J. Appl. Prob. **51,** 1133–1153 (2014)
Printed in England

MARKOV TAIL CHAINS

A. JANSSEN,* *University of Hamburg*
J. SEGERS,** *Université Catholique de Louvain*

Abstract

The extremes of a univariate Markov chain with regularly varying stationary marginal distribution and asymptotically linear behavior are known to exhibit a multiplicative random walk structure called the tail chain. In this paper we extend this fact to Markov chains with multivariate regularly varying marginal distributions in $\mathbb{R}^d$. We analyze both the forward and the backward tail process and show that they mutually determine each other through a kind of adjoint relation. In a broader setting, we will show that even for non-Markovian underlying processes a Markovian forward tail chain always implies that the backward tail chain is also Markovian. We analyze the resulting class of limiting processes in detail. Applications of the theory yield the asymptotic distribution of both the past and the future of univariate and multivariate stochastic difference equations conditioned on an extreme event.

Keywords: Autoregressive conditional heteroskedasticity; extreme value distribution; (multivariate) Markov chain; multivariate regular variation; random walk; stochastic difference equation; tail chain; tail-switching potential

2010 Mathematics Subject Classification: Primary 60G70; 60J05
Secondary 60G10; 60H25; 62P05

1. Introduction

Consider a discrete-time, $\mathbb{R}^d$-valued random process $\{X_t : t = 0, 1, 2, \ldots\}$ defined by the recursive equation

$$X_t = \Phi(X_{t-1}, \varepsilon_t), \qquad t = 1, 2, \ldots, \tag{1.1}$$

where

(C1) $\varepsilon_1, \varepsilon_2, \ldots$ are independent and identically distributed (i.i.d.) random elements of a measurable space $(\mathbb{E}, \mathcal{E})$ and independent of X_0, and

(C2) Φ is a measurable function from $\mathbb{R}^d \times \mathbb{E}$ to $\mathbb{R}^d$.

If the process $\{X_t\}$ happens to be stationary, it will be assumed to be defined for all integer t. The distribution of X_0 is assumed to be multivariate regularly varying.

The aim of this paper is to analyze the special structure of weak limits of the finite-dimensional distributions of the process conditional on $\|X_0\|$ being large, where $\|\cdot\|$ denotes the Euclidean norm. More precisely, we will investigate the weak limits, called the forward tail chain, of vectors of the form $(X_0, \ldots, X_t)$ given that $\|X_0\|$ exceeds a high threshold. In addition, if the process is stationary we will extend this to find the so-called back-and-forth

Received 2 May 2013; revision received 3 February 2014.
* Postal address: Department of Mathematics, University of Hamburg, Bundesstrasse 55, 20146 Hamburg, Germany. Email address: anja.janssen@math.uni-hamburg.de
** Postal address: Institut de Statistique, Université Catholique de Louvain, Voie du Roman Pays 20, B-1348 Louvain-la-Neuve, Belgium. Email address: johan.segers@uclouvain.be

tail chain, which corresponds to the weak limits of vectors of the form $(X_{-s}, \dots, X_t)$ given that $\|X_0\|$ is large. A close relation of these processes to multivariate regular variation of the whole process has been analyzed in Basrak and Segers (2009). In this paper we are interested in the special form of the processes, in particular the Markovian structure of both the forward and the backward process and how they necessarily determine each other.

The process $\{X_t\}$ is obviously a discrete-time homogeneous Markov chain. On the other hand, every homogeneous discrete-time Markov chain $\{X_t\}$ on $\mathbb{R}^d$ can be represented as in (1.1), (C1), and (C2), as shown by Kifer (1986). Of course, for a given Markov chain $\{X_t\}$ the above representation is not unique. The way in which Markov chains are defined is often through a recursive equation; all the examples in Goldie (1991, pp. 126–127), for instance, are of this type. The chain is stationary if and only if the random vectors $X_1 = \Phi(X_0, \varepsilon_1)$ and X_0 are equal in law.

In Smith (1992) and Perfekt (1994), excursions of a univariate Markov chain over a high threshold following an extreme event are shown to behave asymptotically and under quite general conditions as a (multiplicative) random walk. The theory has been extended to multivariate Markov chains in Perfekt (1997) and to higher-order Markov chains in Yun (1998), (2000). More recently, Resnick and Zeber (2013) analyzed the topic with a focus on the convergence of Markov kernels, adding a criterion to distinguish between extreme and nonextreme states of a Markov chain as the threshold rises. The random walk representation is useful from a statistical perspective because it gives an understanding of how to model the extremes of certain time series (Bortot and Coles (2000), Coles *et al.* (1994), Smith *et al.* (1997)). A useful and widely investigated class of processes, for which the random walk structure is quite revealing, are the stationary solutions to certain stochastic difference equations, including squared (generalized) autoregressive conditionally heteroskedastic (ARCH/GARCH) processes as a special case (see Basrak *et al.* (2002b), Gomes *et al.* (2004), and de Haan *et al.* (1989)).

One limitation of the theory of Smith (1992), Perfekt (1994), and Resnick and Zeber (2013) is that it is specialized to univariate, nonnegative Markov chains. Similarly, Perfekt (1997) considers only the upper extremes of a multivariate Markov chain. When extending the theory to real-valued and higher-dimensional chains, we have to keep in mind that the extremes may be both positive or negative and that extreme values of X_t may depend not only on $\|X_{t-1}\|$ but also on $X_{t-1}/\|X_{t-1}\|$. We will focus on the simplest case of the extension, which deals with real-valued univariate Markov chains, where an extreme value of X_t may depend on the sign of X_{t-1}. This can be seen, for instance, in the time series of logreturns of prices of financial securities in periods of high volatility. The observation of this so-called leverage effect has lead to the formulation of asymmetric extensions of GARCH models (compare, for example, Zivot (2009)). For such Markov chains with a tail-switching potential, the random walk representation of excursions over high thresholds breaks down in the sense that the distribution of the multiplicative increment now depends in general on the sign of the chain in the previous step. In Bortot and Coles (2003), a more general representation is postulated that involves four transition mechanisms rather than one, corresponding to the four cases of transitions to and from upper or lower extreme states.

The novelty of this paper is two-fold: firstly, to explicitly state the random walk representation in the general $\mathbb{R}^d$-valued case; and, secondly, in the stationary case, to study the joint distribution of the forward and backward tail chain, coined the *back-and-forth tail chain*. Throughout, some remarkable simplifications in the (univariate) real-valued case will be studied in more detail. In particular, in the univariate case the backward tail chain is again a random walk which is, in some sense, dual the forward tail chain. Besides the assumption that the distribution of X_0 is regularly varying, the only condition is a relatively easy-to-check statement on the asymptotic behavior of $\Phi(x, \cdot)$ for large $\|x\|$.

The outline of the paper is as follows. The forward tail chain of a possibly nonstationary $\mathbb{R}^d$-valued Markov chain is studied in Section 2. In Section 3 we examine the backward tail chain, where, for stationary Markov chains, the tail chain can be extended backwards in time. Section 4 describes a kind of adjoint relation between distributions which is motivated by a general property of tail processes of stationary processes. In Section 5 we show that a certain class of processes, coined back-and-forth tail chains, which are derived from this adjoint distribution, form exactly the class of tail processes which arise in our Markovian setting. Finally, Section 6 provides some examples of the theory, including an application to stationary solutions of (multivariate) stochastic difference equations.

To conclude this section, let us fix some notation. We write $(x)_+ = \max(x, 0)$ for the positive part of $x \in \mathbb{R}$ and $(x)_- = \min(x, 0)$ for the negative part. The transpose of a matrix A is denoted by $A^\top$. The law of a random vector X is denoted by $\mathcal{L}(X)$ and weak convergence of probability measures is denoted by $\Rightarrow$. The probability measure degenerate at a point x is denoted by δ_x, and $\mathrm{Unif}(E)$ denotes the uniform distribution on a compact set E. The indicator of an event A is denoted by $\mathbf{1}_A(\cdot)$. We write $\bar{\mathbb{R}}$ for $\mathbb{R} \cup \{-\infty, \infty\}$, $\mathbb{S}^{d-1}$ for $\{x \in \mathbb{R}^d : \|x\| = 1\}$, and 0 for a vector (of suitable dimension) which consists of all 0s. Let $\mathbb{Z}$ be the set of integers and $\mathbb{N}_0$ be the set of nonnegative integers.

2. Forward tail chains

Let $X_0, X_1, X_2, \ldots$ be a homogeneous Markov chain as in (1.1), (C1), and (C2), but not necessarily stationary. The focus of this section is on the weak limits of the finite-dimensional distributions of the process conditionally on $\|X_0\|$ being large (Theorem 2.1). Two conditions are required: Condition 2.1 on the tails of X_0, and Condition 2.2 on the asymptotics of $x \mapsto \Phi(x, e)$ for large $\|x\|$. See, for instance, Resnick (2007) for details on multivariate regular variation.

Condition 2.1. *The distribution of X_0 is multivariate regularly varying on $\bar{\mathbb{R}}^d \setminus \{0\}$, that is, there exists a nondegenerate probability measure Υ on $\mathbb{S}^{d-1}$ (called the spectral measure) and an $\alpha > 0$ such that*

$$\lim_{x\to\infty} \mathbb{P}\bigg(\|X_0\| > ux, \frac{X_0}{\|X_0\|} \in S \mid \|X_0\| > x\bigg) = u^{-\alpha}\Upsilon(S)$$

for all Borel sets $S \subset \mathbb{S}^{d-1}$ which satisfy $\Upsilon(\partial S) = 0$ and $u \geq 1$.

The second condition states that the function Φ in (1.1) is asymptotically homogeneous in x for large values of $\|x\|$.

Condition 2.2. *There exists a measurable map $\phi \colon \mathbb{S}^{d-1} \times \mathbb{E} \mapsto \mathbb{R}^d$ such that, for all $e \in \mathbb{E}$,*

$$\lim_{x\to\infty} x^{-1}\Phi(xs(x), e) = \phi(s, e) \quad \text{whenever } s(x) \to s \text{ in } \mathbb{S}^{d-1}. \tag{2.1}$$

Moreover, if $\mathbb{P}(\phi(s, \varepsilon_1) = 0) > 0$ for some $s \in \mathbb{S}^{d-1}$ then $\mathbb{P}(\varepsilon_1 \in \mathbb{W}) = 1$ also, where $\mathbb{W}$ is a measurable subset of $\mathbb{E}$ such that, for all $e \in \mathbb{W}$,

$$\sup_{\|y\|\leq x} \|\Phi(y, e)\| = O(x), \qquad x \to \infty. \tag{2.2}$$

https://doi.org/10.1239/jap/1421763332 Published online by Cambridge University Press

We extend the domain of the limit function ϕ in (2.1) to $\mathbb{R}^d \times \mathbb{E}$ by setting

$$\phi(v, e) = \begin{cases} \|v\|\phi\big(\frac{v}{\|v\|}, e\big) & \text{if } v \neq 0, \\ 0 & \text{if } v = 0. \end{cases} \tag{2.3}$$

Lemma 2.1. *If Condition 2.2 holds then*

$$\lim_{x\to\infty} x^{-1}\Phi(xv(x), e) = \phi(v, e) \tag{2.4}$$

whenever $v(x) \to v \in \mathbb{R}^d \setminus \{0\}$ *and* $e \in \mathbb{E}$. *If* $\mathbb{P}(\phi(s, \varepsilon_1) = 0) > 0$ *for some* $s \in \mathbb{S}^{d-1}$ *then (2.4) also holds for* $v(x) \to v = 0$ *and* $e \in \mathbb{W}$.

Proof. If $v(x) \to v \in \mathbb{R}^d \setminus \{0\}$ then both $\|v(x)\| \to \|v\|$ and $v(x)/\|v(x)\| \to v/\|v\|$. Thus,

$$\lim_{x\to\infty} \frac{\Phi(xv(x), e)}{x} = \lim_{x\to\infty} \|v(x)\| \frac{\Phi(x\|v(x)\|(v(x)/\|v(x)\|), e)}{x\|v(x)\|} = \|v\|\phi\bigg(\frac{v}{\|v\|}, e\bigg)$$

which, by (2.3), gives (2.4). The case $v(x) \to 0$ follows from (2.2).

Theorem 2.1. *Let* $\{X_t : t \in \mathbb{N}_0\}$ *be given by (1.1), (C1), and (C2). If Conditions 2.1 and 2.2 hold, then for every integer* $t \geq 0$, *as* $x \to \infty$,

$$\mathcal{L}\bigg(\frac{\|X_0\|}{x}, \frac{X_0}{\|X_0\|}, \frac{X_1}{\|X_0\|}, \ldots, \frac{X_t}{\|X_0\|} \,\bigg|\, \|X_0\| > x\bigg) \Rightarrow \mathcal{L}(Y, M_0, M_1, \ldots, M_t)$$

with

$$M_j = \phi(M_{j-1}, \varepsilon_j), \qquad j = 1, 2, \ldots, \tag{2.5}$$

and

(i) $Y, M_0, \varepsilon_1, \varepsilon_2, \ldots$ *are independent with* ε_t *as in (1.1) and (C1),*

(ii) $\mathbb{P}(Y > y) = y^{-\alpha}$ *for* $y \geq 1$,

(iii) $\mathcal{L}(M_0) = \Upsilon$.

We call $\{M_t : t \in \mathbb{N}_0\}$ *the forward tail chain of* $\{X_t : t \in \mathbb{N}_0\}$.

Proof. The argument is by induction on t. The $t = 0$ case is a straightforward consequence of Condition 2.1. So let t be a positive integer and let $f : \mathbb{R} \times (\mathbb{R}^d)^{t+1} \to \mathbb{R}$ be bounded and continuous. We have to show that

$$\lim_{x\to\infty} \mathbb{E}\bigg[f\bigg(\frac{\|X_0\|}{x}, \frac{X_0}{\|X_0\|}, \ldots, \frac{X_t}{\|X_0\|}\bigg) \,\bigg|\, \|X_0\| > x\bigg] = \mathbb{E}[f(Y, M_0, \ldots, M_t)]. \tag{2.6}$$

By (1.1), if $X_0 \neq 0$,

$$\frac{X_t}{\|X_0\|} = \frac{\Phi(X_{t-1}, \varepsilon_t)}{\|X_0\|} = \frac{\Phi(x((\|X_0\|/x)(X_{t-1}/\|X_0\|)), \varepsilon_t)}{x(\|X_0\|/x)}.$$

Hence,

$$\mathbb{E}\bigg[f\bigg(\frac{\|X_0\|}{x}, \frac{X_0}{\|X_0\|}, \ldots, \frac{X_t}{\|X_0\|}\bigg) \,\bigg|\, \|X_0\| > x\bigg] \tag{2.7}$$

$$= \mathbb{E}\bigg[g_x\bigg(\frac{\|X_0\|}{x}, \frac{X_0}{\|X_0\|}, \ldots, \frac{X_{t-1}}{\|X_0\|}\bigg) \,\bigg|\, \|X_0\| > x\bigg], \tag{2.8}$$

where

$$g_x(y, x_0, \dots, x_{t-1}) = \mathbb{E}\left[f\left(y, x_0, \dots, x_{t-1}, \frac{\Phi(x y x_{t-1}, \varepsilon_t)}{xy} \right) \right] \tag{2.9}$$

(note that the expectation is taken with respect to the distribution of ε_t). Define

$$g(y, x_0, \dots, x_{t-1}) = \mathbb{E}[f(y, x_0, \dots, x_{t-1}, \phi(x_{t-1}, \varepsilon_t))]. \tag{2.10}$$

By (2.5),

$$\mathbb{E}[f(Y, M_0, \dots, M_t)] = \mathbb{E}[g(Y, M_0, \dots, M_{t-1})]. \tag{2.11}$$

In view of the identities (2.8) and (2.11), the limit relation in (2.6) will follow if we can show that

$$\mathbb{E}\left[g_x\left(\frac{\|X_0\|}{x}, \frac{X_0}{\|X_0\|}, \dots, \frac{X_{t-1}}{\|X_0\|} \right) \,\middle|\, \|X_0\| > x \right] \to \mathbb{E}[g(Y, M_0, \dots, M_{t-1})] \tag{2.12}$$

as $x \to \infty$. In turn, (2.12) will follow from the induction hypothesis and an extension of the continuous mapping theorem (van der Vaart (1998, Theorem 18.11)) provided

$$\lim_{x\to\infty} g_x(y(x), x_0(x), \dots, x_{t-1}(x)) = g(y, x_0, \dots, x_{t-1}) \tag{2.13}$$

whenever $y(x) \to y$ and $x_i(x) \to x_i$ as $x \to \infty$ with $(y, x_0, \dots, x_{t-1})$ ranging over a set $E \subset \mathbb{R} \times (\mathbb{R}^d)^t$ with $\mathbb{P}((Y, M_0, \dots, M_{t-1}) \in E) = 1$. From the definitions of g_x and g in (2.9) and (2.10), respectively, (2.13) is implied by

$$\lim_{x\to\infty} \frac{\Phi(x w(x), v)}{x} = \phi(w, v) \tag{2.14}$$

whenever $\lim_{x\to\infty} w(x) = w$ and where w and v range over sets that receive probability one by the distributions of M_{t-1} and ε_1, respectively. Since (2.14) is ensured by Condition 2.2 and Lemma 2.1, the statement follows. This concludes the proof.

3. Backward tail processes

From now on, the process $\{X_t\}$ in (1.1), (C1), and (C2) is assumed to be strictly stationary. A necessary and sufficient condition for stationarity is that

$$\mathcal{L}(\Phi(X_0, \varepsilon_1)) = \mathcal{L}(X_0). \tag{3.1}$$

It may be highly nontrivial to find the law for X_0 that solves (3.1). But even when the stationary distribution does not admit an explicit expression, its tails may, in many cases, be found by the theory originally developed in Kesten (1973), Letac (1986), and Goldie (1991). For recent results on specific models; see, for instance, Klüppelberg and Pergamenchtchikov (2003), (2004), De Saporta *et al.* (2004), Mirek (2011), Buraczewski *et al.* (2012), and Collamore and Vidyashankar (2013).

If the process $\{X_t\}$ is stationary, then by Kolmogorov's extension theorem and changing the probability space if necessary, the range of t can without loss of generality be assumed to be the set of all integers, $\mathbb{Z}$; recall that we are interested in distributional properties only, and not in almost-sure properties.

Our aim is to extend Theorem 2.1 and find the asymptotic distribution of the random vector $(X_{-s}, \dots, X_t)$ conditionally on $\|X_0\| > x$ as $x \to \infty$, for all integer s and t (Corollary 5.1). According to Basrak and Segers (2009, Theorem 2.1), if the underlying process is stationary, the existence of a forward tail process ($t \in \mathbb{N}_0$) is enough to guarantee the existence of the tail process as a whole ($t \in \mathbb{Z}$).

https://doi.org/10.1239/jap/1421763332 Published online by Cambridge University Press

Proposition 3.1. *Let $\{X_t : t \in \mathbb{Z}\}$ be a stationary Markov chain with distribution determined by (1.1), (C1), (C2), and (3.1). If Conditions 2.1 and 2.2 hold then there exists a process $\{M_t : t \in \mathbb{Z}\}$ such that, as $x \to \infty$,*

$$\mathcal{L}\left(\frac{X_{-s}}{\|X_0\|}, \dots, \frac{X_0}{\|X_0\|}, \dots, \frac{X_t}{\|X_0\|} \,\middle|\, \|X_0\| > x\right) \Rightarrow \mathcal{L}(M_{-s}, \dots, M_0, \dots, M_t)$$

for all integers $s, t \geq 0$.

Proof. The proof follows from our Theorem 2.1 and Theorem 2.1 in Basrak and Segers (2009), combined with a continuous mapping argument.

We call the process $\{M_t : t \in \mathbb{Z}\}$ the *spectral (tail) process* of $\{X_t : t \in \mathbb{Z}\}$, in accordance with the definition of the process $\{\Theta_t : t \in \mathbb{Z}\}$ in Basrak and Segers (2009).

Basrak and Segers (2009) also state an important property of the limiting process.

Proposition 3.2. *Let $\{X_t : t \in \mathbb{Z}\}$ be a stationary Markov chain with distribution determined by (1.1), (C1), (C2), and (3.1), and spectral process $\{M_t : t \in \mathbb{Z}\}$. Then for all $s, t \geq 0$ and for all bounded and measurable $f : (\mathbb{R}^d)^{s+t+1} \to \mathbb{R}$ satisfying $f(y_{-s}, \dots, y_t) = 0$ whenever $y_{-s} = 0$,*

$$\mathbb{E}[f(M_{-s}, \dots, M_t)] = \mathbb{E}\left[f\left(\frac{M_0}{\|M_s\|}, \dots, \frac{M_{s+t}}{\|M_s\|}\right)\|M_s\|^\alpha \mathbf{1}_{\{M_s \neq 0\}}\right]. \tag{3.2}$$

Proof. It follows directly from our Proposition 3.1 and Theorem 3.1 in Basrak and Segers (2009) that

$$\mathbb{E}[f(M_{-s-i}, \dots, M_{t-i})] = \mathbb{E}\left[f\left(\frac{M_{-s}}{\|M_i\|}, \dots, \frac{M_t}{\|M_i\|}\right)\|M_i\|^\alpha \mathbf{1}_{\{M_i \neq 0\}}\right] \tag{3.3}$$

holds for all bounded and continuous $f : (\mathbb{R}^d)^{t+s+1} \to \mathbb{R}$ satisfying $f(y_{-s}, \dots, y_t) = 0$ whenever $y_0 = 0$ (instead of $y_{-s} = 0$) and all $i \in \mathbb{Z}$. We have added the indicator function on the right-hand side for greater clarity. Let s, t, and f be the same as in the statement of Proposition 3.2. Apply (3.3) to the indices $(\underline{s}, \underline{t}, \underline{i}) = (0, t+s, s)$ to arrive at (3.2); note that $\underline{s} + 1 + \underline{t} = s + 1 + t$ and $f(x_{-\underline{s}}, \dots, x_{\underline{t}}) = 0$ as soon as $x_0 = 0$. Thus, for functions f which are additionally assumed to be continuous, the statement follows directly.

For the general case, we set for abbreviation $\mathbb{A}^* := (\mathbb{R}^d)^{s+t+1} \setminus (\{0\} \times (\mathbb{R}^d)^{s+t})$. Furthermore, let μ denote the restriction of the law of $(M_{-s}, \dots, M_t)$ to $\mathbb{A}^*$ and let ν denote the measure on $\mathbb{A}^*$ defined by

$$\nu(f) = \mathbb{E}\left[f\left(\frac{M_0}{\|M_s\|}, \dots, \frac{M_{s+t}}{\|M_s\|}\right)\|M_s\|^\alpha \mathbf{1}_{\{M_s \neq 0\}}\right]$$

for all bounded and continuous f on $\mathbb{A}^*$. In order to show (3.2) for a general bounded and measurable f with $f(y_{-s}, \dots, y_t) = 0$ if $y_{-s} = 0$ it is sufficient to show that μ and ν coincide. The closed sets of $(\mathbb{R}^d)^{s+t+1}$ which are bounded away from $\{0\} \times (\mathbb{R}^d)^{s+t}$ are a π-system generating $\mathbb{B}(\mathbb{A}^*)$. Indicator functions of closed sets A can be written as pointwise limits of continuous functions with values in $[0, 1]$. If A is bounded away from $\{0\} \times (\mathbb{R}^d)^{s+t}$ we can choose these approximating continuous functions in such a way that they vanish on $\{0\} \times (\mathbb{R}^d)^{s+t}$. Thus, by dominated convergence $\mu(A) = \nu(A)$ for all sets A of a generating π-system and, therefore, $\mu = \nu$ on the Borel sets of $\mathbb{A}^*$ (Billingsley (1968, Theorem 2.2)), concluding the proof.

By Lemma 2.2 in Basrak and Segers (2009) it follows that the distribution of $\{M_t : t \in \mathbb{Z}\}$ is uniquely determined by the distribution of $\{M_t : t \in \mathbb{N}_0\}$ (and $\alpha > 0$). We will use (3.2) to analyze the structure of the spectral process with a special focus on the backward process $\{M_{-t} : t \in \mathbb{N}_0\}$. At the heart of the connection between the forward and backward processes is an adjoint relation between the laws of (M_0, M_1) and (M_0, M_{-1}), which we will examine in the next section.

4. An adjoint relation between distributions

A special case of the equality (3.2) is

$$\mathbb{E}[f(M_{-1}, M_0)] = \mathbb{E}\biggl[f\biggl(\frac{M_0}{\|M_1\|}, \frac{M_1}{\|M_1\|}\biggr)\|M_1\|^\alpha \mathbf{1}_{\{M_1 \neq 0\}}\biggr] \tag{4.1}$$

for all $f : (\mathbb{R}^d)^2 \to \mathbb{R}$ satisfying $f(y_0, y_1) = 0$ whenever $y_0 = 0$. Starting from a given distribution of (M_0, M_1) we will characterize the distributions of (M_{-1}, M_0) which satisfy (4.1). For such an adjoint distribution to exist, the distribution (M_0, M_1) cannot be chosen arbitrarily from the distributions on $\mathbb{S}^{d-1} \times \mathbb{R}^d$. We therefore introduce the following set of 'admissible' distributions.

Definition 4.1. For $\alpha \in (0, \infty)$, let $\mathcal{M}_\alpha = \mathcal{M}_{\alpha,d}$ be the set of all probability measures $\mathbb{P}$ on $\mathbb{S}^{d-1} \times \mathbb{R}^d$ such that

$$\int_{\mathbb{S}^{d-1}\times(\mathbb{R}^d\setminus\{0\})} \mathbf{1}_S\biggl(\frac{m}{\|m\|}\biggr)\|m\|^\alpha \mathbb{P}(\mathrm{d}s, \mathrm{d}m) \leq \mathbb{P}(S \times \mathbb{R}^d) \tag{4.2}$$

for every Borel set $S \subset \mathbb{S}^{d-1}$. We call $\mathcal{M}_\alpha$ the set of *admissible distributions* for $\alpha > 0$.

Note that for $\mathbb{P} \in \mathcal{M}_\alpha$ we have

$$\int_{\mathbb{S}^{d-1}\times\mathbb{R}^d} \|m\|^\alpha \mathbb{P}(\mathrm{d}s, \mathrm{d}m) \leq 1.$$

Now we make the aforementioned notion of an 'adjoint' distribution more concise.

Definition 4.2. For $\mathbb{P} \in \mathcal{M}_\alpha$, define a signed Borel measure $\mathbb{P}^*$ on $\mathbb{S}^{d-1} \times \mathbb{R}^d$ by

$$\mathbb{P}^*(S \times \{0\}) = \mathbb{P}(S \times \mathbb{R}^d) - \int_{\mathbb{S}^{d-1}\times(\mathbb{R}^d\setminus\{0\})} \mathbf{1}_S\biggl(\frac{m}{\|m\|}\biggr)\|m\|^\alpha \mathbb{P}(\mathrm{d}s, \mathrm{d}m), \tag{4.3}$$

$$\mathbb{P}^*(E) = \int_{\mathbb{S}^{d-1}\times(\mathbb{R}^d\setminus\{0\})} \mathbf{1}_E\biggl(\frac{m}{\|m\|}, \frac{s}{\|m\|}\biggr)\|m\|^\alpha \mathbb{P}(\mathrm{d}s, \mathrm{d}m), \tag{4.4}$$

for Borel sets $S \subset \mathbb{S}^{d-1}$ and $E \subset \mathbb{S}^{d-1} \times (\mathbb{R}^d \setminus \{0\})$. We call $\mathbb{P}^*$ the *adjoint measure of* $\mathbb{P}$ *in* $\mathcal{M}_\alpha$.

Lemma 4.1. *Let* $\mathbb{P} \in \mathcal{M}_\alpha$ *and let* $\mathbb{P}^*$ *be the same as in Definition 4.2.*

(i) $\mathbb{P}^*$ *is a probability measure and the marginal distributions induced by* $\mathbb{P}$ *and* $\mathbb{P}^*$ *on* $\mathbb{S}^{d-1}$ *are the same.*

(ii) *For every measurable function* $f\colon \mathbb{S}^{d-1} \times (\mathbb{R}^d \setminus \{0\}) \to \mathbb{R}$,

$$\int_{\mathbb{S}^{d-1}\times(\mathbb{R}^d\setminus\{0\})} f(s^*, m^*)\mathbb{P}^*(\mathrm{d}s^*,\, \mathrm{d}m^*) = \int_{\mathbb{S}^{d-1}\times(\mathbb{R}^d\setminus\{0\})} f\left(\frac{m}{\|m\|}, \frac{s}{\|m\|}\right)\|m\|^\alpha \mathbb{P}(\mathrm{d}s,\, \mathrm{d}m) \tag{4.5}$$

in the sense that if one integral exists then so does the other, and they are the same.

(iii) $\mathbb{P}^* \in \mathcal{M}_\alpha$.

(iv) $(\mathbb{P}^*)^* = \mathbb{P}$.

Proof. (i) By (4.2), $\mathbb{P}^*$ is a nonnegative Borel measure. Let S be a Borel subset of $\mathbb{S}^{d-1}$. We have

$$\mathbb{P}^*(S \times \mathbb{R}^d) = \mathbb{P}^*(S \times \{0\}) + \mathbb{P}^*(S \times (\mathbb{R}^d \setminus \{0\})).$$

Applying (4.3) to the first term on the right-hand side and applying (4.4) with $E = S \times (\mathbb{R}^d \setminus \{0\})$ to the second term on the right-hand side yields

$$\mathbb{P}^*(S \times \mathbb{R}^d) = \mathbb{P}(S \times \mathbb{R}^d).$$

It follows that $\mathbb{P}^*$ is a probability measure (take $S = \mathbb{S}^{d-1}$) on $\mathbb{S}^{d-1} \times \mathbb{R}^d$ inducing the same marginal distribution on $\mathbb{S}^{d-1}$ as $\mathbb{P}$.

(ii) By (4.4), Lemma 4.1(ii) holds for indicator functions $\mathbf{1}_E$ of Borel subsets E of $\mathbb{S}^{d-1} \times (\mathbb{R}^d \setminus \{0\})$. The extension to general bounded, measurable functions follows from the definition of the integral.

(iii) Let S be a Borel subset of $\mathbb{S}^{d-1}$. We will apply Lemma 4.1(ii) to the function

$$f(s, m) = \mathbf{1}_S\left(\frac{m}{\|m\|}\right)\|m\|^\alpha \quad \text{for} \quad (s, m) \in \mathbb{S}^{d-1} \times (\mathbb{R}^d \setminus \{0\}).$$

We find that

$$\begin{aligned}
&\int_{\mathbb{S}^{d-1}\times(\mathbb{R}^d\setminus\{0\})} \mathbf{1}_S\left(\frac{m^*}{\|m^*\|}\right)\|m^*\|^\alpha \mathbb{P}^*(\mathrm{d}s^*,\, \mathrm{d}m^*) \\
&\quad = \int_{\mathbb{S}^{d-1}\times(\mathbb{R}^d\setminus\{0\})} f(s^*, m^*)\mathbb{P}^*(\mathrm{d}s^*,\, \mathrm{d}m^*) \\
&\quad = \int_{\mathbb{S}^{d-1}\times(\mathbb{R}^d\setminus\{0\})} f\left(\frac{m}{\|m\|}, \frac{s}{\|m\|}\right)\|m\|^\alpha \mathbb{P}(\mathrm{d}s,\, \mathrm{d}m) \\
&\quad = \int_{\mathbb{S}^{d-1}\times(\mathbb{R}^d\setminus\{0\})} \mathbf{1}_S\left(\frac{s/\|m\|}{\|(s/\|m\|)\|}\right)\left\|\frac{s}{\|m\|}\right\|^\alpha \|m\|^\alpha \mathbb{P}(\mathrm{d}s,\, \mathrm{d}m) \\
&\quad = \mathbb{P}(S \times (\mathbb{R}^d \setminus \{0\})) \\
&\quad \le \mathbb{P}(S \times \mathbb{R}^d) \\
&\quad = \mathbb{P}^*(S \times \mathbb{R}^d),
\end{aligned}$$

where we applied (i) in the last step.

(iv) Let $Q = (\mathbb{P}^*)^*$. We already know that Q is a probability measure on $\mathbb{S}^{d-1} \times \mathbb{R}^d$, that $Q \in \mathcal{M}_\alpha$, and that the marginal induced by Q on $\mathbb{S}^{d-1}$ coincides with the one of $\mathbb{P}^*$ and thus

https://doi.org/10.1239/jap/1421763332 Published online by Cambridge University Press

with the one of $\mathbb{P}$. Let f be a nonnegative, measurable function on $\mathbb{S}^{d-1} \times (\mathbb{R}^d \setminus \{0\})$. Define the nonnegative, measurable function g on $\mathbb{S}^{d-1} \times (\mathbb{R}^d \setminus \{0\})$ by

$$g(s, m) = f\left(\frac{m}{\|m\|}, \frac{s}{\|m\|}\right)\|m\|^\alpha \quad \text{for } (s, m) \in \mathbb{S}^{d-1} \times (\mathbb{R}^d \setminus \{0\}).$$

We have

$$g\left(\frac{m}{\|m\|}, \frac{s}{\|m\|}\right)\|m\|^\alpha = f\left(\frac{s/\|m\|}{\|(s/\|m\|)\|}, \frac{m/\|m\|}{\|(s/\|m\|)\|}\right)\left\|\frac{s}{\|m\|}\right\|^\alpha \|m\|^\alpha = f(s, m). \tag{4.6}$$

By Lemma 4.1(ii) applied first to Q and f and then to $\mathbb{P}^*$ and g, we have

$$\begin{aligned}
&\int_{\mathbb{S}^{d-1}\times(\mathbb{R}^d\setminus\{0\})} f(s, m) Q(\mathrm{d}s,\, \mathrm{d}m) \\
&\quad = \int_{\mathbb{S}^{d-1}\times(\mathbb{R}^d\setminus\{0\})} f\left(\frac{m^*}{\|m^*\|}, \frac{s^*}{\|m^*\|}\right)\|m^*\|^\alpha \mathbb{P}^*(\mathrm{d}s^*,\, \mathrm{d}m^*) \\
&\quad = \int_{\mathbb{S}^{d-1}\times(\mathbb{R}^d\setminus\{0\})} g(s^*, m^*)\mathbb{P}^*(\mathrm{d}s^*,\, \mathrm{d}m^*) \\
&\quad = \int_{\mathbb{S}^{d-1}\times(\mathbb{R}^d\setminus\{0\})} g\left(\frac{m}{\|m\|}, \frac{s}{\|m\|}\right)\|m\|^\alpha \mathbb{P}(\mathrm{d}s,\, \mathrm{d}m) \\
&\quad = \int_{\mathbb{S}^{d-1}\times(\mathbb{R}^d\setminus\{0\})} f(s, m)\mathbb{P}(\mathrm{d}s,\, \mathrm{d}m),
\end{aligned}$$

where we used (4.6) in the last step. It follows that Q and $\mathbb{P}$ coincide on $\mathbb{S}^{d-1} \times (\mathbb{R}^d \setminus \{0\})$. As Q and $\mathbb{P}$ also induce the same marginal distributions on $\mathbb{S}^{d-1}$, it follows that they must also coincide on $\mathbb{S}^{d-1} \times \{0\}$. As a consequence, Q is equal to $\mathbb{P}$.

The next lemma shows that the class $\mathcal{M}_\alpha$ and the adjoint relation on it arise naturally in the context of regularly varying Markov chains.

Lemma 4.2. *Let $\{X_t : t \in \mathbb{Z}\}$ be a stationary Markov chain with distribution determined by (1.1), (C1), (C2), and (3.1). If Conditions 2.1 and 2.2 hold then $\mathcal{L}(M_0, M_1)$ belongs to $\mathcal{M}_\alpha$ and its adjoint is equal to $\mathcal{L}(M_0, M_{-1})$.*

Proof. To prove admissibility, we have to show that

$$\mathbb{E}\left[\mathbf{1}_S\left(\frac{M_1}{\|M_1\|}\right)\|M_1\|^\alpha\right] \le \mathbb{P}(M_0 \in S) \quad \text{for every Borel set } S \subset \mathbb{S}^{d-1}. \tag{4.7}$$

Let f be a bounded, nonnegative and continuous function on $\mathbb{S}^{d-1}$. We will show that

$$\mathbb{E}\left[f\left(\frac{M_1}{\|M_1\|}\right)\|M_1\|^\alpha\right] \le \mathbb{E}[f(M_0)]. \tag{4.8}$$

Equation (4.8) implies (4.7) for closed sets S because the indicator function of a closed set S can be written as the pointwise limit of a decreasing sequence of continuous functions taking values in the interval [0, 1]. From this we arrive at (4.7) for an arbitrary Borel set S by invoking an increasing sequence of closed sets S_n contained in S such that $\mathbb{E}[\mathbf{1}_{S_n}(M_1/\|M_1\|)\|M_1\|^\alpha]$

and $\mathbb{P}(M_0 \in S_n)$ converge to $\mathbb{E}[\mathbf{1}_S(M_1/\|M_1\|)\|M_1\|^\alpha]$ and $\mathbb{P}(M_0 \in S)$, respectively (see, for instance, Theorem 1.1 in Billingsley (1968, p. 7)).

Let $\delta > 0$. By stationarity of $\{X_t : t \in \mathbb{Z}\}$ and by definition of the spectral process $\{M_t : t \in \mathbb{Z}\}$, we have

$$\begin{aligned}
\mathbb{E}[f(M_0)] &= \lim_{x\to\infty} \mathbb{E}\left[f\left(\frac{X_1}{\|X_1\|}\right) \mid \|X_1\| > x \right] \\
&\geq \limsup_{x\to\infty} \mathbb{E}\left[\mathbf{1}_{\{\|X_0\|>\delta x\}} f\left(\frac{X_1}{\|X_1\|}\right) \mid \|X_1\| > x \right] \\
&= \limsup_{x\to\infty} \frac{\mathbb{P}[\|X_0\| > \delta x]}{\mathbb{P}[\|X_1\| > x]} \mathbb{E}\left[f\left(\frac{X_1}{\|X_1\|}\right) \mathbf{1}_{\{\|X_1\|>x\}} \mid \|X_0\| > \delta x \right] \\
&= \delta^{-\alpha} \mathbb{E}\left[f\left(\frac{M_1}{\|M_1\|}\right) \mathbf{1}_{\{Y\|M_1\|>\delta^{-1}\}} \right].
\end{aligned}$$

In the last line, Y is a Pareto(α) random variable, independent of M_1. As $\mathbb{P}(Y\|M_1\| = \delta^{-1}) = 0$ by continuity of the law of Y, the last equality in the above display follows from the continuous mapping theorem.

Since the distribution of $Y^{-\alpha}$ is uniform on the interval (0, 1), we have

$$\begin{aligned}
\delta^{-\alpha} \mathbb{E}\left[f\left(\frac{M_1}{\|M_1\|}\right) \mathbf{1}_{\{Y\|M_1\|>\delta^{-1}\}} \right] &= \delta^{-\alpha} \mathbb{E}\left[f\left(\frac{M_1}{\|M_1\|}\right) \mathbf{1}_{\{\delta^\alpha\|M_1\|^\alpha>Y^{-\alpha}\}} \right] \\
&= \delta^{-\alpha} \mathbb{E}\left[\mathbb{E}\left[f\left(\frac{M_1}{\|M_1\|}\right) \mathbf{1}_{\{\delta^\alpha\|M_1\|^\alpha>Y^{-\alpha}\}} \mid M_1 \right]\right] \\
&= \delta^{-\alpha} \mathbb{E}\left[f\left(\frac{M_1}{\|M_1\|}\right) \min(\delta^\alpha\|M_1\|^\alpha, 1) \right] \\
&= \mathbb{E}\left[f\left(\frac{M_1}{\|M_1\|}\right) \min(\|M_1\|^\alpha, \delta^{-\alpha}) \right].
\end{aligned}$$

We obtain that for every $\delta > 0$,

$$\mathbb{E}[f(M_0)] \geq \mathbb{E}\left[f\left(\frac{M_1}{\|M_1\|}\right) \min(\|M_1\|^\alpha, \delta^{-\alpha}) \right].$$

Take the limit as $\delta \to 0$ and apply the monotone convergence theorem to obtain (4.8).

Next we show that the adjoint of $\mathcal{L}(M_0, M_1)$ is equal to $\mathcal{L}(M_0, M_{-1})$. We have to check the two equations

$$\mathbb{P}((M_0, M_{-1}) \in S \times \{0\}) = \mathbb{P}(M_0 \in S) - \mathbb{E}\left[\mathbf{1}_{\mathbb{R}^d\setminus\{0\}}(M_1) \mathbf{1}_S\left(\frac{M_1}{\|M_1\|}\right) \|M_1\|^\alpha \right], \tag{4.9}$$

$$\mathbb{P}((M_0, M_{-1}) \in E) = \mathbb{E}\left[\mathbf{1}_{\mathbb{R}^d\setminus\{0\}}(M_1) \mathbf{1}_E\left(\frac{M_1}{\|M_1\|}, \frac{M_0}{\|M_1\|}\right) \|M_1\|^\alpha \right], \tag{4.10}$$

for all Borel sets $S \subset \mathbb{S}^{d-1}$ and $E \subset \mathbb{S}^{d-1} \times (\mathbb{R}^d\setminus\{0\})$. Since the first component M_0 is common to both laws it is sufficient to check only the second equation, (4.10). Set $f(m_{-1}, m_0) =$

$\mathbf{1}_E(m_0, m_{-1})$ on $\mathbb{R} \times \mathbb{S}^{d-1}$. Note that $f(0, m_0) = 0$. Apply (4.1) to f, to obtain

$$\begin{aligned}\mathbb{P}((M_0, M_{-1}) \in E) &= \mathbb{E}[f(M_{-1}, M_0)] \\ &= \mathbb{E}\bigg[f\bigg(\frac{M_0}{\|M_1\|}, \frac{M_1}{\|M_1\|}\bigg)\|M_1\|^\alpha \mathbf{1}_{\{M_1 \neq 0\}}\bigg] \\ &= \mathbb{E}\bigg[\mathbf{1}_{\mathbb{R}^d \setminus \{0\}}(M_1)\mathbf{1}_E\bigg(\frac{M_1}{\|M_1\|}, \frac{M_0}{\|M_1\|}\bigg)\|M_1\|^\alpha\bigg],\end{aligned}$$

which gives (4.10) as required.

Remark 4.1. The determination of the adjoint measure is particularly simple for probability measures $\mathbb{P}$ such that

$$\int_{\mathbb{S}^{d-1} \times \mathbb{R}^d} \|m\|^\alpha \mathbb{P}(\mathrm{d}s,\ \mathrm{d}m) = 1, \tag{4.11}$$

since in this case $\mathbb{P}^*(\mathbb{S}^{d-1} \times \{0\}) = 0$ by (4.3) and $\mathbb{P}^*$ is completely described by (4.4).

Remark 4.2. We call a measure $\mathbb{P} \in \mathcal{M}_\alpha$ *self-adjoint* if $\mathbb{P}^* = \mathbb{P}$. An example for such a distribution in the case of $d = 1$ and $\alpha = 1$ is given by $\mathbb{P} = \mathcal{L}(1, Y)$, where $Y = \exp(X - \frac{1}{2})$ for a standard normally distributed X (compare Segers (2007, Example 3.2)).

Definition 4.2 and Lemma 4.1 generalize Proposition 3.1 in Segers (2007) to the multivariate case. Examples 3.2–3.4 in Segers (2007) illustrate the adjoint relation for laws on $\{-1, +1\} \times \mathbb{R}$. We conclude the section with a multivariate example.

Example 4.1. Let $\alpha > 0$ and let $\mathbb{P}$ be the law of (C, RQC), with C, R and Q independent, C taking values in $\mathbb{S}^{d-1}$, R a positive random variable with $\mathbb{E}[R^\alpha] = 1$, and Q a random orthogonal $d \times d$ matrix, that is $Q^\top = Q^{-1}$ almost surely; also assume that the laws of C and QC are the same (compare Example 6.1). We easily verify that $\mathbb{P} \in \mathcal{M}_\alpha$ and (4.11) holds, so that the adjoint law $\mathbb{P}^*$ is concentrated on $\mathbb{S}^{d-1} \times (\mathbb{R}^d \setminus \{0\})$. Thus, from (4.4) we derive that for Borel sets $S \subset \mathbb{S}^{d-1}$ and $T \subset \mathbb{R}^d \setminus \{0\}$,

$$\mathbb{P}^*(S \times T) = \mathbb{E}\bigg[\mathbf{1}_S(QC)\mathbf{1}_T\bigg(\frac{C}{R}\bigg)R^\alpha\bigg]. \tag{4.12}$$

Additionally, if we assume that C is uniformly distributed on $\mathbb{S}^{d-1}$ (which readily implies $\mathcal{L}(C) = \mathcal{L}(QC)$ for any law of Q), then

$$\begin{aligned}\mathbb{E}[\mathbf{1}_S(QC)\mathbf{1}_T(C)] &= \mathbb{E}\bigg[\int_{\mathbb{R}^{d\times d}} \mathbf{1}_S(qC)\mathbf{1}_T(C)\mathbb{P}^Q(dq)\bigg] \\ &= \mathbb{E}\bigg[\int_{\mathbb{R}^{d\times d}} \mathbf{1}_S(C)\mathbf{1}_T(q'C)\mathbb{P}^Q(dq)\bigg] \\ &= \mathbb{E}[\mathbf{1}_S(C)\mathbf{1}_T(Q'C)]\end{aligned}$$

and it follows from (4.12) that $\mathbb{P}^*$ is the law of $(C^*, R^*Q^*C^*)$, with C^*, R^*, and Q^* independent, $\mathcal{L}(C^*) = \mathcal{L}(C)$, $\mathcal{L}(Q^*) = \mathcal{L}(Q')$, and the law of $R^* > 0$ given by $\mathbb{E}[f(R^*)] = \mathbb{E}[f(1/R)R^\alpha]$ for measurable functions f on $(0, \infty)$.

https://doi.org/10.1239/jap/1421763332 Published online by Cambridge University Press

5. Back-and-forth tail chains and the spectral process

In this section we will analyze a certain class of discrete-time processes which are constructed from a pair of adjoint distributions. We will see that this class of processes fulfills (3.2) for all $s, t \geq 0$.

Definition 5.1. A d-dimensional discrete-time process $\{M_t : t \in \mathbb{Z}\}$ is called a *back-and-forth tail chain* with index $\alpha > 0$, notation BFTC(α), if the following properties hold.

(i) $\mathcal{L}(M_0, M_1)$ and $\mathcal{L}(M_0, M_{-1})$ belong to $\mathcal{M}_\alpha$ and are adjoint.

(ii) The forward process $\{M_t : t \in \mathbb{N}_0\}$ is a Markov chain with respect to the filtration $\sigma(M_s, -\infty < s \leq t)$, $t \geq 0$, and the Markov kernel satisfies

$$\mathbb{P}(M_t \in \cdot \mid M_{t-1} = x_{t-1}) = \begin{cases} \delta_0(\cdot) & \text{if } x_{t-1} = 0, \\ \mathbb{P}\Big(\|x_{t-1}\| M_1 \in \cdot \Big| M_0 = \frac{x_{t-1}}{\|x_{t-1}\|} \Big) & \text{if } x_{t-1} \neq 0. \end{cases}$$

(iii) The backward process $\{M_{-t} : t \in \mathbb{N}_0\}$ is a Markov chain with respect to the filtration $\sigma(M_{-s}, -\infty < s \leq t)$, $t \geq 0$, and the Markov kernel satisfies

$$\begin{aligned} &\mathbb{P}(M_{-t} \in \cdot \mid M_{-t+1} = x_{-t+1}) \\ &\quad = \begin{cases} \delta_0(\cdot) & \text{if } x_{-t+1} = 0, \\ \mathbb{P}\Big(\|x_{-t+1}\| M_{-1} \in \cdot \Big| M_0 = \frac{x_{-t+1}}{\|x_{-t+1}\|} \Big) & \text{if } x_{-t+1} \neq 0. \end{cases} \end{aligned}$$

Clearly, $\{M_t : t \in \mathbb{Z}\}$ is a BFTC(α) if and only if $\{M_{-t} : t \in \mathbb{Z}\}$ is a BFTC(α). The distribution of a BFTC(α) is completely determined by an admissible law of (M_0, M_1) (and $\alpha > 0$).

The fact that the distributions $\mathbb{P} = \mathcal{L}(M_0, M_1)$ and $\mathbb{P}^* = \mathcal{L}(M_0, M_{-1})$ are adjoint in $\mathcal{M}_\alpha$ implies that for every measurable function $f : \mathbb{R}^d \times \mathbb{S}^{d-1} \to \mathbb{R}$ such that $f(0, s) = 0$ for all $s \in \mathbb{S}^{d-1}$, we have

$$\begin{aligned} \mathbb{E}[f(M_{-1}, M_0)] &= \int_{\mathbb{S}^{d-1} \times (\mathbb{R}^d \setminus \{0\})} f(m, s) \mathbb{P}^*(\mathrm{d}s, \mathrm{d}m) && (5.1) \\ &= \int_{\mathbb{S}^{d-1} \times (\mathbb{R}^d \setminus \{0\})} f(s/\|m\|, m/\|m\|) \|m\|^\alpha \mathbb{P}(\mathrm{d}s, \mathrm{d}m) \\ &= \mathbb{E}\bigg[f\bigg(\frac{M_0}{\|M_1\|}, \frac{M_1}{\|M_1\|} \bigg) \|M_1\|^\alpha \mathbf{1}_{\{M_1 \neq 0\}} \bigg], && (5.2) \end{aligned}$$

in the sense that if one expectation exists, then so does the other, the two expectations being equal. This corresponds to (4.1) which originally motivated the definition of an adjoint distribution. The above formula is the special case $s = 1$ and $t = 0$ of the following result.

Proposition 5.1. *Let* $\{M_t : t \in \mathbb{Z}\}$ *be a* BFTC(α). *For all integers* $s, t \geq 0$ *and for all measurable functions* $f : (\mathbb{R}^d)^{s+1+t} \to \mathbb{R}$ *vanishing on* $\{0\} \times (\mathbb{R}^d)^{s+t}$, *the* $s + 1$ *numbers*

$$\mathbb{E}\bigg[f\bigg(\frac{M_{-s+i}}{\|M_i\|}, \ldots, \frac{M_{t+i}}{\|M_i\|} \bigg) \|M_i\|^\alpha \mathbf{1}_{\{M_i \neq 0\}} \bigg], \qquad i = 0, \ldots, s, \tag{5.3}$$

are all the same, in the sense that if one integral exists then they all exist and they are equal.

Proof. For $s = 0$ there is nothing to prove, so assume that $s \geq 1$. By definition of the integral, it is sufficient to consider the case where f is nonnegative, in which case the expectations in (5.3) are always well defined, possibly equal to ∞.

Reduction to the case $i \in \{0, 1\}$. Suppose first that we can show that the numbers corresponding to $i = 0$ and $i = 1$ in (5.3) are equal, that is (note that $\|M_0\| = 1$),

$$\mathbb{E}[f(M_{-s}, \ldots, M_t)] = \mathbb{E}\biggl[f\biggl(\frac{M_{-s+1}}{\|M_1\|}, \ldots, \frac{M_{t+1}}{\|M_1\|}\biggr)\|M_1\|^{\alpha}\mathbf{1}_{\{M_1 \neq 0\}}\biggr]. \tag{5.4}$$

Take arbitrary $i = 0, \ldots, s-1$. Note that

$$\mathbb{E}\biggl[f\biggl(\frac{M_{-s+i}}{\|M_i\|}, \ldots, \frac{M_{t+i}}{\|M_i\|}\biggr)\|M_i\|^{\alpha}\mathbf{1}_{\{M_i \neq 0\}}\biggr] = \mathbb{E}[g(M_{-s+i}, \ldots, M_{t+i})]$$

for a measurable function $g\colon (\mathbb{R}^d)^{s+1+t} \to \mathbb{R}$ that vanishes as soon as its first d-tuple of arguments is 0. By (5.4) applied to $\tilde{s} = s - i$ and $\tilde{t} = t + i$, we find

$$\mathbb{E}[g(M_{-s+i}, \ldots, M_{t+i})] = \mathbb{E}\biggl[g\biggl(\frac{M_{-s+i+1}}{\|M_1\|}, \ldots, \frac{M_{t+i+1}}{\|M_1\|}\biggr)\|M_1\|^{\alpha}\mathbf{1}_{\{M_1 \neq 0\}}\biggr].$$

By definition of g, if $M_1 \neq 0$ then

$$\begin{aligned} g&\biggl(\frac{M_{-s+i+1}}{\|M_1\|}, \ldots, \frac{M_{t+i+1}}{\|M_1\|}\biggr) \\ &= f\biggl(\frac{M_{-s+i+1}/\|M_1\|}{\|(M_{i+1}/\|M_1\|)\|}, \ldots, \frac{M_{t+i+1}/\|M_1\|}{\|(M_{i+1}/\|M_1\|)\|}\biggr)\biggl\|\frac{M_{i+1}}{\|M_1\|}\biggr\|^{\alpha}\mathbf{1}_{\{M_{i+1} \neq 0\}} \\ &= f\biggl(\frac{M_{-s+i+1}}{\|M_{i+1}\|}, \ldots, \frac{M_{t+i+1}}{\|M_{i+1}\|}\biggr)\frac{\|M_{i+1}\|^{\alpha}}{\|M_1\|^{\alpha}}\mathbf{1}_{\{M_{i+1} \neq 0\}}. \end{aligned}$$

Combine the previous three displays to see that

$$\begin{aligned} \mathbb{E}&\biggl[f\biggl(\frac{M_{-s+i}}{\|M_i\|}, \ldots, \frac{M_{t+i}}{\|M_i\|}\biggr)\|M_i\|^{\alpha}\mathbf{1}_{\{M_i \neq 0\}}\biggr] \\ &= \mathbb{E}\biggl[f\biggl(\frac{M_{-s+i+1}}{\|M_{i+1}\|}, \ldots, \frac{M_{t+i+1}}{\|M_{i+1}\|}\biggr)\|M_{i+1}\|^{\alpha}\mathbf{1}_{\{M_1 \neq 0, M_{i+1} \neq 0\}}\biggr]. \end{aligned}$$

By definition of the forward chain $(M_t)_{t \geq 0}$, we have $M_{i+1} = 0$ as soon as $M_1 = 0$. As a consequence, we may suppress the event $\{M_1 \neq 0\}$ in the indicator function on the right-hand side, and, thus,

$$\begin{aligned} \mathbb{E}&\biggl[f\biggl(\frac{M_{-s+i}}{\|M_i\|}, \ldots, \frac{M_{t+i}}{\|M_i\|}\biggr)\|M_i\|^{\alpha}\mathbf{1}_{\{M_i \neq 0\}}\biggr] \\ &= \mathbb{E}\biggl[f\biggl(\frac{M_{-s+i+1}}{\|M_{i+1}\|}, \ldots, \frac{M_{t+i+1}}{\|M_{i+1}\|}\biggr)\|M_{i+1}\|^{\alpha}\mathbf{1}_{\{M_{i+1} \neq 0\}}\biggr]. \end{aligned}$$

We conclude that in order to show (5.3), it is enough to show (5.4). We will show (5.4) by induction on $s \geq 1$.

https://doi.org/10.1239/jap/1421763332 Published online by Cambridge University Press

Proof of (5.4) if $s = 1$. We have to show that

$$\mathbb{E}[f(M_{-1}, \dots, M_t)] = \mathbb{E}\left[f\left(\frac{M_0}{\|M_1\|}, \dots, \frac{M_{t+1}}{\|M_1\|} \right) \|M_1\|^\alpha \mathbf{1}_{\{M_1 \neq 0\}} \right]. \tag{5.5}$$

We will proceed by induction on $t \geq 0$.

The $t = 0$ case is nothing more than the adjoint relation between the laws of (M_0, M_1) and (M_0, M_{-1}); see (5.2).

Let $t \geq 1$ and let (5.5) be fulfilled for $t - 1$. By the Markov property,

$$\mathbb{E}[f(M_{-1}, \dots, M_t)] = \mathbb{E}[g(M_{-1}, \dots, M_{t-1})]$$

with

$$g(m_{-1}, \dots, m_{t-1}) = \mathbb{E}[f(m_{-1}, \dots, m_{t-1}, M_t) \mid M_{t-1} = m_{t-1}].$$

As $g(0, m_0, \dots, m_{t-1}) = 0$, we can apply the induction hypothesis, yielding

$$\mathbb{E}[g(M_{-1}, \dots, M_{t-1})] = \mathbb{E}\left[g\left(\frac{M_0}{\|M_1\|}, \dots, \frac{M_t}{\|M_1\|} \right) \|M_1\|^\alpha \mathbf{1}_{\{M_1 \neq 0\}} \right].$$

The defining property of a BFTC implies that for every $c > 0$, for every integer $r \geq 1$, and for every nonnegative, measurable function h on $\mathbb{R}^d$,

$$\mathbb{E}\left[h(cM_r) \,\middle|\, M_{r-1} = \frac{m}{c} \right] = \begin{cases} h(0) & \text{if } m = 0, \\ \mathbb{E}\left[h(\|m\| M_1) \,\middle|\, M_0 = \frac{m}{\|m\|} \right] & \text{if } m \neq 0, \end{cases} \tag{5.6}$$

the right-hand side not depending on the scaling constant c nor on the time index r. It follows that if $m_1 \neq 0$,

$$\begin{aligned} g\left(\frac{m_0}{\|m_1\|}, \dots, \frac{m_t}{\|m_1\|} \right) &= \mathbb{E}\left[f\left(\frac{m_0}{\|m_1\|}, \dots, \frac{m_t}{\|m_1\|}, M_t \right) \,\middle|\, M_{t-1} = \frac{m_t}{\|m_1\|} \right] \\ &= \mathbb{E}\left[f\left(\frac{m_0}{\|m_1\|}, \dots, \frac{m_t}{\|m_1\|}, \frac{M_{t+1}}{\|m_1\|} \right) \,\middle|\, M_t = m_t \right]. \end{aligned}$$

We find that, on the event $\{M_1 \neq 0\}$, by the Markov property,

$$g\left(\frac{M_0}{\|M_1\|}, \dots, \frac{M_t}{\|M_1\|} \right) = \mathbb{E}\left[f\left(\frac{M_0}{\|M_1\|}, \dots, \frac{M_t}{\|M_1\|}, \frac{M_{t+1}}{\|M_1\|} \right) \,\middle|\, M_0, \dots, M_t \right].$$

We can conclude that

$$\begin{aligned} \mathbb{E}[f(M_{-1}, \dots, M_t)] &= \mathbb{E}[g(M_{-1}, \dots, M_{t-1})] \\ &= \mathbb{E}\left[g\left(\frac{M_0}{\|M_1\|}, \dots, \frac{M_t}{\|M_1\|} \right) \|M_1\|^\alpha \mathbf{1}_{\{M_1 \neq 0\}} \right] \\ &= \mathbb{E}\left[f\left(\frac{M_0}{\|M_1\|}, \dots, \frac{M_t}{\|M_1\|}, \frac{M_{t+1}}{\|M_1\|} \right) \|M_1\|^\alpha \mathbf{1}_{\{M_1 \neq 0\}} \right], \end{aligned}$$

as required.

Proof of (5.4) for general $s \geq 1$. The case $s = 1$ was treated above. So let $s \geq 2$ and let (5.4) hold for $s - 1$. By the Markov property, we have

$$\mathbb{E}[f(M_{-s}, \dots, M_t)] = \mathbb{E}[g(M_{-s+1}, \dots, M_t)]$$

with $g\colon (\mathbb{R}^d)^{s+t} \to \mathbb{R}$ a nonnegative, measurable function defined by

$$g(m_{-s+1}, \dots, m_t) = \mathbb{E}[f(M_{-s}, m_{-s+1}, \dots, m_t) \mid M_{-s+1} = m_{-s+1}].$$

Conditionally on $M_{-s+1} = 0$, we have $M_{-s} = 0$, and, thus, $f(M_{-s}, \dots) = 0$ too. It follows that $g(0, m_{-s+2}, \dots, m_t) = 0$. By the induction hypothesis, we therefore have

$$\mathbb{E}[g(M_{-s+1}, \dots, M_t)] = \mathbb{E}\bigg[g\bigg(\frac{M_{-s+2}}{\|M_1\|}, \dots, \frac{M_{t+1}}{\|M_1\|}\bigg)\|M_1\|^\alpha \mathbf{1}_{\{M_1 \neq 0\}}\bigg].$$

As for the forward chain in (5.6), we have for every nonnegative, measurable function h on $\mathbb{R}^d$ and every $c > 0$,

$$\mathbb{E}\bigg[h(cM_{-r})\,\bigg|\, M_{-r+1} = \frac{m}{c}\bigg] = \begin{cases} h(0) & \text{if } m = 0, \\ \mathbb{E}\Big[h(\|m\| M_{-1})\,\Big|\, M_0 = \frac{m}{\|m\|}\Big] & \text{if } m \neq 0, \end{cases}$$

the right-hand side not depending on the scaling constant $c > 0$ nor on the time index $r = 1, 2, \dots$. It follows that for $m_1 \neq 0$, we have

$$\begin{aligned} &g\bigg(\frac{m_{-s+2}}{\|m_1\|}, \dots, \frac{m_{t+1}}{\|m_1\|}\bigg) \\ &\quad = \mathbb{E}\bigg[f\bigg(M_{-s}, \frac{m_{-s+2}}{\|m_1\|}, \dots, \frac{m_{t+1}}{\|m_1\|}\bigg)\,\bigg|\, M_{-s+1} = \frac{m_{-s+2}}{\|m_1\|}\bigg] \\ &\quad = \mathbb{E}\bigg[f\bigg(\frac{M_{-s+1}}{\|m_1\|}, \frac{m_{-s+2}}{\|m_1\|}, \dots, \frac{m_{t+1}}{\|m_1\|}\bigg)\,\bigg|\, M_{-s+2} = m_{-s+2}\bigg]. \end{aligned}$$

Invoking the Markov property again, we conclude that

$$\begin{aligned} \mathbb{E}[f(M_{-s}, \dots, M_t)] &= \mathbb{E}[g(M_{-s+1}, \dots, M_t)] \\ &= \mathbb{E}\bigg[g\bigg(\frac{M_{-s+2}}{\|M_1\|}, \dots, \frac{M_{t+1}}{\|M_1\|}\bigg)\|M_1\|^\alpha \mathbf{1}_{\{M_1 \neq 0\}}\bigg] \\ &= \mathbb{E}\bigg[f\bigg(\frac{M_{-s+1}}{\|M_1\|}, \dots, \frac{M_{t+1}}{\|M_1\|}\bigg)\|M_1\|^\alpha \mathbf{1}_{\{M_1 \neq 0\}}\bigg], \end{aligned}$$

as required. This concludes the proof of Proposition 5.1.

The following proposition connects BFTCs and spectral processes.

Proposition 5.2. *Let $\{Y_t : t \in \mathbb{Z}\}$ be an $\mathbb{R}^d$-valued process and let $\{M_t : t \in \mathbb{Z}\}$ be an $\mathbb{R}^d$-valued* BFTC*(α). If*

$$\mathcal{L}(Y_0, \dots, Y_t) = \mathcal{L}(M_0, \dots, M_t) \quad \textit{for all } t \geq 0 \tag{5.7}$$

and if

$$\mathbb{E}[f(Y_{-s}, \dots, Y_t)] = \mathbb{E}\bigg[f\bigg(\frac{Y_0}{\|Y_s\|}, \dots, \frac{Y_{s+t}}{\|Y_s\|}\bigg)\|Y_s\|^\alpha \mathbf{1}_{\{Y_s \neq 0\}}\bigg] \tag{5.8}$$

for all $s, t \geq 0$ *and for all bounded and measurable* $f\colon (\mathbb{R}^d)^{s+t+1} \to \mathbb{R}$ *satisfying* $f(y_{-s}, \dots, y_t) = 0$ *whenever* $y_{-s} = 0$*, then*

$$\mathcal{L}(Y_{-s}, \dots, Y_t) = \mathcal{L}(M_{-s}, \dots, M_t) \quad \textit{for all } s, t \geq 0. \tag{5.9}$$

Proof. The proof relies on the fact that both the process $\{Y_t : t \in \mathbb{Z}\}$ which satisfies (5.8) and the BFTC(α) are uniquely determined by their forward process. Our proof is by induction on s. For $s = 0$, (5.9) is equal to the assumption (5.7) for all $t \geq 0$. For the induction step, assume that (5.9) holds for a fixed value of $\tilde{s} = s - 1 \geq 0$ and all $t \geq 0$. Let $f\colon (\mathbb{R}^d)^{s+t+1} \to \mathbb{R}$ be a bounded continuous function. Write

$$f(y_{-s}, \dots, y_t) = f_1(y_{-s}, \dots, y_t) + f_2(y_{-s}, \dots, y_t)$$

with

$$\begin{aligned} f_1(y_{-s}, \dots, y_t) &= f(0, y_{-s+1}, \dots, y_t), \\ f_2(y_{-s}, \dots, y_t) &= f(y_{-s}, y_{-s+1}, \dots, y_t) - f(0, y_{-s+1}, \dots, y_t), \end{aligned}$$

and note that $f_2(0, y_{-s+1}, \dots, y_t) = 0$, while the value of f_1 does not depend on the first coordinate of the argument. Then

$$\begin{aligned} &\mathbb{E}[f(Y_{-s}, \dots, Y_t)] \\ &\quad = \mathbb{E}[f_1(Y_{-s}, \dots, Y_t)] + \mathbb{E}[f_2(Y_{-s}, \dots, Y_t)] \\ &\quad = \mathbb{E}[f_1(Y_{-s}, \dots, Y_t)] + \mathbb{E}\bigg[f_2\bigg(\frac{Y_0}{\|Y_s\|}, \dots, \frac{Y_{s+t}}{\|Y_s\|}\bigg)\|Y_s\|^\alpha \mathbf{1}_{\{Y_s \neq 0\}}\bigg] \\ &\quad = \mathbb{E}[f_1(M_{-s}, \dots, M_t)] + \mathbb{E}\bigg[f_2\bigg(\frac{M_0}{\|M_s\|}, \dots, \frac{M_{s+t}}{\|M_s\|}\bigg)\|M_s\|^\alpha \mathbf{1}_{\{M_s \neq 0\}}\bigg], \end{aligned}$$

where both the induction hypothesis and (5.8) and (5.9) have been used. Since $\{M_t : t \in \mathbb{Z}\}$ is a BFTC(α), we may apply Proposition 5.1 for $i = s$ and $i = 0$ (note that $\|M_0\| = 1$), so that the above expression is equal to

$$\mathbb{E}[f_1(M_{-s}, \dots, M_t)] + \mathbb{E}[f_2(M_{-s}, \dots, M_t)] = \mathbb{E}[f(M_{-s}, \dots, M_t)],$$

which completes the induction step and the proof.

Remark 5.1. Proposition 5.2 can be read in the following way: every spectral process $\{M_t : t \in \mathbb{Z}\}$ with a forward process (meaning: $\{M_t : t \in \mathbb{N}_0\}$) which has a BFTC($\alpha$) structure, automatically has a BFTC(α)-backward-distribution as well. This means that a Markovian structure in the forward spectral process (which may also arise in settings where the underlying process is non-Markovian) is enough to secure a Markovian structure of the backward spectral process as well.

Corollary 5.1. *Let* $\{X_t : t \in \mathbb{Z}\}$ *be a stationary Markov chain with distribution determined by (1.1), (C1), (C2), and (3.1). Then the corresponding spectral process* $\{M_t : t \in \mathbb{Z}\}$ *is a* BFTC(α).

We call $\{M_{-t} : t \in \mathbb{N}_0\}$ *the* backward tail chain of $\{X_t : t \in \mathbb{Z}\}$ *and* $\{M_t : t \in \mathbb{Z}\}$ *the* tail chain of $\{X_t : t \in \mathbb{Z}\}$.

Proof. The existence of a corresponding spectral process follows from Proposition 3.1. Furthermore, it follows from Theorem 2.1 that the forward process $\{M_t : t \in \mathbb{N}_0\}$ is equal in law to the forward process of a BFTC(α). By Proposition 5.2 the statement follows.

Remark 5.2. Since the forward and backward tail chains of a process $\{X_t : t \in \mathbb{Z}\}$ are uniquely determined by the laws of (M_0, M_1) and (M_0, M_{-1}), respectively, it follows that the backward tail chain is equal in distribution to the forward tail chain if and only if the law of (M_0, M_1) is self-adjoint (compare Remark 4.2). This is, for example, the case if the process $\{X_t : t \in \mathbb{Z}\}$ is a time reversible Markov chain and fulfills the assumptions of Corollary 5.1.

More generally, since the existence of a forward tail process ensures joint regular variation of (X_0, X_1) (compare Basrak and Segers (2009, Corollary 3.2)), the resulting limiting spectral measure of the $2d$-dimensional vector $(X_{0,1}, \dots, X_{0,d}, X_{1,1}, \dots, X_{1,d})$ and the law of (M_0, M_1) uniquely determine each other. Therefore, the backward tail chain is equal in distribution to the forward tail chain if and only if the spectral measure of $(X_{0,1}, \dots, X_{0,d}, X_{1,1}, \dots, X_{1,d})$ is equal to the spectral measure of $(X_{1,1}, \dots, X_{1,d}, X_{0,1}, \dots, X_{0,d})$. For $d = 1$, this simply means that the spectral measure of (X_0, X_1) is symmetric.

In the univariate case, BFTCs have an additional structure which generalizes a multiplicative random walk in that the distribution of the increment depends on the sign of the process in its current state (Segers (2007)). The random walk structure of the forward tail chain was first observed in Smith (1992) for one-sided extremes and extended to allow for both positive and negative extremes in Bortot and Coles (2003).

6. Examples for BFTCs

We conclude the paper with some examples of BFTCs for multivariate Markov processes. For univariate examples; see Segers (2007, Section 7).

Example 6.1. Let $(A_t, B_t), t \in \mathbb{Z}$, be i.i.d. with $A_t \in \mathbb{R}^{d\times d}$ and $B_t \in \mathbb{R}^d$. The stationary distribution and asymptotic behavior of the corresponding random difference equation

$$X_t = A_t X_{t-1} + B_t, \qquad t \in \mathbb{Z}, \tag{6.1}$$

has been studied in the seminal work by Kesten (1973). Let us assume that the distribution of (A_t, B_t) satisfies the technical, but mild assumptions of Theorems A and B or Theorem 6 in Kesten (1973) (where the first two theorems deal with the nonnegative case, i.e. all components of $A_t, t \in \mathbb{Z}$, are nonnegative almost surely, and the last theorem treats the general case). Together with the results in Boman and Lindskog (2009) this implies that the stationary distribution of X_t for (6.1) is multivariate regularly varying in the nonnegative case. In the general case, multivariate regular variation follows if $\kappa_1 > 0$ in Kesten (1973, Equation (4.8)), is not an integer, compare Basrak *et al.* (2002a). Let Υ denote the spectral measure and $\alpha > 0$ the index of regular variation of the stationary distribution of X_t. It can be shown that

$$\mathbb{E}\left[f\left(\frac{AC}{\|AC\|}\right)\|AC\|^\alpha\right] = \mathbb{E}[f(C)] \tag{6.2}$$

for all bounded, continuous funtions f on $\mathbb{S}^{d-1}$, where $C \in \mathbb{S}^{d-1}$ has distribution Υ and $A \in \mathbb{R}^{d\times d}$ is independent of C with $\mathcal{L}(A) = \mathcal{L}(A_1)$; compare Basrak and Segers (2009).

Due to the linear structure of (6.1), Theorem 2.1 applies with $\mathbb{P}(Y > y) = y^{-\alpha}, y > 1, \mathcal{L}(M_0) = \Upsilon$ and $\phi(M_{j-1}, \varepsilon_j) = \varepsilon_j M_{j-1}$, where the $\varepsilon_j \in \mathbb{R}^{d\times d}, j = 1, 2, \dots,$ are i.i.d. with $\mathcal{L}(\varepsilon_j) = \mathcal{L}(A_1)$. In order to find the distribution of the backward tail chain we note that Remark 4.1 applies to this example by (6.2). So the law $\mathbb{P}^*$ of (M_0, M_{-1}) is given by

$$\mathbb{P}^*(E) = \mathbb{E}\left[\mathbf{1}_E\left(\frac{AC}{\|AC\|}, \frac{C}{\|AC\|}\right)\|AC\|^\alpha\right]$$

for all Borel sets $E \subset \mathbb{S}^{d-1} \times \mathbb{R}^d$.

Additional assumptions about $\mathcal{L}(A)$ allow us to simplify this characterization. Let us assume that A has a multiplicative form as in Example 4.1, i.e. $A = RQ$ for a positive random variable R with $\mathbb{E}[R^\alpha] = 1$ and Q is an orthogonal matrix independent of R. Additionally, we may assume that R has a density on $\mathbb{R}_+$ and that the support of the law of Q is equal to the orthogonal group in dimension d. In this case, the spectral measure Υ is the uniform distribution on $\mathbb{S}^{d-1}$ (compare Buraczewski *et al.* (2009, p. 390)), $\alpha > 0$ is the index of regular variation and

$$\mathbb{E}\bigg[f\bigg(\frac{AC}{\|AC\|}\bigg)\|AC\|^\alpha\bigg] = \mathbb{E}[f(QC)R^\alpha] = \mathbb{E}[f(QC)]\mathbb{E}[R^\alpha] = \mathbb{E}[f(C)]$$

holds for all bounded, continuous functions f on $\mathbb{S}^{d-1}$ with $C \sim \mathrm{Unif}(\mathbb{S}^{d-1})$. Since $\mathcal{L}(C) = \mathcal{L}(QC)$, all assumptions of Example 4.1 are met and the adjoint measure $\mathbb{P}^*$ is determined by (4.12) and equal to the law of $(C^*, R^*Q^*C^*)$ with R^*, Q^*, C^* independent, $\mathcal{L}(C^*) = \mathrm{Unif}(\mathbb{S}^{d-1})$, $\mathcal{L}(Q^*) = \mathcal{L}(Q')$, and R^* has density $f_{R^*}(y) = f_R(y^{-1})y^{-(2+\alpha)}$, $y > 0$, where f_R denotes the density of R. Thus, both the forward and the backward tail chains have a simple multiplicative structure:

$$M_t = M_0A_1 \cdot \ldots \cdot A_t, \qquad M_{-t} = M_0A_{-1} \cdot \ldots \cdot A_{-t}, \qquad t \geq 1,$$

with $A_1, A_2, \ldots$ as above and $A_{-1}, A_{-2}, \ldots$ i.i.d. with the same distribution as R^*Q^*, all independent of each other and of $M_0 \sim \mathrm{Unif}(\mathbb{S}^{d-1})$.

Example 6.2. While the preceding example dealt with random difference equations where the random increment B_t has a relatively light tail (Kesten (1973) assumes that $E(\|B_1\|^\alpha) < \infty$), the following example deals with AR(1) processes where the innovations themselve are regularly varying. Let

$$X_t = AX_{t-1} + B_t, \qquad t \in \mathbb{Z}, \tag{6.3}$$

where A is a deterministic $\mathbb{R}^{d\times d}$-matrix and $B_t \in \mathbb{R}^d$, $t \in \mathbb{Z}$, are i.i.d. and multivariate regularly varying with index $\alpha > 0$ and spectral measure λ on $\mathbb{S}^{d-1}$. For extensions to random but light-tailed random matrices A_t; see, for instance, Hult and Samorodnitsky (2008).

If $\sup_{x\in\mathbb{S}^{d-1}} \|A^m x\| < 1$ for some positive integer m then (6.3) has the stationary solution

$$X_t = \sum_{n=0}^{\infty} A^n B_{t-n}, \qquad t \in \mathbb{Z}.$$

It has been shown in Meinguet and Segers (2010) that in this case the stationary distribution of X_t is multivariate regularly varying as well, with the same index α and spectral measure $\Upsilon = \sum_{n=0}^{\infty} p_n\lambda_n$, where

$$p_n := \frac{c_n}{\sum_{k=0}^{\infty} c_k} \quad \text{with} \quad c_n := \int_{\mathbb{S}^{d-1}} \|A^n\theta\|^\alpha \lambda(\mathrm{d}\theta), \qquad n \in \mathbb{N}_0,$$

and where λ_n is the spectral measure of A^nB_1, provided $c_n > 0$, i.e.

$$\lambda_n(f) := \frac{1}{c_n}\int_{\mathbb{S}^{d-1}} f\bigg(\frac{A^n s}{\|A^n s\|}\bigg)\|A^n s\|^\alpha \lambda(\mathrm{d}s), \qquad n \in \mathbb{N}_0 \text{ if } c_n > 0,$$

for all bounded, continuous functions f on $\mathbb{S}^{d-1}$ (Meinguet and Segers (2010, Example 9.3)). The spectral process $\{M_t : t \in \mathbb{Z}\}$ in Proposition 3.1 is of the form

$$M_{-N+t} = \begin{cases} A^t\Theta, & t = 0, 1, 2, \ldots, \\ 0, & t = -1, -2, \ldots \end{cases} \tag{6.4}$$

https://doi.org/10.1239/jap/1421763332 Published online by Cambridge University Press

for a random integer N with $\mathbb{P}(N = n) = p_n$, $n \in \mathbb{N}_0$, and a random vector Θ with distribution

$$\mathbb{P}(\Theta \in E \mid N = n) = \frac{1}{c_n} \int_{\mathbb{S}^{d-1}} \mathbf{1}_E\biggl(\frac{s}{\|A^n s\|}\biggr) \|A^n s\|^\alpha \lambda(\mathrm{d}s)$$

for $n \in \mathbb{N}_0$ and Borel sets $E \in \mathbb{R}^d$. Here, the forward tail chain has a deterministic multiplicative structure with $M_0 \sim \Upsilon$ and $M_n = AM_{n-1}$ for $n \geq 1$. The backward process is Markovian as well, by Corollary 5.1. This is clear if we look at (6.4) and note that $M_{-(n+h)} = 0$ if $M_{-n} = 0$ for all $h \geq 1, n \geq 1$. Furthermore, if $M_{-n} \neq 0$ then $(M_{-n+1}, \ldots, M_0) = (AM_{-n}, \ldots, A^n M_{-n})$ contains no more information about $M_{-(n+1)}$ than M_{-n}.

The distribution of (M_0, M_{-1}) is adjoint to the one of $(M_0, M_1) = (M_0, AM_0)$. By (4.4) and since $M_0 \sim \Upsilon$, we find, for every Borel set $E \subset \mathbb{S}^{d-1} \times (\mathbb{R}^d \setminus \{0\})$,

$$\begin{aligned}\mathbb{P}((M_0, M_{-1}) \in E) &= \mathbb{E}\biggl[\mathbf{1}_E\biggl(\frac{M_1}{\|M_1\|}, \frac{M_0}{\|M_1\|}\biggr)\|M_1\|^\alpha\biggr] \\ &= \frac{1}{\sum_{k=0}^\infty c_k} \sum_{n \geq 0} \int_{\mathbb{S}^{d-1}} \mathbf{1}_E\biggl(\frac{A^{n+1}s}{\|A^{n+1}s\|}, \frac{A^n s}{\|A^{n+1}s\|}\biggr)\|A^{n+1}s\|^\alpha \lambda(\mathrm{d}s).\end{aligned}$$

Choosing $E = S \times (\mathbb{R}^d \setminus \{0\})$ for a Borel set $S \subset \mathbb{S}^{d-1}$ yields, upon taking complements with respect to $\{M_0 \in S\}$ and noting that $\|s\| = 1$ for $s \in \mathbb{S}^{d-1}$,

$$\mathbb{P}(M_0 \in S, M_{-1} = 0) = \frac{1}{\sum_{k=0}^\infty c_k} \lambda(S). \tag{6.5}$$

In particular, $\mathbb{P}(M_{-1} = 0) = p_0 = \mathbb{P}(N = 0)$. The backward tail chain now follows from Definition 5.1(iii) together with the distribution of (M_0, M_{-1}).

In the special case that A is invertible, we find from (6.4) that $M_{-(t+1)}$ is equal to either $A^{-1}M_{-t}$ or 0 with conditional probabilities depending on $M_{-t}/\|M_{-t}\|$: if $M_{-t} = 0$ then $M_{-(t+1)} = 0$ too, while if $M_{-t} = x \neq 0$ then

$$M_{-(t+1)} = \begin{cases} A^{-1}x & \text{with probability } 1 - \mathbb{P}\Bigl(M_{-1} = 0 \Bigm| M_0 = \frac{x}{\|x\|}\Bigr), \\ 0 & \text{with probability } \mathbb{P}\Bigl(M_{-1} = 0 \Bigm| M_0 = \frac{x}{\|x\|}\Bigr). \end{cases}$$

To derive a concrete form of the backward Markov kernel, let us assume that λ has a Lebesgue density f_λ on $\mathbb{S}^{d-1}$. Then all measures λ_n and, thus, Υ have Lebesgue densities as well and (6.5) gives

$$\mathbb{P}(M_{-1} = 0 \mid M_0 = s) = \frac{1}{\sum_{k=0}^\infty c_k} \frac{f_\lambda(s)}{f_\Upsilon(s)}$$

for all $s \in \mathbb{S}^{d-1}$ such that $f_\Upsilon(s) > 0$.

Acknowledgements

The authors thank Richard Davis and Holger Drees for helpful discussions. Furthermore, they wish to thank the organisers, especially Paul Doukhan, of the workshop ‘Extremes and risk management’ which took place during September 2012 at the university of Cergy-Pontoise. The authors would like to thank the anonymous referee for helpful comments and suggestions.

Anja Janßen was supported by DFG (DFG project JA 2160/1-1). Johan Segers was supported by contract 'Projet d'Actions de Recherche Concertées' number 12/17-045 of the 'Communauté française de Belgique' and by IAP research network grant number P7/06 of the Belgian government (Belgian Science Policy).

References

Basrak, B. and Segers, J. (2009). Regularly varying multivariate time series. *Stoch. Process. Appl.* **119,** 1055–1080. (Erratum: **121** (2011), 896–898.)

Basrak, B., Davis, R. A. and Mikosch, T. (2002a). A characterization of multivariate regular variation. *Ann. Appl. Prob.* **12,** 908–920.

Basrak, B., Davis, R. A. and Mikosch, T. (2002b). Regular variation of GARCH processes. *Stoch. Process. Appl.* **99,** 95–115.

Billingsley, P. (1968). *Convergence of Probability Measures*. John Wiley, New York.

Boman, J. and Lindskog, F. (2009). Support theorems for the Radon transform and Cramér–Wold theorems. *J. Theoret. Prob.* **22,** 683–710.

Bortot, P. and Coles, S. (2000). A sufficiency property arising from the characterization of extremes of Markov chains. *Bernoulli* **6,** 183–190.

Bortot, P. and Coles, S. (2003). Extremes of Markov chains with tail switching potential. *J. R. Statist. Soc. B* **65,** 851–867.

Buraczewski, D., Damek, E. and Mirek, M. (2012). Asymptotics of stationary solutions of multivariate stochastic recursions with heavy tailed inputs and related limit theorems. *Stoch. Process. Appl.* **122,** 42–67.

Buraczewski, D. *et al.* (2009). Tail-homogeneity of stationary measures for some multidimensional stochastic recursions. *Prob. Theory Relat. Fields* **145,** 385–420.

Coles, S. G., Tawn, J. A. and Smith, R. L. (1994). A seasonal Markov model for extremely low temperatures. *Environmetrics* **5,** 221–239.

Collamore, J. F. and Vidyashankar, A. N. (2013). Tail estimates for stochastic fixed point equations via nonlinear renewal theory. *Stoch. Process. Appl.* **123,** 3378–3429.

De Haan, L., Resnick, S. I., Rootzén, H. and De Vries, C. G. (1989). Extremal behaviour of solutions to a stochastic difference equation with applications to ARCH processes. *Stoch. Process. Appl.* **32,** 213–224.

De Saporta, B., Guivarc'h, Y. and Le Page, E. (2004). On the multidimensional stochastic equation $Y_{n+1} = A_n Y_n + B_n$. *C. R. Math Acad. Sci. Paris* **339,** 499–502.

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

Goldie, C. M. (1991). Implicit renewal theory and tails of solutions of random equations. *Ann. Appl. Prob.* **1,** 126–166.

Gomes, I. M., De Haan, L. and Pestana, D. (2004). Joint exceedances of the ARCH process. *J. Appl. Prob.* **41,** 919–926. (Correction: **41** (2004), 919–926.)

Hult, H. and Samorodnitsky, G. (2008). Tail probabilities for infinite series of regularly varying random vectors. *Bernoulli* **14,** 838–864.

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

Kifer, Y. (1986). *Ergodic Theory of Random Transformations*. Birkhäuser, Boston, MA.

Klüppelberg, C. and Pergamenchtchikov, S. (2003). Renewal theory for functionals of a Markov chain with compact state space. *Ann. Prob.* **31,** 2270–2300.

Klüppelberg, C. and Pergamenchtchikov, S. (2004). The tail of the stationary distribution of a random coefficient AR(q) model. *Ann. Appl. Prob.* **14,** 971–1005.

Letac, G. (1986). A contraction principle for certain Markov chains and its applications. In *Random Matrices and Their Applications* (Brunswick, Maine, 1984; Contemp. Math. **50**), American Mathematical Society, Providence, RI, pp. 263–273.

Meinguet, T. and Segers, J. (2010). Regularly varying time series in Banach spaces. Preprint. Available at http://arxiv.org/abs/1001.3262.

Mirek, M. (2011). Heavy tail phenomenon and convergence to stable laws for iterated Lipschitz maps. *Prob. Theory Relat. Fields* **151,** 705–734.

Perfekt, R. (1994). Extremal behaviour of stationary Markov chains with applications. *Ann. Appl. Prob.* **4,** 529–548.

Perfekt, R. (1997). Extreme value theory for a class of Markov chains with values in $\mathbb{R}^d$. *Adv. Appl. Prob.* **29,** 138–164.

Resnick, S. I. (2007). *Heavy-Tail Phenomena. Probabilistic and Statistical Modeling*. Springer, New York.

Resnick, S. I. and Zeber, D. (2013). Asymptotics of Markov kernels and the tail chain. *Adv. Appl. Prob.* **45,** 186–213.

SEGERS, J. (2007). Multivariate regular variation of heavy-tailed Markov chains. Discussion Paper 0703, Institut de statistique, Université catholique de Louvain. Available at http://uk.arxiv.org/abs/math/0701411.

SMITH, R. L. (1992). The extremal index for a Markov chain. *J. Appl. Prob.* **29,** 37–45.

SMITH, R. L., TAWN, J. A. AND COLES, S. G. (1997). Markov chain models for threshold exceedances. *Biometrika* **84,** 249–268.

VAN DER VAART, A. W. (1998). *Asymptotic Statistics*. Cambridge University Press.

YUN, S. (1998). The extremal index of a higher-order stationary Markov chain. *Ann. Appl. Prob.* **8,** 408–437.

YUN, S. (2000). The distributions of cluster functionals of extreme events in a dth-order Markov chain. *J. Appl. Prob.* **37,** 29–44.

ZIVOT, E. (2009). Practical issues in the analysis of univariate GARCH models. In *Handbook of Financial Time Series*, Springer, Berlin, pp. 113–155.