H. Hence, $\{\xi_j : j = 1, 2, \dots\}$ is a sequence of iid $N(0,1)$ random variables. Writing the function f as, $f(t) = \sum_{j=1}^{\infty} f_j \psi_j(t)$, where $f_j = \langle f, \psi_j \rangle_{L^2(T)}$ and using the definition of $\mathcal{I}_X$ we find,

$$\mathcal{I}_X(f) = R^{-1} \circ V(f) = R^{-1} \circ V \left(\sum_{j=1}^{\infty} f_j \psi_j \right) = R^{-1} \left(\sum_{j=1}^{\infty} f_j \Psi_j \right) = \sum_{j=1}^{\infty} f_j \xi_j.$$

Since $f \in L^2(T)$ we have that the sequence $\{f_j : j = 1, 2, \dots\}$ is in $l^2(\mathbb{N})$ and $S_n \triangleq \sum_{j=1}^{n} f_j \xi_j$, is a square integrable martingale so convergence follows from the martingale convergence theorem. This proves (1).

(2) First we show the result for an elementary adapted process u^n and then we use a limit argument for general adapted processes u. Suppose that u^n is an elementary adapted process of the form $u^n(t) = \sum_{i=1}^{n} F_i 1_{(t_i, t_{i+1}]}(t)$, where $F_i \in \mathcal{S}$ and F_i is $\mathcal{F}_{t_i}^X$-measurable. Then $U^n = Vu^n$ is of the form $U^n(t) = \sum_{i=1}^{n} F_i (V 1_{(t_i, t_{i+1}]})(t)$ which is in $\mathrm{Dom}_2(\delta^X)$. By Definition 5.26 and Lemma 5.25 we have for any $G \in \mathbb{D}_{2,1}^X$,

$$\mathbb{E}\Big[G\delta^X (V \sum_{i=1}^{n} F_i 1_{(t_i, t_{i+1}]}) \Big] = \mathbb{E}\Big[G\delta^W (\sum_{i=1}^{n} F_i 1_{(t_i, t_{i+1}]}) \Big] \tag{5.23}$$

$$= E\Big[\langle \nabla^W G, \sum_{i=1}^{n} F_i 1_{(t_i, t_{i+1}]} \rangle_{L^2(T)} \Big]$$

$$= \mathbb{E}\Big[\sum_{i=1}^{n} \big(G F_i X(V 1_{(t_i, t_{i+1}]}) - G \langle \nabla^W F_i, 1_{(t_i, t_{i+1}]} \rangle_{L^2(T)} \big) \Big]. \tag{5.24}$$

By (1) we have $X(V 1_{(t_i, t_{i+1}]}) = \delta^X(V 1_{(t_i, t_{i+1}]})$ and since F_i is $\mathcal{F}_{t_i}^X$-measurable we have, as in Remark 5.28 (ii),

$$\mathbb{E}[G \sum_{i=1}^{n} \langle \nabla^W F_i, 1_{(t_i, t_{i+1}]} \rangle_{L^2(T)}] = 0.$$

Inserting this into (5.23) we obtain

$$\mathbb{E}\Big[G\delta^X (V \sum_{i=1}^{n} F_i 1_{(t_i, t_{i+1}]}) \Big] = \mathbb{E}\Big[G \sum_{i=1}^{n} F_i \delta^X (V 1_{(t_i, t_{i+1}]}) \Big]$$

By (1), $\delta^X(V 1_{(t_i, t_{i+1}]}) = \sum_{j=1}^{\infty} f_j^i \xi_j$ where $f_j^i = \int_T 1_{(t_i, t_{i+1}]}(t) \psi_j(t) \mathrm{d}t$ so we get,

$$\delta^X (V \sum_{i=1}^{n} F_i 1_{(t_i, t_{i+1}]}) = \sum_{j=1}^{\infty} \sum_{i=1}^{n} \xi_j \int_T F_i 1_{(t_i, t_{i+1}]}(t) \psi_j(t) \mathrm{d}t = \sum_{j=1}^{\infty} \xi_j u_j^n.$$

Any adapted process in $L^2(\Omega \times T)$ can be approximated by a sequence u^n of simple adapted processes in the norm of $L^2(\Omega \times T)$. Furthermore, $u^n \to u$ in $L^2(\Omega \times$

T) implies that $u_j^n = \langle u^n, \psi_j\rangle_{L^2(T)} \to u_j$ in $L^2(\Omega)$ uniformly in j. Since $Vu \in \mathrm{Dom}_2(\delta^X)$ and δ^X is continuous on $\mathrm{Dom}_2(\delta^X)$ we get that $\delta^X(Vu^n) \to \delta^X(Vu)$ in $L^2(\Omega)$. We can now identify the limit to obtain

$$\delta^X(Vu) = \sum_{j=1}^{\infty} u_j \xi_j,$$

which completes the proof of (2).
(3) First note that we have

$$\mathbb{D}_{2,1}^X(\mathcal{H}) = V(\mathbb{D}_{2,1}^W(L^2(T))) \subset V(\mathrm{Dom}_2(\delta^W)) = \mathrm{Dom}_2(\delta^X)$$

so the integral $\delta^X(Vu)$ exists. Recall that by the proof of (1) we have that $\xi_j = \delta^X(\Psi_j) = X(\Psi_j)$ is standard normal and the ξ_j's are independent. We approximate Vu by $U^n = \sum_{j=1}^n u_j \Psi_j$. Then $U^n \to Vu$ in $\mathbb{D}_{2,1}^X(\mathcal{H})$. Now for any $F \in \mathbb{D}_{2,1}^X$ we have

$$\begin{aligned}
\mathbb{E}[F\delta^X(U^n)] &= \mathbb{E}\Big[\langle \nabla F, \sum_{j=1}^n u_j \Psi_j\rangle_{\mathcal{H}}\Big] \\
&= \mathbb{E}\Big[\sum_{j=1}^n F u_j X(\Psi_j) - \sum_{j=1}^n F\langle \nabla^X u_j, \Psi_j\rangle_{\mathcal{H}}\Big] \\
&= \mathbb{E}\Big[F\sum_{j=1}^n u_j \delta^X(\Psi_j) - F\sum_{j=1}^n \langle \nabla_{\Psi_j}^X U^n, \Psi_j\rangle_{\mathcal{H}}\Big] \\
&= \mathbb{E}\Big[F\sum_{j=1}^n u_j \xi_j - \langle \nabla_{\Psi_j}^X U^n, \Psi_j\rangle_{\mathcal{H}})\Big].
\end{aligned}$$

Hence, $\delta^X(U^n) = \sum_{j=1}^n u_j\xi_j - \langle \nabla_{\Psi_j}^X U^n, \Psi_j\rangle_{\mathcal{H}}$. Since $U^n \to Vu$ in $\mathbb{D}_{2,1}^X(\mathcal{H})$ and δ^X is continuous on $\mathrm{Dom}_2(\delta^X)$ it follows that $\delta^X(U^n) \to \delta^X(Vu)$ in $L^2(\Omega)$. Furthermore $\nabla_{\Psi_j}^X U^n = \nabla_{\Psi_j}^X Vu$ for $j \le n$ so identifying the limit we obtain

$$\delta^X(Vu) = \sum_{j=1}^{\infty} \big[u_j\xi_j - \langle \nabla_{\Psi_j}^X Vu, \Psi_j\rangle_{\mathcal{H}}\big].$$

This completes the proof. □

We have the following consequence of the above definition of the stochastic integral.

Proposition 5.30. *Assume that $u = \{u(t,\omega) : t \in T, \omega \in \Omega\}$ is a process in $L^2(T\times\Omega)$ such that Vu is in $\mathrm{Dom}_2(\delta^X)$ and that u is adapted to $\{\mathcal{F}_t^X : t \in T\}$. Then $M = \{M_t : t \in T\}$ defined by $M_t \triangleq \delta^X(V(u1_{[0,t]}))$ is a square integrable continuous martingale w.r.t. $\{\mathcal{F}_t^X : t \in T\}$ with associated Doob-Meyer process $\langle M\rangle_t = \int_0^t u(s)^2 \mathrm{d}s$.*

See also Theorem 4.7 in Decreusefond and Üstünel [9] in the case of the fractional Brownian motion.

Proof. Since u is adapted we have $\delta^X(V(u1_{[0,t]})) = \delta^W(u1_{[0,t]})$ which coincides with the Itô integral $\int_0^t u(s)\mathrm{d}W_s$. This is a continuous martingale with Doob-Meyer process $\int_0^t u(s)^2\mathrm{d}s$. □

5.5 Applications to parameter estimation

In this section we will study estimation of drift parameters in some stochastic models driven by Volterra type processes. Throughout this section we assume that the index set T is a compact interval $[0,\tau]$. We will start to develop theory for estimation of drift parameters in a fairly general setting using a Girsanov transformation. Then, the following subsections we will apply the theoretical results to some particular models involving the fractional Brownian motion. We start by deriving a Girsanov formula for the Volterra process X.

Theorem 5.31 (Girsanov transformation). *Let $X = \{X_t : t \in [0,\tau]\}$ be a Volterra process on $(\Omega, \mathcal{F}, \mathbb{P}_0)$ with kernel V and let $a = \{a(t,\omega) : t \in [0,\tau], \omega \in \Omega\}$ be a process in $L^2([0,\tau]\times\Omega)$ with Va in $\mathrm{Dom}_2(\delta^X)$ and $a(t,\cdot)$ $\mathcal{F}_t^X$-measurable. Define the probability measure $\mathbb{P}_\theta$, $\theta \in \mathbb{R}$, by*

$$\frac{\mathrm{d}\mathbb{P}_\theta}{\mathrm{d}\mathbb{P}_0}\Big|_{\mathcal{F}_t^X} = \exp\Big(\theta\delta^X(V(a1_{[0,t]})) - \frac{\theta^2}{2}\int_0^t a(s)^2\mathrm{d}s\Big).$$

Then the process $Y = \{Y_t : t \in [0,\tau]\}$ defined by

$$Y(t) = X(t) - \theta\int_0^t V(t,s)a(s)\mathrm{d}s,$$

is a Volterra process with kernel V under $\mathbb{P}_\theta$.

Proof. Since $W = \{\mathcal{I}_X(1_{[0,t]}) : t \in [0,\tau]\}$ is a Brownian motion under $\mathbb{P}_0$ and $\delta^X(V(a1_{[0,t]})) = \int_0^t a(s)\mathrm{d}W_s$ it follows from the Girsanov theorem for Brownian motion (see e.g. Liptser and Shiryayev [20] Theorem 6.3 p. 232) that $\widetilde{W} = \{\widetilde{W}_t : t \in [0,\tau]\}$ given by

$$\widetilde{W}_t = W_t - \theta\int_0^t a(s)\mathrm{d}s, \quad \widetilde{W}_0 = 0,$$

is a Brownian motion under $\mathbb{P}_\theta$. Therefore, we can represent Y as

$$Y_t = \int_0^t V(t,s)\mathrm{d}\widetilde{W}_s,$$

which is a Volterra process with kernel V under $\mathbb{P}_\theta$. □

Let us now consider parameter estimation for processes defined on the positive half-axis $[0,\infty)$. We will assume that the processes are observed on $[0,\tau]$ and use the Girsanov transformation to derive maximum likelihood estimators for the unknown parameter θ.

Suppose that $X = \{X_t : t \in [0,\infty)\}$ is a Volterra type process with kernel V defined on $[0,\infty)\times[0,\infty)$ satisfying Hypothesis 1 on every interval $[0,\tau]\times[0,\tau]$, $0 < \tau < \infty$. We assume that X is defined on the probability space $(\Omega,\mathcal{F})$ and has distribution $\mathbb{P}_0$. Let $a = \{a(t,\omega) : t \geq 0, \omega \in \Omega\}$ be a process such that $\{a(t,\omega) : t \in [0,\tau], \omega \in \Omega\}$ belongs to $L^2([0,\tau]\times\Omega)$ with $Va1_{[0,\tau]}$ in $\mathrm{Dom}_2(\delta^X)$ for every $0 < \tau < \infty$. Suppose further that $a(t,\cdot)$ is $\mathcal{F}_t^X$-measurable. Consider the class of models $\{\mathbb{P}_\theta : \theta \in \Theta\}$, $\Theta \subset \mathbb{R}$, $\operatorname{int}\Theta \neq \emptyset$ given by

$$\frac{\mathrm{d}\mathbb{P}_\theta}{\mathrm{d}\mathbb{P}_0}\Big|_{\mathcal{F}_t^X} = \exp\Big(\theta\delta^X(V(a1_{[0,t]})) - \frac{\theta^2}{2}\int_0^t a(s)^2\mathrm{d}s\Big). \tag{5.25}$$

To simplify notations we will write $A(t,\omega) = (Va(\cdot,\omega))(t) = \int_0^t V(t,s)a(s,\omega)\mathrm{d}s$ throughout this section. By the Girsanov formula given above we have the representation

$$X(t) = \theta A(t) + \widetilde{X}(t)$$

where $\widetilde{X}$ is a Volterra type process with kernel V under $\mathbb{P}_\theta$. Our objective is to estimate θ based on observing $\{X_t : t \in [0,\tau]\}$. Let us briefly comment on this assumption.

From a practical point of view it is unrealistic to assume that the whole trajectory of the process is observed. In practice one would like to obtain an estimator based on discrete observations of the process. In the case of classical diffusion processes this problem has been thoroughly investigated by e.g. Bibby and Sørensen [4] using martingale estimation functions. We will however not follow this approach here.

To construct an estimator of θ we maximize the likelihood (5.25) to get

$$\widehat{\theta}_\tau = \frac{\delta^X(V(a1_{[0,\tau]}))}{\int_0^\tau a(s)^2\mathrm{d}s}. \tag{5.26}$$

Recall that the process W defined by $W_t = \delta^X(V1_{[0,t]})$ is a Brownian motion under $\mathbb{P}_0$ and in particular $\delta^X(Va1_{[0,t]}) = \delta^W(a1_{[0,t]}) = \int_0^t a(s)\mathrm{d}W_s$. Then we see that

$$\widehat{\theta}_\tau = \frac{\int_0^\tau a(s)\mathrm{d}W_s}{\int_0^\tau a(s)^2\mathrm{d}s}. \tag{5.27}$$

Furthermore, the Girsanov transformation can be written as

$$\frac{\mathrm{d}\mathbb{P}_\theta}{\mathrm{d}\mathbb{P}_0}\Big|_{\mathcal{F}_\tau^X} = \exp\Big(\theta\int_0^\tau a(s)\mathrm{d}W_s - \frac{\theta^2}{2}\int_0^\tau a(s)^2\mathrm{d}s\Big)$$

which falls into the framework of the exponential families of stochastic processes (see e.g. Chapter 5 in Küchler and Sørensen [17]). Thus, the properties of the estimator $\widehat{\theta}_\tau$ is derived by straightforward application of the results in Küchler and Sørensen [17]. Denote by M_τ the $\mathbb{P}_0$-martingale $\delta^X(V(a1_{[0,\tau]}))$ with Doob-Meyer process $\langle M\rangle_\tau = \int_0^\tau a(s)^2 \mathrm{d}s$.

Theorem 5.32. *Suppose that X, a, $\widehat{\theta}_\tau$ and M are as above and $\langle M\rangle_\tau \to \infty$ $\mathbb{P}_\theta$-a.s. as $\tau \to \infty$. Then the following statements hold.*

(1) *(Strong consistency)* $\widehat{\theta}_\tau \to \theta \ \mathbb{P}_\theta - a.s$ *as* $\tau \to \infty$.

(2) *(Law of the iterated logarithm)*

$$\limsup_{\tau\to\infty} \frac{\langle M\rangle_\tau |\widehat{\theta}_\tau - \theta|}{(2\log\log\langle M\rangle_\tau)^{1/2}} = 1, \quad \mathbb{P}_\theta\text{-a.s.}$$

Moreover, if a is deterministic, then $\widehat{\theta}_\tau \sim N(\theta, 1/\langle M\rangle_\tau)$.

Proof. (1) By Proposition 5.30 M is a continuous martingale w.r.t. $\mathbb{P}_0$ with associated Doob-Meyer process

$$\langle M\rangle_\tau = \int_0^\tau a(s)^2 \mathrm{d}s.$$

Then $M_\tau - \theta\langle M\rangle_\tau$ is a $\mathbb{P}_\theta$ martingale with the same Doob-Meyer process $\langle M\rangle_\tau$ which tends to infinity $\mathbb{P}_\theta$ almost surely as $\tau \to \infty$, by assumption. From the law of large numbers for martingales it follows that, as $\tau \to \infty$,

$$\frac{M_\tau - \theta\langle M\rangle_\tau}{\langle M\rangle_\tau} \to 0, \quad \text{on } \{\langle M\rangle_\infty = \infty\}.$$

Hence, $\widehat{\theta}_\tau \to \theta$ $\mathbb{P}_\theta$-a.s.
(2) This follows from law of the iterated logarithm for martingales (see e.g. Lepingle [19] Theorem 3).
The last assertion follows since, if a is deterministic, then under $\mathbb{P}_\theta$ we have $M_\tau \sim N(\theta, \langle M\rangle_\tau)$.

□

We can also derive a central limit theorem under some extra assumptions (see Küchler and Sørensen [17]).

Theorem 5.33. *Suppose $\theta \in \operatorname{int}\Theta$ and that there exists an increasing positive non-random function $\varphi_\theta(\tau)$ such that under $\mathbb{P}_\theta$, $\langle M\rangle_\tau/\varphi_\theta(\tau) \to \eta^2(\theta)$ in probability as $\tau \to \infty$, where $\eta^2(\theta)$ is a finite non-negative random variable for which $\mathbb{P}_\theta(\eta^2(\theta) > 0) > 0$. Then, under $\mathbb{P}_\theta$*

$$[\langle M\rangle_\tau^{1/2}(\widehat{\theta}_\tau - \theta), \langle M\rangle_\tau/\varphi_\theta(\tau)] \to N(0,1) \times F_\theta$$

weakly as $\tau \to \infty$ conditionally on $\{\eta^2(\theta) > 0\}$, where F_θ is the conditional distribution of $\eta^2(\theta)$ given $\{\eta^2(\theta) > 0\}$.

Proof. This is a direct application of Theorem 5.2.2 p. 49 in Küchler and Sørensen [17]. □

We denote by $\{\psi_j : j = 1, 2, \dots\}$ an orthonormal basis of $L^2([0,\tau])$ and $\{\Psi_j : j = 1, 2, \dots\}$ the associated basis in the reproducing kernel Hilbert space $\mathcal{H} = V(L^2([0,\tau]))$. Then according to Proposition 5.29, we can rewrite the estimator as

$$\widehat{\theta}_\tau = \frac{\sum_{j=1}^{\infty} a_j \xi_j}{\sum_{j=1}^{\infty} a_j^2} \tag{5.28}$$

giving us a spectral representation of the estimator.

In order to compute $\widehat{\theta}_\tau$ we need to compute the coefficients a_j and ξ_j. The a_j's can be computed simply as $a_j = \langle A, \Psi_j \rangle_{\mathcal{H}}$ but the ξ_j's are more difficult to compute since the process X is typically not in $\mathcal{H}$. The next result solves this problem using the adjoint operator. We denote by V^* the adjoint of V with respect to the $L^2(T)$ inner product, i.e. for $f, g \in L^2(T)$

$$\langle Vf, g \rangle_{L^2(T)} = \langle f, V^* g \rangle_{L^2(T)}$$

and V^{-1*} denotes the adjoint of V^{-1}.

Proposition 5.34. *Let $\{X_t : t \in T\}$ be a Volterra type process with kernel V. Assume that $\{\psi_j : j = 1, 2, \dots\}$ is an orthonormal basis in $L^2(T)$ and $\psi_j \in V^*(L^2(T))$, $j \geq 1$. Then, the coefficients $\{\xi_j : j = 1, 2, \dots\}$ in the representation (5.18) are given by*

$$\xi_j = \int_T X(t)(V^{-1*}\psi_j)(t)\mathrm{d}t, \quad j = 1, 2, \dots \tag{5.29}$$

Proof. By the proof of Theorem 5.16 we have that $\xi_j = R^{-1}\Psi_j(t)$, where $\{\Psi_j : j = 1, 2, \dots\}$ is an orthonormal basis in the reproducing kernel Hilbert space. For processes of the form (5.13) we have that the reproducing kernel Hilbert space is $V(L^2(T))$ and hence $\Psi_j = V\psi_j$ is an orthonormal basis in this space. We need to prove that $R(\xi_j)(t) = \Psi_j(t)$ for ξ_j given by (5.29). Indeed, using integration by parts,

$$\begin{aligned} R(\xi_j)(t) &= \mathbb{E}\left(\xi_j X(t)\right) = \mathbb{E}\left(X(t)\int_T X(s)(V^{-1*}\psi_j)(s)\mathrm{d}s\right) \\ &= \int_T \rho(t,s)(V^{-1*}\psi_j)(s)\mathrm{d}s = \int_T (V^{-1}\rho(t,\cdot))(s)\psi_j(s)\mathrm{d}s \\ &= \int_T V(t,s)\psi_j(s)\mathrm{d}s = (V\psi_j)(t) = \Psi_j(t). \end{aligned}$$

□

In the following subsections we will give two concrete examples where this estimation technique is applied. The first example studies the case of a fractional Brownian motion with deterministic linear drift which was considered by Norros et. al. in [23]. The second example studies the fractional Ornstein-Uhlenbeck type process considered by Kleptsyna and Le Breton in [15]. We will show how the general results derived so far applies in these two cases.

5.5.1 Deterministic drift

Let us consider the family $\{\mathbb{P}_\theta : \theta \in \Theta\}$ of models given by

$$\frac{\mathrm{d}\mathbb{P}_\theta}{\mathrm{d}\mathbb{P}_0}\Big|_{\mathcal{F}_t^X} = \exp\Big(\theta\delta^X(K_H(a1_{[0,t]})) - \frac{\theta^2}{2}\int_0^t a(s)^2\mathrm{d}s\Big) \tag{5.30}$$

where X is the fractional Brownian motion and $A(t) = (K_H(a1_{[0,t]}))(t) = t$. Then the maximum likelihood estimator of θ is given as in (5.27)

$$\widehat{\theta}_\tau = \frac{\delta^X(K_H(a1_{[0,\tau]}))}{\int_0^\tau a(s)^2\mathrm{d}s} = \frac{\int_0^\tau a(s)\mathrm{d}W_s}{\int_0^\tau a(s)^2\mathrm{d}s}. \tag{5.31}$$

Note that using Table 9.1 in Samko et. al. [26] we find that

$$a(s) = (K_H^{-1}A)(s) = \sqrt{V_H}\frac{\Gamma(3/2-H)}{\Gamma(2-2H)}s^{1/2-H},$$

with V_H as in (5.15). Hence a is deterministic and thus $\widehat{\theta}_\tau$ is normally distributed under $\mathbb{P}_\theta$ with mean θ and variance $\langle M\rangle_\tau = \int_0^\tau a(s)^2\mathrm{d}s$. We can explicitly compute the variance as

$$\begin{aligned}\mathrm{Var}_\theta(\widehat{\theta}_\tau) &= \int_0^\tau a(s)^2\mathrm{d}s = \tau^{2-2H}\frac{1}{V_H}\frac{(2-2H)\Gamma(2-2H)^2}{\Gamma(3/2-H)^2}\\ &= \tau^{2-2H}\frac{\pi H(1-2H)(2-2H)\Gamma(2-2H)}{\cos(\pi H)\Gamma(3/2-H)^2}\end{aligned}$$

In Norros et. al. [23] this particular example is studied. We will now show that our approach leads to the same results in this case. Their estimator is given by

$$\overline{\theta}_\tau = \frac{M'_\tau}{\langle M'\rangle_\tau},$$

where $M'_t = \int_0^t k_H(t,s)\mathrm{d}B_s^H$, $k_H(t,s) = c_1 s^{1/2-H}(t-s)^{1/2-H}1_{(0,t)}(s)$ which can be rewritten as (see Norros et. al. [23] p. 584) $\frac{c_H}{2H}\int_0^t s^{1/2-H}\mathrm{d}W_s$ where

$$\begin{aligned}c_1 &= \frac{1}{H\Gamma(3/2-H)\Gamma(H+1/2)},\\ c_H &= \Big(\frac{2H\Gamma(3/2-H)}{\Gamma(H+1/2)\Gamma(2-2H)}\Big)^{1/2}.\end{aligned}$$

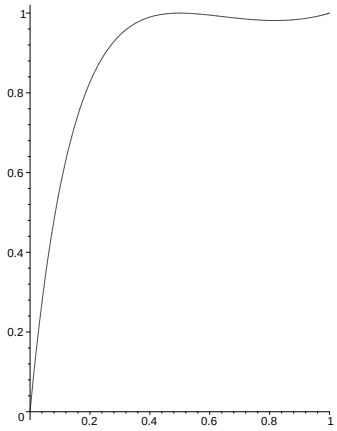

Figure 5.1. The variance of the estimator $\widehat{\theta}_1$.

Since $a(s) = \sqrt{V_H}\frac{\Gamma(3/2-H)}{\Gamma(2-2H)}s^{1/2-H}$ we get that $\frac{c_H}{2H}\int_0^\tau s^{1/2-H}\mathrm{d}W_s = \delta^W(a1_{[0,\tau]}) = \delta^{B^H}(K_H(a1_{[0,\tau]})) = \delta^X(K_H(a1_{[0,\tau]}))$ and then also $\langle M'\rangle_\tau = \langle M\rangle_\tau$. It follows that $\overline{\theta}_\tau = \widehat{\theta}_\tau$.

We can compare this estimator at $\tau = 1$ with the sample mean $X_\tau/\tau = X_1$ which is unbiased and has variance 1 for all $H \in (0,1)$. The variance of the estimator $\widehat{\theta}_\tau$, as a function of H, is illustrated in Figure 5.1. We conclude from this figure that the estimator $\widehat{\theta}_\tau$ has significantly lower variance than the sample mean for small values of H, has the same variance for $H = 1/2$ and has negligible lower variance for $H \in (1/2, 1)$.

5.5.2 The fractional Ornstein-Uhlenbeck type process.

Let us consider the family $\{\mathbb{P}_\theta : \theta \in \Theta\}$ of models given by

$$\frac{\mathrm{d}\mathbb{P}_\theta}{\mathrm{d}\mathbb{P}_0}\Big|_{\mathcal{F}_t^X} = \exp\left(\theta\delta^X(K_H(a1_{[0,t]})) - \frac{\theta^2}{2}\int_0^t a(s)^2\mathrm{d}s\right) \tag{5.32}$$

where X is the fractional Brownian motion and the drift is given by $A(t,\omega) = (K_H a(\cdot,\omega))(t) = \int_0^t X(s,\omega)\mathrm{d}s$. This means that we must take $a(\cdot,\omega) = (K_H^{-1} \circ I^1 X(\cdot,\omega))(\cdot)$. To verify that this choice is possible we need to check that $I^1X(\cdot,\omega) \in \mathcal{H}$ for $\mathbb{P}_\theta$-a.e. $\omega \in \Omega$. For every $\theta \in \Theta$ the measure $\mathbb{P}_\theta$ is absolutely continuous w.r.t. $\mathbb{P}_0$ so it is sufficient to check this for $\theta = 0$, i.e. when X is the fractional Brownian motion. We denote by $\mathrm{Hol}^\alpha([0,\tau])$ the Hölder space of order α over the interval $[0,\tau]$. For every $\alpha < H$ we have $X \in \mathrm{Hol}^\alpha([0,\tau])$ $\mathbb{P}_0$-a.s. so $I^1X \in \mathrm{Hol}^{\alpha+1}([0,\tau])$ $\mathbb{P}_0$-a.s. For $\beta < \alpha+1 < H+1$ we have $\mathrm{Hol}^{\alpha+1}([0,\tau]) \subset I^\beta(L^2([0,\tau]))$

so in particular $I^1 X \in I^{H+1/2}(L^2([0,\tau]))$ $\mathbb{P}_0$-a.s. Hence, the process a is well defined. Using the Girsanov transformation we see that X has the representation

$$X(t) = \theta \int_0^t X(s)\mathrm{d}s + \widetilde{B}^H(t),$$

where $\widetilde{B}^H$ is a fractional Brownian motion under $\mathbb{P}_\theta$.

The maximum likelihood estimator of θ is given as in (5.26)

$$\widehat{\theta}_\tau = \frac{\delta^X(K_H(a1_{[0,\tau]}))}{\int_0^\tau a(s)^2\mathrm{d}s}$$

Let us derive the relation to the results in Le Breton and Kleptsyna [15] where this problem is studied for $H \geq 1/2$ as an extension of Norros et. al. [23]. The authors introduce the process Q defined by

$$\int_0^t Q(s)\mathrm{d}w_s^H = \int_0^t k_H(t,s)X(s)\mathrm{d}s,$$

where

$$w^H(t) = \frac{\Gamma(3/2-H)}{2H\Gamma(3-2H)\Gamma(H+1/2)} t^{2-2H}.$$

The estimator $\overline{\theta}_\tau$ of θ is then derived as

$$\overline{\theta}_\tau = \frac{\int_0^\tau Q(s)\mathrm{d}M'_s}{\int_0^\tau Q^2(s)\mathrm{d}w_s^H},$$

with M' as in Section 5.5.1. To show that the estimators $\widehat{\theta}_\tau$ and $\overline{\theta}_\tau$ coincides it is sufficient to show that $\int_0^\tau Q(s)\mathrm{d}M'_s = \delta^X(K_H(a1_{[0,\tau]}))$. Indeed,

$$\int_0^\tau Q(s)\mathrm{d}M'_s = \int_0^\tau Q(s)s^{1/2-H}\mathrm{d}W_s = \delta^W(Q(s)s^{1/2-H}) = \delta^X(K_H(Q(s)s^{1/2-H}))$$

and using (12) in Kleptsyna, Le Breton and Roubaud [16] we find

$$(K_H Q(s)s^{1/2-H})(t) = \int_0^t K_H(t,s)Q(s)s^{1/2-H}\mathrm{d}s = \int_0^t X(s)\mathrm{d}s.$$

Thus we have established that $\widehat{\theta}_\tau = \overline{\theta}_\tau$. It is proved in Kleptsyna and Le Breton [15] Proposition 2.2 that $\int_0^\tau Q^2(s)\mathrm{d}w_s^H \to \infty$ $\mathbb{P}_\theta$-a.s. and since $\langle M\rangle_\tau = \int_0^\tau a(s)^2\mathrm{d}s = \int_0^\tau Q^2(s)\mathrm{d}w_s^H$ Theorem 5.32 applies and we derive consistency and the law of the iterated logarithm of the estimator. In Kleptsyna and Le Breton [15] the bias and mean square error is also determined using Laplace transforms.

References

[1] Abry, P., Flandrin, P., Taqqu, M. and Veitch, D. (2000) Wavelets for the analysis, estimation and synthesis of scaling data, in: K. Park and W. Willinger, eds., *Self-Similar Network Traffic and Performance Evaluation*, Wiley, New York, 39–88.

[2] Alòs, E., Mazet, O. and Nualart, D. (2000) Stochastic calculus with respect to Gaussian processes, *Ann. Probab.*, Vol. 29 No. 2, 766–801.

[3] Benassi, A., Jaffard S. and Roux, D. (1997) Elliptic Gaussian random processes, *Rev. Mat. Iberoamericana*, Vol. 13 No. 1, 19–90.

[4] Bibby, B.M. and Sørensen, M. (1995) On estimation for discretely observed diffusions: A review, Research Report No. 331, Department of Theoretical Statistics, University of Aarhus.

[5] Carmona, P., Coutin, L. and Montseny, G. (2000) Approximation of some Gaussian processes, *Stat. Inference Stoch. Process.*, Vol. 3, 161–171.

[6] Comte, F. (1996) Simulation and estimation of long memory continuous time models, *J. Time Ser. Anal.*, Vol. 17 No. 1, 19–36.

[7] Comte, F. and Renault, E. (1996) Long memory continuous time models, *J. Econometrics*, Vol. 73, 101–149.

[8] Decreusefond, L. (2000) A Skorohod-Stratonovich integral for the fractional Brownian motion, *Proceedings of the 7-th Workshop on Stochastic Analysis and Related Fields.*

[9] Decreusefond, L. and Üstünel, A.S. (1999) Stochastic Analysis of the Fractional Brownian Motion, *Potential Anal.*, Vol. 10, 177–214.

[10] Dudley, R. and Norvaisa, R. (1998) An introduction to p-variation and Young integrals, Tech. rep. 1, MaPhySto.

[11] Feyel, D. and la Paradelle, A. (1999) On Fractional Brownian Processes, *Potential Anal.*, Vol. 10, 273–288.

[12] Grenander, U, (1981) *Abstract Inference*, John Wiley & Sons.

[13] Janson, S. (1997) *Gaussian Hilbert Spaces*, Cambridge University Press.

[14] Kallianpur, G. (1971) Abstract Wiener Processes and Their Reproducing Kernel Hilbert Spaces, *Z. Wahrsch. Verw. Gebiete*, Vol. 17, 113–123.

[15] Kleptsyna M.L. and Le Breton, A. (2001) Statistical analysis of the fractional Ornstein-Uhlenbeck type process, *Stat. Inference Stoch. Process.*, Vol. 5 No. 3, 229–248.

[16] Kleptsyna, M.L., Le Breton A. and Roubaud, M.-C. (2000) Parameter Estimation and Optimal Filtering for Fractional Type Stochastic Systems, *Stat. Inference Stoch. Process.*, Vol. 3 No. 1-2, 173–182.

[17] Küchler, U. and Sørensen, M. (1997) *Exponential Families of Stochastic Processes*, Springer Verlag.

[18] Ledoux, M. and M. Talagrand, M. (1991) *Probability in Banach Spaces*, Springer Verlag.

[19] Lepingle, D. (1978) Sur le comportement asymptotique des martingales locale, *Séminaire de Probabilités XII, Lecture Notes in Maths.* 649, 148–161.

[20] Liptser, R.S. and Shiryayev, A.N. (1977) *Statistics of Random processes I General Theory*, Springer Verlag.

[21] Meyer, Y., Sellan, F. and Taqqu, M. (1999) Wavelets, Generalized White Noise and Fractional Integration: The Synthesis of Fractional Brownian Motion, *J. Fourier Anal. Appl.*, Vol. 5 No. 5, 465–494.

[22] Norros, I., Mannersalo, P. and Wang, J.L. (1999) Simulation of Fractional Brownian Motion with Conditionalized Random Midpoint Displacement, *Advanced Performance Analysis*, Vol. 2 No. 1, 77–101.

[23] Norros, I., Valkeila, E. and Virtamo, J. (1999) An elementary approach to a Girsanov formula and other analytical results on fractional Brownian motions, *Bernoulli*, Vol. 5 No. 4, 571–588.

[24] Nualart, D. (1995) *The Malliavin Calculus and Related Topics*, Springer Verlag.

[25] Pipiras, V. and Taqqu, M. (2000) Integration questions related to fractional Brownian motion, *Probab. Theory Related Fields*, Vol. 118, 251–291.

[26] Samko, S.G., Kilbas, A.A. and Marichev, O.I. (1987) *Fractional Integrals and Derivatives, Theory and Applications*, Gordon and Breach Science Publishers.

[27] Samorodnitsky, G. and Taqqu, M. (1994) *Stable Non-Gaussian Random Processes*, Chapman and Hall.

[28] Steele, J.M. (2001) *Stochastic Calculus and Financial Applications*, Springer Verlag.

[29] Üstünel, A.S. (1995) An Introduction to Analysis on Wiener Space, *Lecture Notes in Maths.* 1610, Springer Verlag.

Chapter 6

Parameter estimation for discretely observed fractional Ornstein-Uhlenbeck process

Henrik Hult (2003) Parameter estimation for the fractional Ornstein-Uhlenbeck process.

Abstract. The fractional Ornstein-Uhlenbeck process is constructed as a classical Ornstein-Uhlenbeck process with the Brownian noise replaced by a fractional Brownian noise. It is a continuous time stationary Gaussian process and has long memory if the Hurst index in the driving fractional Brownian motion is greater than 1/2. We study the problem of estimating unknown parameters in this model based on discrete observations of the process. The idea of Whittle estimation is well adapted to this particular estimation problem as the covariance function is complicated whereas the spectral density has a simple form. This allows us to derive consistency, asymptotic normality and asymptotic efficiency of our estimator. The theoretical results are illustrated by simulations.

2000 *Mathematics Subject Classification.* 62M09 (primary); 62M15, 60G15 (secondary).
Keywords and phrases. Fractional Brownian motion; Parameter estimation; Whittle estimation; Long-range dependence.
Acknowledgments. The author want to thank Boualem Djehiche for suggesting the topic of this paper and for comments on the manuscript.

6.1 Introduction

Stationary time series with long memory such as the fractional Gaussian noise and the fractionally integrated ARMA processes are currently used as stochastic models in various applications including telecommunications, hydrodynamics and economics. A stationary time series $\{X_j : j \in \mathbb{Z}\}$ is said to have long memory if the sum of correlations diverges, i.e. $\sum_k \rho(k) = \infty$, where $\rho(k) = \mathrm{Cov}(X_{j+k}, X_j)/\mathrm{Var}(X_j)$. In continuous time the fractional Ornstein-Uhlenbeck process is a natural analogue to these models. It is constructed as a classical Ornstein-Uhlenbeck process with the Brownian noise replaced by a fractional Brownian noise. It can be defined the stationary solution to the equation

$$X_t - X_s = -\theta \int_s^t X_u \mathrm{d}u + \sigma(B_t^H - B_s^H), \quad s < t,$$

where $\sigma > 0$, $\theta > 0$, $H \in (0,1)$ and $\{B_t^H : t \in \mathbb{R}\}$ is the fractional Brownian motion with Hurst index H. The increments of the fractional Brownian motion have long memory if $H > 1/2$ and this property is transfered to the fractional Ornstein-Uhlenbeck process. The fractional Ornstein-Uhlenbeck process has recently received some attention in the literature. Cheridito, Kawaguchi and Maejima [3] derives some of its properties and show, contrary to the classical case, that it can not be derived as the stationary process obtained by the Lamperti transformation of fractional Brownian motion for $H \neq 1/2$. Kleptsyna and Le Breton [7] and Hult [6] study the fractional Ornstein-Uhlenbeck process via a Girsanov transformation of the fractional Brownian motion and discuss the use of the Girsanov transformation in the context of parameter estimation.

In this paper we use the following setup. Let $\{X_t : t \in \mathbb{R}\}$ be the fractional Ornstein-Uhlenbeck process with $H > 1/2$ and suppose we have at hand discrete observations $\mathbf{x} = (x_1, \ldots, x_N)'$ of $\{X_{j\Delta} : j = 1, \ldots, N\}$, with $\Delta > 0$ being the time between consecutive observations. Based on these observations we want to estimate all the unknown parameters $\Psi = (\sigma^2, \theta, H)$. Since the process is Gaussian the exact log-likelihood may be explicitly computed as

$$\log L(\mathbf{x}; \Psi) \propto -\frac{1}{2} \log |\Sigma_{\Delta,N}| - \frac{1}{2} \mathbf{x}' \Sigma_{\Delta,N}^{-1} \mathbf{x},$$

where $\Sigma_{\Delta,N}$ is the covariance matrix of $(X_\Delta, \ldots, X_{N\Delta})'$. A problem here is that the expression for the covariance function for the fractional Ornstein-Uhlenbeck process is complicated and the inversion of $\Sigma_{\Delta,N}$ may be computationally demanding if the number of observations N is large. A way around these problems is to use an idea introduced by Whittle [11] and further developed for Gaussian sequences with long memory by Fox and Taqqu [5] and Dahlhaus [4]. This approach is also discussed in some detail in Beran [1]. The idea is to approximate the exact likelihood by replacing $\log |\Sigma_{\Delta,N}|$ by $N(2\pi)^{-1} \int_{-\pi}^{\pi} \log f_\Delta(\lambda; \Psi) \mathrm{d}\lambda$ and $\Sigma_{\Delta,N}^{-1}$ by a matrix A where

$$A_{jk} = (2\pi)^{-2} \int_{-\pi}^{\pi} \frac{1}{f_\Delta(\lambda; \Psi)} e^{i(j-k)\lambda} \mathrm{d}\lambda,$$

and f_Δ is the spectral density for the Gaussian sequence $\{X_{j\Delta} : j \in \mathbb{Z}\}$. In the case of the fractional Ornstein-Uhlenbeck process this spectral density is easily computed and this provides a satisfactory solution to the estimation problem. Using results by Fox and Taqqu [5] and Dahlhaus [4] we see that the derived estimator is consistent, asymptotically normal and asymptotically efficient as $N \to \infty$. The results are also illustrated by simulations.

The paper is organized as follows. In Section 6.2 the fractional Ornstein-Uhlenbeck process is introduced and we derive its spectral density and its covariance function. In Section 6.3 we study the problem of parameter estimation based on discrete observations and show that the asymptotic results derived by Fox and Taqqu [5] and Dahlhaus [4] may be applied. Section 6.4 contains numerical illustrations and simulations.

6.2 The fractional Ornstein-Uhlenbeck process

In this section we introduce the fractional Ornstein-Uhlenbeck process and study some of its properties. The fractional Ornstein-Uhlenbeck is constructed as a classical Ornstein-Uhlenbeck process with the Brownian noise replaced a fractional Brownian noise. The fractional Brownian motion $\{B_t^H : t \in \mathbb{R}\}$ is a continuous Gaussian process with mean zero and covariance function

$$r_{B^H}(t,s) = \frac{1}{2}(|t|^{2H} + |s|^{2H} - |t-s|^{2H}),$$

where $H \in (0,1)$ is called the Hurst index. For $H = 1/2$ is coincides with the standard Brownian motion. The classical Ornstein-Uhlenbeck process is defined as the process $\{Y_t : t \in \mathbb{R}\}$ given by

$$Y_t = \sigma \int_{-\infty}^{t} e^{-\theta(t-u)} \mathrm{d}B_u^{1/2}, \quad t \in \mathbb{R},$$

where $\sigma > 0$, $\theta > 0$. For arbitrary $H \in (0,1)$ the stochastic integral

$$\sigma \int_{-\infty}^{t} e^{-\theta(t-u)} \mathrm{d}B_u^{H}(\omega), \quad t \in \mathbb{R},$$

is well defined as a pathwise Riemann-Stiltjes integral (see Cheridito, Kawaguchi and Maejima [3]) and it is the almost surely continuous process which satisfy the equation

$$X_t(\omega) - X_s(\omega) = -\theta \int_s^t X_u(\omega)\mathrm{d}u + \sigma(B_t^H(\omega) - B_s^H(\omega)), \quad s < t.$$

This motivates the following definition.

Definition 6.1. The fractional Ornstein-Uhlenbeck process $\{X_t : t \in \mathbb{R}\}$ is the stochastic process given by

$$X_t(\omega) = \sigma \int_{-\infty}^{t} e^{-\theta(t-u)} \mathrm{d}B_u^H(\omega), \quad t \in \mathbb{R},$$

where $\sigma > 0$, $\theta > 0$, $H \in (0,1)$ and $\{B_t^H : t \in \mathbb{R}\}$ is the fractional Brownian motion.

For every $t \in \mathbb{R}$ the functions $f_t(u) = \sigma \exp\{-\theta(t-u)\} 1_{\{u \le t\}}$ belongs to the space of integrands (see Pipiras and Taqqu [9])

$$\widetilde{\Lambda}^{H-1/2} = \{f : f \in L^2(\mathbb{R}), \int_{-\infty}^{\infty} |\widehat{f}(\xi)|^2 |\xi|^{1-2H} \mathrm{d}\xi < \infty\},$$

where

$$\widehat{f}(\xi) = \int_{-\infty}^{\infty} e^{iu\xi} f(u) \mathrm{d}u.$$

Indeed,

$$\widehat{f}_t(\xi) = \int_{-\infty}^{t} e^{iu\xi} \sigma e^{-\theta(t-u)} \mathrm{d}u = \sigma \frac{e^{it\xi}}{\theta + i\xi}.$$

Hence the stochastic integrals $I(f_t) = \int_{-\infty}^{\infty} f_t(u) \mathrm{d}B_u^H$ may be consistently defined as limits of elementary functions and $I(f_t)$ is Gaussian with $\mathbb{E}(I(f_t)) = 0$ and

$$\mathbb{E}(I(f_t) I(f_s)) = \frac{C(H)}{2\pi} \int_{-\infty}^{\infty} \widehat{f}_t(\xi) \overline{\widehat{f}_s(\xi)} |\xi|^{1-2H} \mathrm{d}\xi,$$

where

$$C(H) = \Gamma(2H+1) \sin(\pi H).$$

This enables the derivation of the spectral density of the fractional Ornstein-Uhlenbeck process. Indeed,

$$\begin{aligned} r(t+s,s) &= \mathbb{E}(X_{t+s} X_s) = \mathbb{E}(I(f_{t+s}) I(f_s)) \\ &= \frac{C(H)}{2\pi} \int_{-\infty}^{\infty} \widehat{f}_{t+s}(\xi) \overline{\widehat{f}_s(\xi)} |\xi|^{1-2H} \mathrm{d}\xi \\ &= \frac{C(H)}{2\pi} \int_{-\infty}^{\infty} \sigma \frac{e^{i(t+s)\xi}}{\theta + i\xi} \sigma \frac{e^{-is\xi}}{\theta - i\xi} |\xi|^{1-2H} \mathrm{d}\xi \\ &= \int_{-\infty}^{\infty} e^{it\xi} \frac{\sigma^2 C(H)}{2\pi} |\xi|^{1-2H} (\theta^2 + \xi^2)^{-1} \mathrm{d}\xi \end{aligned}$$

Hence, we can identify the spectral density as

$$f(\xi) = \sigma^2 C(H) (2\pi)^{-1} |\xi|^{1-2H} (\theta^2 + \xi^2)^{-1}. \tag{6.1}$$

By inverting the spectral density (see e.g. Oberhettinger [8] p. 6) we obtain an alternative expression for its covariance function

$$r(t) = \sigma^2 C(H)\Big[\frac{1}{2}\theta^{-2H}\sec(\pi(1-2H)/2)\cosh(\theta t) \tag{6.2}$$
$$+ \pi^{-\frac{1}{2}}2^{-2H-1}t^{2H}\Gamma(-H)[\Gamma(H+1/2)]^{-1}\,{}_1F_2\Big(1; H+\frac{1}{2}, H+1, \frac{1}{4}\theta^2t^2\Big)\Big],$$

where ${}_1F_2$ is the Gauss hypergeometric function. We have just proved the following result.

Proposition 6.2. *The spectral density and the covariance function of the fractional Ornstein-Uhlenbeck process are given by* (6.1) *and* (6.2) *respectively.*

The process $\{X_t : t \in \mathbb{R}\}$ can also be given a spectral representation

$$X_t = (\sigma^2 C(H)/2\pi)^{\frac{1}{2}} \int_{-\infty}^{\infty} e^{it\xi}(\theta^2+\xi^2)^{-\frac{1}{2}}|\xi|^{\frac{1}{2}-H}\widetilde{M}(\mathrm{d}\xi) \tag{6.3}$$

where $\widetilde{M}$ is a complex Gaussian measure (see e.g. Samorodnitsky and Taqqu [10]). Indeed, the covariance function is then given by

$$r(t) = \mathbb{E}(X_tX_0) = \sigma^2C(H)(2\pi)^{-1}\int_{-\infty}^{\infty} e^{it\xi}(\theta^2+\xi^2)^{-1}|\xi|^{1-2H}\mathrm{d}\xi = \int_{-\infty}^{\infty} e^{it\xi}f(\xi)\mathrm{d}\xi.$$

The spectral density of the fractional Ornstein-Uhlenbeck process satisfies

$$f(\xi) \sim C(H)\sigma^2(2\pi)^{-1}\theta^{-2}|\xi|^{1-2H}, \quad \text{as } |\xi| \to 0.$$

Hence, using a Tauberian theorem (Bingham, Goldie and Teugels [2] Theorem 4.10.3 p. 240) its covariance function satisfies

$$r(t) \sim \sigma^2H(2H-1)\theta^{-2}|t|^{2H-2}, \quad \text{as } |t| \to \infty.$$

We may conclude that the fractional Ornstein-Uhlenbeck process has long memory for $H \in (1/2, 1)$ since $\sum_k r(k)$ diverges (see also Cheridito, Kawaguchi and Maejima [3]).

An alternative construction of the fractional Ornstein-Uhlenbeck process is to consider a Girsanov transformation of the fractional Brownian motion. If $\{\sigma X_t : t \in \mathbb{R}\}$ is a fractional Brownian motion under the probability measure $\mathbb{P}_0$ then σX is a fractional Ornstein-Uhlenbeck process under the measure $\mathbb{P}_\theta$ given by

$$\frac{\mathrm{d}\,\mathbb{P}_\theta}{\mathrm{d}\,\mathbb{P}_0}\Big|_{\mathcal{F}_t^X} = \exp\Big(\theta M_t - \frac{\theta^2}{2}\langle M\rangle_t\Big),$$

where $M_t = \int_0^t (K_H^{-1} \circ I^1X)(s)\mathrm{d}B(s)$ is a martingale and B is a Brownian motion. Here I^1 denotes the integration operator and K_H is the kernel in the representation of the fractional Brownian motion (see e.g. Hult [6])

$$B_t^H = \int_0^t K_H(t,s)\mathrm{d}B_s.$$

6.3 Parameter estimation based on discrete observations

Let $\{X_t : t \in \mathbb{R}\}$ be a fractional Ornstein-Uhlenbeck process with unknown parameters $\Psi = (\sigma^2, \theta, H)$. We will throughout the rest of the paper assume that $H > 1/2$. Suppose we have made discrete observations $(x_1, \ldots, x_N)$ of the process $\{X_{j\Delta} : j = 1, \ldots, N\}$ with $\Delta > 0$. That is x_j is the observed value of $X_{j\Delta}$, $j = 1, \ldots, N$. Based on these observations we want to estimate the unknown parameters.

This is a natural situation that one encounters in practice. We have a sample of discrete observations and want to estimate parameters for the continuous time stochastic process generating the sample. In the case of diffusion processes this problem have been well studied and one of the fundamental examples in this context appears to be the classical Ornstein-Uhlenbeck process. The Markov-structure of diffusion processes is essential for most of the methods used and therefore studying this problem for non-Markovian models as in this paper seems challenging. On the other hand, for Gaussian processes estimation techniques have been quite extensively developed and it will become apparent that some of these techniques works fine for estimation of the unknown parameters for the fractional Ornstein-Uhlenbeck process.

In analogy to the classical case a natural starting point for estimating unknown parameters would be to use the likelihood given by the Girsanov transformation for estimating the unknown parameters. Let us comment on this approach. First of all the Girsanov transformation gives the likelihood for continuous observations of the process. In practice we have discrete observations so we need to approximate the integrals appearing in the likelihood by discrete analogues. This requires the time between observations, Δ, to be small which is not always the case. Furthermore, the Girsanov transformation gives the likelihood as a function of θ. It can not provide us with insights for estimating H which in most practical situations will be unknown. This differs from the classical case where $H = 1/2$ is known. Hence, although the Girsanov transformation is theoretically interesting, it does not seem very practical to use it in the context of parameter estimation for the fractional Ornstein-Uhlenbeck process.

On the other hand, since the fractional Ornstein-Uhlenbeck process is Gaussian the log-likelihood for the discrete observations $\mathbf{x} = (x_1, \ldots, x_N)'$ of $\{X_{j\Delta} : j = 1, \ldots, N\}$ may be explicitly computed.

$$\log L(\mathbf{x}; \Psi) \propto -\frac{1}{2} \log |\Sigma_{\Delta,N}| - \frac{1}{2} \mathbf{x}' \Sigma_{\Delta,N}^{-1} \mathbf{x},$$

where $(\Sigma_{\Delta,N})_{jk} = r((j-k)\Delta)$. The problem with this approach is the computational effort needed to maximize the exact log-likelihood. First have to compute $\Sigma_{\Delta,N}$ which is complicated since this involves computing the hypergeometric function ${}_1F_2$. Then, $\Sigma_{\Delta,N}$ must be inverted which may be computationally demanding

if the number of observations N is large. A way around the computational problems is to use an approximation of the exact likelihood. Since the spectral density of the fractional Ornstein-Uhlenbeck process is known we may consider the approximation of the exact likelihood suggested by Whittle (see Whittle [11]) and further developed for Gaussian sequences with long memory by Fox and Taqqu in [5] and Dahlhaus [4]. This approach is well suited for the current problem and allows us in a straightforward way to estimate all the unknown parameters. The exact log-likelihood is approximated by replacing $\log|\Sigma_{\Delta,N}|$ by $N(2\pi)^{-1}\int_{-\pi}^{\pi}\log f_\Delta(\lambda;\Psi)\mathrm{d}\lambda$ and $\Sigma_{\Delta,N}^{-1}$ by the matrix A where

$$A_{jk}=(2\pi)^{-2}\int_{-\pi}^{\pi}\frac{1}{f_\Delta(\lambda;\Psi)}e^{i(j-k)\lambda}\mathrm{d}\lambda.$$

Here f_Δ denotes the spectral density for the sequence $\{X_{j\Delta}: j\in\mathbb{Z}\}$. It is given by

$$f_\Delta(\lambda)=\frac{\sigma^2C(H)\Delta^{2H}}{2\pi}\sum_{k=-\infty}^{\infty}\frac{|\lambda+2\pi k|^{1-2H}}{(\Delta\theta)^2+(\lambda+2\pi k)^2}.$$

Indeed, using the spectral representation (6.3) we have

$$\begin{aligned}
r(j\Delta)=\mathbb{E}(X_{j\Delta}X_0)&=\sigma^2C(H)(2\pi)^{-1}\int_{-\infty}^{\infty}e^{ij\Delta\xi}\frac{|\xi|^{1-2H}}{\theta^2+\xi^2}\mathrm{d}\xi\\
&=\sigma^2C(H)(2\pi)^{-1}\sum_{k=-\infty}^{\infty}\int_{(-\pi+2\pi k)/\Delta}^{(\pi+2\pi k)/\Delta}e^{ij\Delta\xi}\frac{|\xi|^{1-2H}}{\theta^2+\xi^2}\mathrm{d}\xi\\
&=\sigma^2C(H)(2\pi)^{-1}\int_{-\pi/\Delta}^{\pi/\Delta}\sum_{k=-\infty}^{\infty}e^{ij\Delta(\xi+2\pi k/\Delta)}\frac{|\xi+2\pi k/\Delta|^{1-2H}}{\theta^2+(\xi+2\pi k/\Delta)^2}\mathrm{d}\xi\\
&=\sigma^2C(H)(2\pi)^{-1}\Delta^{2H}\int_{-\pi}^{\pi}e^{ij\lambda}\sum_{k=-\infty}^{\infty}\frac{|\lambda+2\pi k|^{1-2H}}{(\Delta\theta)^2+(\lambda+2\pi k)^2}\mathrm{d}\lambda.
\end{aligned}$$

Next we parameterize the model so that $f_\Delta(\lambda)=\sigma_*^2f_*(\lambda)$ with

$$\begin{aligned}
f_*(\lambda)&=\frac{K(\theta,H)}{\sigma^2C(H)}f_\Delta(\lambda),\quad \sigma_*^2=\sigma^2\frac{C(H)}{K(\theta,H)}\\
K(\theta,H)&=\sigma^2C(H)\exp\Big(-(2\pi)^{-1}\int_{-\pi}^{\pi}\log f_\Delta(\lambda)\mathrm{d}\lambda\Big)
\end{aligned}$$

Then, $\int_{-\pi}^{\pi}\log f_*(\lambda)\mathrm{d}\lambda=0$ and the approximative log-likelihood becomes

$$\log L^W(\mathbf{x},\Psi)\propto-\frac{N}{2}\log\sigma_*^2-\frac{1}{2\sigma_*^2}\mathbf{x}A(\theta,H)\mathbf{x}.$$

To maximize the approximate log-likelihood we see that it is sufficient to minimize

$$\sigma_N^2(\theta,H)=\frac{\mathbf{x}'A(\theta,H)\mathbf{x}}{N}$$

with respect to θ, H to find the estimators $\hat{\theta}, \hat{H}$ and then put $\hat{\sigma}_*^2 = \sigma_N^2(\hat{\theta}, \hat{H})$. For computational purposes we note that

$$\sigma_N^2(\theta, H) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{1}{f_*(\lambda; \theta, H)} I_N(\lambda) \mathrm{d}\lambda$$

where $I_N(\lambda) = |\sum_{j=1}^N e^{ij\lambda} X_{j\Delta}|^2/(2\pi N)$. Furthermore, the asymptotic results will hold if we replace the integral by a discrete sum (see e.g. Dahlhaus [4])

$$\sigma_N^2(\theta, H) = \sum_{k=-\lceil N/2 \rceil}^{\lfloor N/2 \rfloor} \frac{1}{f_*(2\pi k/N; \theta, H)} I_N(2\pi k/N).$$

Next we give a result on the consistency and asymptotic normality of the estimators.

Theorem 6.3. *Let $\Psi^* = (\Psi_1^*, \Psi_2^*, \Psi_3^*) = (\sigma_*^2, \theta, H)$. The estimators $\hat{\sigma}_*^2(N)$, $\hat{\theta}(N)$ and $\hat{H}(N)$ for the unknown parameters of the fractional Ornstein-Uhlenbeck process satisfies with probability one*

$$\lim_{N\to\infty} \hat{\sigma}_*^2(N) = \sigma_*^2, \ \lim_{N\to\infty} \hat{\theta}(N) = \theta \text{ and } \lim_{N\to\infty} \hat{H}(N) = H.$$

Moreover, the vector $\sqrt{N}(\hat{\Psi}^*(N) - \Psi^*)$ converges in distribution to a normal law with mean zero and covariance $4\pi W^{-1}(\Psi^*)$ where,

$$W(\theta, H)_{jk} = \int_{-\pi}^{\pi} f_\Delta(\lambda; \Psi^*) \frac{\partial^2}{\partial \Psi_j^* \partial \Psi_k^*} \frac{1}{f_\Delta(\lambda; \Psi^*)} \mathrm{d}\lambda.$$

Finally, the estimator $\hat{\Psi}^*(N)$ is an asymptotically efficient estimate of Ψ^* in the sense of Fisher.

Proof. The consistency follows from Theorem 1 in Fox and Taqqu [5], asymptotic normality from Theorem 2.1 in Dahlhaus [4] and the asymptotic efficiency from Theorem 4.1 in Dahlhaus [4] if we verify the conditions (A1)-(A7) in Dahlhaus [4]. To verify conditions (A1)-(A6) we may apply the arguments of Lemma 6 in Fox and Taqqu [5] and instead verify the following conditions. For each $\delta > 0$ there exist constants $C_0(\delta) > 0$ and $C(\delta) > 0$ such that

(B.1) $f_\Delta(\lambda; \Psi^*)$ is continuous at all (λ, Ψ^*), $\lambda \neq 0$, and

$$f_\Delta(\lambda; \Psi^*) \geq C_0(\delta) |\lambda|^{1-2H+\delta}$$

(B.2) $f_\Delta(\lambda; \Psi^*) \leq C(\delta) |\lambda|^{1-2H-\delta}$

(B.3) $\frac{\partial}{\partial\Psi_j^*}f_\Delta(\lambda;\Psi^*)$, $\frac{\partial^2}{\partial\Psi_j^*\partial\Psi_k^*}f_\Delta(\lambda;\Psi^*)$ and $\frac{\partial^3}{\partial\Psi_j^*\partial\Psi_k^*\partial\Psi_l^*}f_\Delta(\lambda;\Psi^*)$ are continuous at all (λ,Ψ^*), $\lambda\neq 0$,

$$\Big|\frac{\partial}{\partial\Psi_j^*}f_\Delta(\lambda;\Psi^*)\Big| \le C(\delta)|\lambda|^{1-2H-\delta},$$
$$\Big|\frac{\partial^2}{\partial\Psi_j^*\partial\Psi_k^*}f_\Delta(\lambda;\Psi^*)\Big| \le C(\delta)|\lambda|^{1-2H-\delta} \text{ and}$$
$$\Big|\frac{\partial^3}{\partial\Psi_j^*\partial\Psi_k^*\partial\Psi_l^*}f_\Delta(\lambda;\Psi^*)\Big| \le C(\delta)|\lambda|^{1-2H-\delta}$$

(B.4) $\frac{\partial}{\partial\lambda}f_\Delta(\lambda;\Psi^*)$, $\frac{\partial^2}{\partial\lambda\partial\Psi_j^*}f_\Delta(\lambda;\Psi^*)$ and $\frac{\partial^3}{\partial\lambda^2\partial\Psi_j^*}f_\Delta(\lambda;\Psi^*)$ are continuous at all (λ,Ψ^*), $\lambda\neq 0$,

$$\Big|\frac{\partial}{\partial\lambda}f_\Delta(\lambda;\Psi^*)\Big| \le C(\delta)|\lambda|^{-2H-\delta},$$
$$\Big|\frac{\partial^2}{\partial\lambda\partial\Psi_j^*}f_\Delta(\lambda;\Psi^*)\Big| \le C(\delta)|\lambda|^{-2H-\delta} \text{ and}$$
$$\Big|\frac{\partial^3}{\partial\lambda^2\partial\Psi_j^*}f_\Delta(\lambda;\Psi^*)\Big| \le C(\delta)|\lambda|^{-1-2H-\delta}$$

The only difference from Lemma 6 in Fox and Taqqu [5] is the condition on the third derivative appearing in (B.3) which is sufficient for verifying the third derivative condition in (A3) in Dahlhaus [4]. To verify these conditions we write $f_\Delta(\lambda;\Psi^*) = \sigma_*^2 K(\theta,H) f_1(\lambda;\theta,H)$ with

$$f_1(\lambda;\Psi^*) = \frac{\Delta^{2H}}{2\pi}\frac{|\lambda|^{1-2H}}{(\Delta\theta)^2+\lambda^2} + \frac{\Delta^{2H}}{2\pi}\sum_{k\neq 0} g_k(\lambda;\Psi^*),$$
$$g_k(\lambda;\Psi^*) = \frac{|\lambda+2\pi k|^{1-2H}}{(\Delta\theta)^2+(\lambda+2\pi k)^2}.$$

Then

$$K(\theta,H) = \exp\Big(-(2\pi)^{-1}\int_{-\pi}^{\pi}\log f_1(\lambda)\mathrm{d}\lambda\Big)$$

and if we verify the conditions (B.1)-(B.4) for f_1 then by the arguments of Lemma 6 in Fox and Taqqu [5] we find that $\int_{-\pi}^{\pi}\log f_\Delta(\lambda)\mathrm{d}\lambda$ is three times continuously differentiable in (θ,H) under the integral sign. Hence $K(\theta,H)$ is also three times continuously differentiable and the conditions are satisfied for f_Δ. A standard theorem on differentiation of series shows that the sum $\sum_{k\neq 0} g_k(\lambda;\Psi^*)$ is three times continuously differentiable at all (λ,Ψ^*). To prove the conditions are satisfied for $\Delta^{2H}(2\pi)^{-1}|\lambda|^{1-2H}((\Delta\theta)^2+\lambda^2)^{-1}$ is elementary. It remains to show condition (A.7) in Dahlhaus [4], i.e.

(A.7) $f_\Delta^{-1}(\lambda;\Psi^*)\frac{\partial}{\partial\lambda}f_\Delta(\lambda;\Psi^*)$ and $\frac{\partial^2}{\partial\lambda^2}f_\Delta^{-1}(\lambda;\Psi^*)$ are continuous at all (λ,Ψ^*), $\lambda\neq 0$ and

$$\Big|\frac{\partial^k}{\partial\lambda^k}f_\Delta^{-1}(\lambda;\Psi^*)\Big|\leq C(\delta)|\lambda|^{2H-1-k-\delta},\quad \text{for } k=0,1,2.$$

Again, by the arguments of Lemma 6 in Fox and Taqqu [5] this follows if we verify that

$$\Big|\frac{\partial^k}{\partial\lambda^k}f_1(\lambda;\Psi^*)\Big|\leq C(\delta)|\lambda|^{1-2H-k-\delta},\quad \text{for } k=0,1,2.$$

Since $\sum_{k\neq 0}g_k(\lambda;\Psi^*)$ is three times continuously differentiable at all (λ,Ψ^*) we only need to prove this for $\Delta^{2H}(2\pi)^{-1}|\lambda|^{1-2H}((\Delta\theta)^2+\lambda^2)^{-1}$ which is elementary. □

6.4 Numerical illustrations

To illustrate the performance of the estimator we have simulated M trajectories of the fractional Ornstein-Uhlenbeck process and recorded the mean and the standard deviation of the estimated parameter values as well as the correlation matrix. To simulate the fractional Ornstein-Uhlenbeck process we have used the spectral representation and the fast Fourier transform. The results are illustrated in the tables below. For each simulation we have $\sigma^2=1$, $\theta=1$ and $\Delta=1$.

	mean	stddev	correlation		
$\hat{\sigma}_*^2(N)$	0.062	0.0087	1	-0.06	0.08
$\hat{\theta}(N)$	1.42	0.71	-0.06	1	0.84
$\hat{H}(N)$	0.62	0.12	0.08	0.84	1

Table 6.1. $\sigma_*^2=0.065$, $\theta=1$, $H=0.6$, $\Delta=1$, $N=100$, M = 50.

	mean	stddev	correlation		
$\hat{\sigma}_*^2(N)$	0.053	0.0090	1	0.21	0.23
$\hat{\theta}(N)$	1.25	0.44	0.21	1	0.71
$\hat{H}(N)$	0.73	0.13	0.23	0.71	1

Table 6.2. $\sigma_*^2=0.053$, $\theta=1$, $H=0.75$, $\Delta=1$, $N=100$, M = 50.

	mean	stddev	correlation		
$\hat{\sigma}_*^2(N)$	0.026	0.0034	1	0.12	0.05
$\hat{\theta}(N)$	0.96	0.34	0.12	1	0.72
$\hat{H}(N)$	0.82	0.10	0.05	0.72	1

Table 6.3. $\sigma_*^2 = 0.027$, $\theta = 1$, $H = 0.9$, $\Delta = 1$, $N = 100$, M = 50.

	mean	stddev	correlation		
$\hat{\sigma}_*^2(N)$	0.065	0.0046	1	0.08	0.16
$\hat{\theta}(N)$	1.21	0.35	0.08	1	0.90
$\hat{H}(N)$	0.63	0.07	0.16	0.90	1

Table 6.4. $\sigma_*^2 = 0.065$, $\theta = 1$, $H = 0.6$, $\Delta = 1$, $N = 500$, M = 50.

	mean	stddev	correlation		
$\hat{\sigma}_*^2(N)$	0.053	0.0038	1	-0.14	-0.02
$\hat{\theta}(N)$	1.18	0.37	-0.14	1	0.91
$\hat{H}(N)$	0.77	0.09	-0.02	0.91	1

Table 6.5. $\sigma_*^2 = 0.053$, $\theta = 1$, $H = 0.75$, $\Delta = 1$, $N = 500$, M = 50.

	mean	stddev	correlation		
$\hat{\sigma}_*^2(N)$	0.027	0.0016	1	-0.13	-0.19
$\hat{\theta}(N)$	1.02	0.26	-0.13	1	0.89
$\hat{H}(N)$	0.89	0.08	-0.19	0.89	1

Table 6.6. $\sigma_*^2 = 0.027$, $\theta = 1$, $H = 0.9$, $\Delta = 1$, $N = 500$, M = 50.

References

[1] Beran, J. (1994) *Statistics for long-memory processes*, Chapman & Hall, New York.

[2] Bingham, N.H., Goldie, C.M. and Teugels, J.L. (1987) *Regular variation*, Cambridge University Press.

[3] Cheridito, P. Kawaguchi, H. and Maejima, M. (2003) Fractional Ornstein-Uhlenbeck processes, *Electron. J. Probab.*, Vol. 8, 1–14.

[4] Dahlhaus, R. (1989) Efficient parameter estimation for self-similar processes, *Ann. Statist.*, Vol. 17 No. 4, 1749–1766.

[5] Fox, R. and Taqqu, M. (1986) Large-sample properties of parameter estimates for strongly dependent stationary Gaussian time series, *Ann. Statist.*, Vol. 14 No. 2, 517–532.

[6] Hult, H. (2003) Approximating some Volterra-type stochastic integrals with applications to parameter estimation, *Stochastic Process. Appl.*, Vol. 105, 1–32.

[7] Kleptsyna, M.L. and Le Breton A. (2002) Statistical inference for the fractional Ornstein-Uhlenbeck process, *Stat. Inference Stoch. Process.*, Vol. 5 No. 3, 229–248.

[8] Oberhettinger, F. (1957) *Tabellen zur Fourier transformen*, Springer Verlag, Berlin.

[9] Pipiras, V. and Taqqu, M. (2000) Integration questions related to fractional Brownian motion, *Probab. Theory Related Fields*, Vol. 118, 251–291.

[10] Samorodnitsky, G. and Taqqu, M. (1994) *Stable non-Gaussian random processes*, Chapman & Hall, New York.

[11] Whittle, P. (1951) *Hypothesis testing in time series analysis*, Hafner, New York.