Technische Universität München

Zentrum Mathematik

# On Markov-Switching Models – Stationarity and Tail Behaviour

Diplomarbeit

von

Robert Josef Stelzer

| | |
|---|---|
| Themenstellerin: | Prof. Dr. Claudia Klüppelberg |
| Betreuer: | Dr. Alexander Lindner |
| Abgabetermin: | 15. Januar 2005 |

Hiermit erkläre ich, dass ich die Diplomarbeit selbstständig angefertigt und nur die angegebenen Quellen verwendet habe.

Buchbach, den 15. Januar 2005,

# Acknowledgements

Taking this opportunity I would like to thank all those people without whom my studies over the last years and especially this thesis would not have been what they have been.

First of all my thanks go to Prof. Claudia Klüppelberg for bringing the topic of this thesis to my attention, for all the helpful advice throughout the time I worked on it and for the excellent working conditions and climate. Moreover, I am most grateful to her for enabling me to spend a semester in Århus a year ago, to attend a MaPhySto course in Copenhagen and spend a week at the University of Århus last autumn and for the many other things throughout the last years.

Dr. Alexander Lindner has done a great job as supervisor of this thesis, for which I would like to thank him very much indeed. Moreover, I am heavily indebted to him for all the helpful advice and encouragement which he has given me over the years.

My special thanks for most helpful and influential discussions on the subjects of this thesis go to Dr. Vicky Fasen, Prof. Charles Goldie, Prof. Thomas Mikosch, Prof. Serguei Pergamenchtchikov and Prof. Gennady Samorodnitsky.

For introducing me into mathematical research and all the helpful things that I learned from the very interesting work together with them I am most grateful to Prof. Søren Asmussen and Prof. Ole Eiler Barndorff-Nielsen.

Moreover, I gratefully acknowledge the privilege of having been supported financially by the State of Bavaria via a BayBFG-scholarship throughout my studies.

Furthermore, a special "thank you very much" is due to all my friends who made life and, in particular, university life much more enjoyable.

Last but not at all least I would like to thank especially my family for simply everything and, of course, a very special "thank you very much" for everything is also due to Christine.

# Contents

# List of Abbreviations

| | |
|---|---|
| a.s. | almost sure |
| a.e. | almost everywhere |
| cf. | confer |
| ch. | chapter |
| cor. | corollary |
| def. | definition |
| e.g. | for example |
| eq. | equation |
| et al. | et alii |
| etc. | et cetera |
| f | and the following one |
| ff | and the following ones |
| i.e. | that is (id est) |
| iff | if and only if |
| i.i.d. | independent and identically distributed |
| p. | page |
| pp. | pages |
| prop. | proposition |
| resp. | respectively |
| rv | random variable |
| s.t. | such that |
| th. | theorem |
| viz. | namely, that is to say (vide licet) |
| w.l.o.g. | without loss of generality |
| w.r.t. | with respect to |

# Chapter 1

# Introduction

In almost all fields of applications, viz. the natural sciences, economics, engineering, medicine, the social sciences, etc., one is often confronted with time series, i.e. sequences of observations on some variables of interest over time. In order to understand the ongoing processes and to be able to make forecasts these have to be analysed and modelled. The penultimate goal, when analysing time series, is usually to find a tractable and parametric stochastic model describing the empirical observations well. In most cases the primary aim is to transform the observed values in such a way that the use of a linear time series model, i.e. an ARMA one (see e.g. Brockwell and Davis (1991)), is possible. A so-called ARMA(p,q) processes is defined via a white noise sequence $Z_t, t \in \mathbb{Z}$, and a set of parameters $\Phi_1, \ldots, \Phi_p, \Theta_1, \ldots, \Theta_q \in \mathbb{R}$. If a stochastic process $X_t$ satisfies

$$X_t - \Phi_1 X_{t-1} - \ldots - \Phi_p X_{t-p} = Z_t + \Theta_1 Z_{t-1} + \ldots + \Theta_q Z_{t-q}, \tag{1.1}$$

then it is called an ARMA(p,q) process.

When fitting some parametric model to empirical data, one is often confronted with the problem that one obtains rather excellent fits over short periods of time, but that the quality of the fits deteriorates considerably when increasing the time span. Moreover, when looking at plots of time series, one often, even with the bare eye, detects breakpoints, i.e. points of time, where the behaviour of the series changes. Using statistical tools one can actually find apparent structural breaks in many observational series.

The purpose of Markov-switching models is to provide a tractable model with nice probabilistic properties that over shorter time horizons behaves like a simple, e.g. an ARMA, model, but allows for structural breaks. Furthermore, such a model should allow for structural breaks at random times, since, if the behaviour has changed in the past, one should clearly allow for this to happen in the future as well in order to obtain realistic forecasts.

To exemplify what Markov-switching is all about, let us consider the so-called Markov-switching ARMA processes. Here the idea is in principle to use (1.1) to describe the evolution of a time series, but to allow the used parameters $\Phi_1, \ldots, \Phi_p, \Theta_1, \ldots, \Theta_q$ to change over time. In order to get a flexible, but still rather tractable model, one employs a Markov chain to describe the variation of the parameters over time. As realizations of such processes look like ones of an ARMA process when considering short time horizons and the parameters change in a Markovian way, the term "Markov-switching ARMA" processes

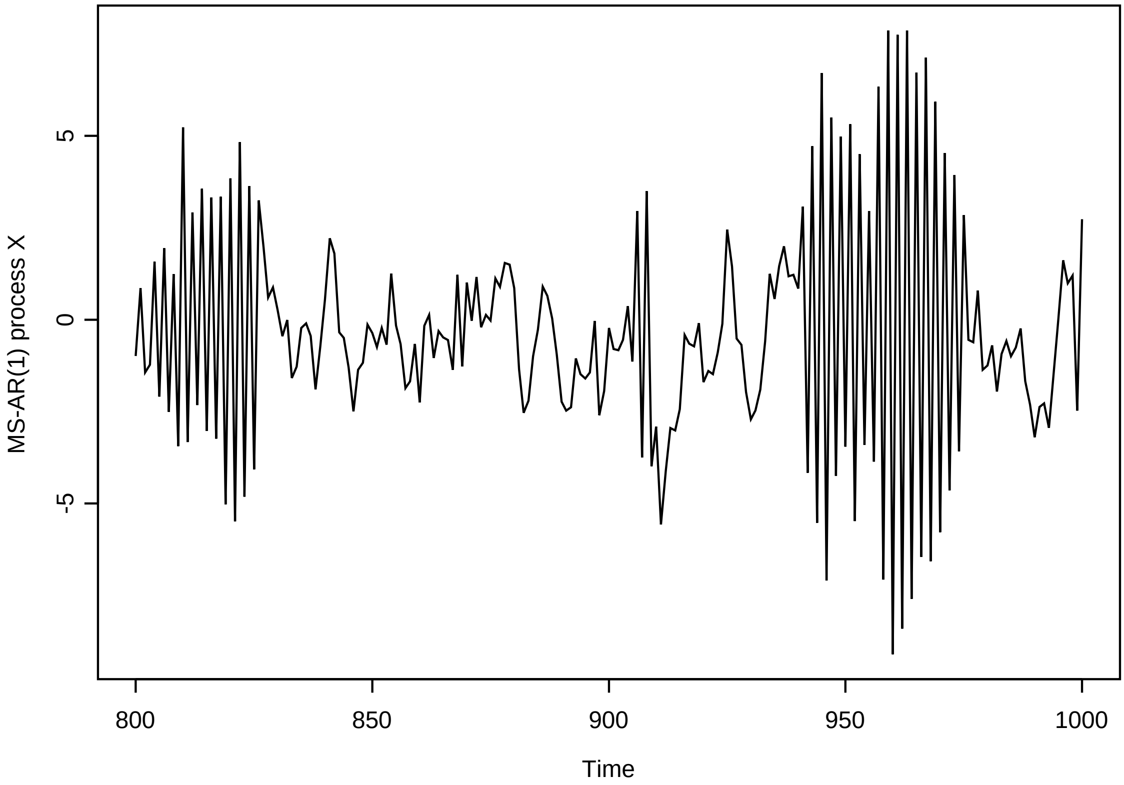


Figure 1.1: MS-AR(1) process $X_t = \Phi_{1t} X_{t-1} + \epsilon_t$ with two regimes (i.e. possible parameter values for $\Phi_{1t}$) $\Phi^{(1)} = -0.9$ and $\Phi^{(2)} = 0.8$. The probability of remaining in the current regime is 0.95 throughout and $\epsilon_t \sim N(0,1)$ is an i.i.d. sequence. As is shown in Section 5.4.1 the process is actually stationary, although the path behaviour changes dramatically at regime shifts. The time horizon considered is 800 to 1000 in order to ensure stationarity.

arises naturally as a name for such models. If the parameter sets for the different times are actually an i.i.d. sequence, the term "random coefficient ARMA" is used to stress that this is a rather special case.

Figure 1.1 depicts a simulation of an MS-AR(1) process. One can clearly observe the structural breaks in the path behaviour due to switches in the driving parameter chain from one state to the other. In between parameter changes the path looks like paths of an AR(1) process corresponding to the respective parameter values. The times of the switches were 826, 867, 873, 903, 909, 941, 979 and 998 and the original parameter at time 800 was $\Phi^{(1)} = -0.9$.

When intending to use GARCH models, one is often confronted with the same problem that the fit over short periods is very good, but the data seems to exhibit structural breaks. Thus, Markov-switching GARCH models enter the scene. A Markov-switching GARCH process is basically again described by the usual GARCH equations, but the parameters change over time in a Markovian way.

Let us now give a brief overview on the results on and use of Markov-switching models in the existing literature. The first area where Markov-switching ARMA processes appear to have been used extensively is electrical engineering (see Tugnait (1982) or Doucet, Logothetis and Krishnamurthy (2000) and references therein), where the first papers on this topic appeared in the 1970s. The main emphasis is on inference and estimation there. Also in the 1970s the interest in random coefficient autoregressions arouse in statistics, as can be seen from the monograph Nicholls and Quinn (1982). In econometrics the two papers Hamilton (1989) and Hamilton (1990) started a whole industry of papers using Markov-switching ARMA models to fit observational series from more or less all the different fields of economics. For an overview over the early work see Hamilton (1994). Krolzig (1997) contains a lot of useful information on Markov-switching processes, the related statistical tools used in econometrics and some detailed studies of actually observed time series. For some recent applications see e.g. Hamilton and Raj (2002). In the statistical literature the theoretical properties of maximum likelihood estimators in Markov-switching models like consistency were first discussed in Francq and Roussignol (1998) and Krishnamurthy and Rydén (1998), who both considered driving chains with finite state space. Only very recently the case of a possibly uncountable, yet compact state space was addressed in Douc, Moulines and Rydén (2004). References to probabilistic work on MS-ARMA and related processes will be mentioned later when studying these processes.

It is noteworthy that the Markov-switching ARMA models we shall consider later fall into the hidden Markov model framework as analysed in the recent paper by Fuh (2004), for example. However, they are more general than the more classical Hidden Markov Models (HMM) as discussed e.g. in Poskitt and Chung (1996) and the references given there.

The use of MS-ARCH models has begun in the econometric papers Cai (1994) and Hamilton and Susmel (1994). General GARCH models have then been studied in Gray (1996) and Dueker (1997). Wong and Li (2001) consider GARCH models with i.i.d. coefficients and Klaassen (2002), Rossi and Gallo (2003) and Haas, Mittnik and Paolella (2004) various different Markov-switching GARCH formulations. Consistency of the maximum likelihood estimator has been considered in Francq, Roussignol and Zakoïan (2001) and we shall study an extension of the Markov-switching GARCH specification given there later.

In this thesis we intend to study Markov-switching processes from a theoretical and probabilistic point of view. The previously known results are rather scattered over the literature and there seems to be no comprehensive treatment on the theoretical probabilistic properties so far. In particular, two somewhat different lines of research have been established. On the one hand Markov-switching models with driving chains of finite state space have been considered by econometricians, statisticians and engineers and employed to model various time series, whereas probabilists on the other hand have studied either random coefficient models or general stochastic recurrence equations with i.i.d. or ergodic input. As far as possible, we try to unite the two lines by studying Markov-switching models with a driving Markov chain that may have a non-finite and even uncountable state space.

The outline of the thesis is as follows. Markov-switching models are not considered before Chapters 5 and 6, as we first develop all the necessary tools for the analysis in order to be able to give a streamlined account on Markov-switching ARMA and GARCH processes.

Chapter 2, in particular, summarizes the employed notions and notations and provides many probabilistic results needed later. Especially Sections 2.1- 2.3 are just rather short overviews. In Section 2.4 we develop the theory of general $L^p$-spaces, i.e. spaces of functions assuming values in some normed space and being $p$-times integrable. This is a topic that is rarely found in the literature, but it is essential for our later studies. Section 2.5 is on ergodicity in the general measure theoretic meaning and related concepts. It lays the necessary foundations for analysing stationarity of Markov-switching processes. Finally, we introduce the necessary notions from general classical Markov chain theory and the concept of strong mixing of stochastic processes in Section 2.6. The results we report there will only be used in Section 5.5. Moreover, the implications of strong mixing for extreme value theory are briefly recalled.

In the third chapter we give first a short account on vague convergence of measures, as we shall employ this notion in order to define multivariate regular variation of random variables and sequences in general. In the second section of this chapter we introduce and analyse several notions of regular variation, which is the essential tool we shall use later to study the tail behaviour. It is noteworthy that the general definition we give and most of the reported results are rather recent. The last section summarizes results on certain combinations and transformations of regularly varying random variables. The very last theorem is a new extension of previously known results and later enables us to analyse regular variation of certain processes with regularly varying noise.

The fourth chapter is on the stochastic difference equation $Y_n = A_n Y_{n-1} + C_n$ with stationary and ergodic input $(A_n, C_n)$. Formulating Markov-switching processes as such a random recurrence equation will be a key step in their probabilistic analysis later on. We first review stationarity and ergodicity results. Then we study the finiteness of moments and conclude by analysing the tail behaviour, where we focus on cases in which regular variation appears in the tails. Our results on the finiteness of moments for such a stochastic recurrence equations in more than one dimension and the analysis of the tail behaviour in the presence of a regularly varying noise, that is not restricted to be non-negative in all components, have, to the best of our knowledge, not been considered elsewhere yet.

Markov-switching ARMA models are considered in detail in Chapter 5. We start by giving a definition of Markov-switching ARMA processes driven by a general state space Markov chain. After a discussion of sufficient stationarity conditions we give sufficient conditions ensuring the existence of moments in Section 5.2. The results on driving Markov chains with general state space and on general moments are not to be found in the existing literature, which only allowed for a finite state space and focused mainly on the first and second order moments. Thereafter we discuss feasible ways of checking the previous conditions in Section 5.3. The norm condition, in particular, seems to be new. The following Section 5.4 explores the relationship between the stationarity of the Markov-switching ARMA process and the stationarity of the ARMA processes related to the individual parameter sets possible. Previously only second order stationarity issues were considered in this respect. In Section 5.5 we give criteria, for when a Markov-Switching

ARMA process is geometrically ergodic, and in the last section of this chapter we again turn to the tail behaviour and examine several cases in which regularly varying tails show up. The general geometric ergodicity results and the ones on the tail behaviour in the presence of a regularly varying noise are again new. Some simulations of real-valued MS-AR(1) processes are to be found in Sections 5.6.1 and 5.6.2.

Finally, we conclude the thesis by looking at Markov-switching GARCH models in Chapter 6. Again we give a definition of such processes driven by general state space Markov chains and a short discussion of sufficient conditions for stationarity and the finiteness of moments.

# Chapter 2

# Preliminaries

In the following we introduce briefly the notation used and give expositions of some general concepts and results employed later in the thesis.

## 2.1 General Notions and Notation

The natural logarithm is abbreviated as log and empty sums, i.e. sums like $\sum_{k=0}^{-1} a_i$, are defined to be equal to the zero element in the appropriate space obvious from the context. Likewise, empty products, i.e. products like $\prod_{k=0}^{-1} a_i$, are understood to be equal to the unit of the appropriate algebra. The same convention is used for products of the form $a_t a_{t-1} \cdots a_{t-k}$ with $k < 0$.

### 2.1.1 Set Operations

$I_A(\cdot)$ stands for the indicator function of some set $A$. The symmetric difference of two sets $A$ and $C$ is as usually defined as $A \Delta C := (A \backslash C) \cup (C \backslash A)$. Algebraic manipulations of sets also have to be understood in their usual meaning, for example, if $A, C$ are two sets, then $A + C = \{a + c : a \in A, c \in C\}$ and $AC = \{ac : a \in A, c \in C\}$ (with the multiplication operation that is obvious from the context).

### 2.1.2 Algebras

The term "algebra" is used both in its algebraic (cf. e.g. Heuser (1992, p.113)) and measure theoretic (see below) definition. From the context it is in the following obvious which definition we refer to. In a unital algebra (in the algebraic sense) $I$ denotes the unit element. Recall that in measure theory one defines an algebra in the following way:

**Definition 2.1 (cf. Bauer (1992, Satz 1.4))** *Let $\Omega$ be a set. A system $\mathscr{A}$ of subsets of $\Omega$ is called* algebra*, if*

$$\begin{aligned} \Omega &\in \mathscr{A}, \\ A \in \mathscr{A} &\Rightarrow A^c \in \mathscr{A} \text{ and} \\ A, B \in \mathscr{A} &\Rightarrow A \cup B \in \mathscr{A}. \end{aligned}$$

As usual we define the complement $A^c$ of some set $A$ as $\Omega\backslash A$. The difference between an algebra and a $\sigma$-algebra is that the later contains countable unions, whereas an algebra only contains finite unions.

### 2.1.3 Expectation Operator

The expectation operator $E(\cdot)$ always operates solely on its argument inside the bracket, which means that e.g. $E(Y)^p$ is understood to mean $(E(Y))^p$. In contrast to this the occasionally used $EY^p$ corresponds to $E(Y^p)$ and $E(Y^p)^r$ to $(E(Y^p))^r$.

### 2.1.4 Product Measures

Suppose that $(E_1, \mathcal{E}_1)$ and $(E_2, \mathcal{E}_2)$ are two measurable spaces and that we have a measure $\nu$ on $(E_1 \times E_2, \sigma(\mathcal{E}_1 \times \mathcal{E}_2)$ that is the product of the two measures $\nu_1$ on $(E_1, \mathcal{E}_1)$ and $\nu_2$ on $(E_2, \mathcal{E}_2)$, then we either use the notation $\nu = \nu_1 \otimes \nu_2$ or a symbolic $dx$ notation, viz. $\nu(dx_1, dx_2) = \nu_1(dx_1)\nu_2(dx_2) = \nu_2(dx_2)\nu_1(dx_1)$, which is sometimes more convenient. Note that the order in which the marginal measures appear in the $dx$ notation does not matter. It should be obvious, how this notation generalizes to products of more than two spaces. In the section on multivariate regular variation we shall consider in particular $\mathbb{R}^{kd}$ and view it as the $k$-fold product of $\mathbb{R}^d$. So a symbolic notation like $\nu(dx_1, dx_2, \ldots, dx_k) = \nu_2(dx_2)\epsilon_{0_{\mathbb{R}^{(k-1)d}}}(dx_1, dx_3, \ldots, dx_k)$, where $\epsilon_0$ is the Dirac measure w.r.t. 0, is equivalent to $\nu = \epsilon_{0_{\mathbb{R}^d}} \otimes \nu_2 \otimes \underbrace{\epsilon_{0_{\mathbb{R}^d}} \otimes \ldots \otimes \epsilon_{0_{\mathbb{R}^d}}}_{(k-2)\ \text{factors}}$.

## 2.2 Vector Spaces, Norms, Metrics, Linear Operators and Matrices

The $m \times n$ matrices over $\mathbb{R}$ ($\mathbb{C}$) will be denoted by $M_{m,n}(\mathbb{R})$ ($M_{m,n}(\mathbb{C})$). If $n = m$, we will use the notation $M_n(\mathbb{R})$ ($M_n(\mathbb{C})$) and $M_n^+(\mathbb{R})$ ($M_n^+(\mathbb{C})$) for the symmetric positive semidefinite matrices. The unit in $M_n$ is denoted by $I_n$. The adjoint (transpose in the real case) of a matrix $M$ will be written as $M^\mathsf{T}$. All matrices defining our models in the chapters to follow will be in $M_n(\mathbb{R})$. However, some manipulations (Schur decomposition etc.) may lead to complex valued matrices. Thus, when necessary, $M_{m,n}(\mathbb{R})$ is simply regarded as a subset of $M_{m,n}(\mathbb{C})$ and identified with linear operators acting on complex vector spaces. This is in particular done when considering the spectrum $\sigma(M)$ of a matrix $M$ and its spectral radius $\rho(M)$.

$\mathbb{R}^d$ is always thought to be equipped with a norm $\|\cdot\|$ and $\|\cdot\|$ will also denote the induced operator norm on $M_d(\mathbb{R})$. Standard norms used on $\mathbb{R}^d$ ($\mathbb{C}^d$) are $\|\cdot\|_1 : (x_1, \ldots, x_d)^\mathsf{T} \mapsto \sum_{i=1}^d |x_i|$, $\|\cdot\|_2 : (x_1, \ldots, x_d)^\mathsf{T} \mapsto \sqrt{\sum_{i=1}^d |x_i|^2}$ and $\|\cdot\|_\infty : (x_1, \ldots, x_d)^\mathsf{T} \mapsto \max_{1 \le i \le d} |x_i|$.

On a finite product $X = \prod_{i=1}^n X_i$ of normed spaces $(X_i, \|\cdot\|_i)$ the norm is understood to be $\|\cdot\| : X \to \mathbb{R}^+, (x_i)_{i=1,\ldots,n} \mapsto \sum_{i=1}^n \|x_i\|_i$, if not indicated otherwise.

All vector spaces are understood to be $\mathbb{R}$-linear or, sometimes, $\mathbb{C}$-linear spaces and the dual space of a topological vector space $X$ is $X^*$.

Recall that two normed spaces $X$ and $Y$ are isomorphic, if there is a bijective linear map $T : X \to Y$ and constants $M \geq m > 0$ such that $m\|x\| \leq \|Tx\| \leq M\|x\| \,\forall x \in X$ (cf. Werner (2002, Def. II.1.9)). $T$ is referred to as an isomorphism. This concept is generalized to metric spaces in the following definition:

**Definition 2.2** *Two metric vector spaces $(X, d)$ and $(Y, d)$ are said to be isomorphic, if there is an isomorphism $T : X \to Y$ and constants $M \geq m > 0$ such that $md(x_1, x_2) \leq d(Tx_1, Tx_2) \leq Md(x_1, x_2)$ for all $x_1, x_2 \in X$. They are said to be isometric, if $m = M = 1$.*

Using the $\epsilon - \delta$ characterization for the continuity of functions between metric spaces, it is clear that a mapping $T$ with the above given properties and its inverse are continuous.

Furthermore, analogously to the familiar definition of a semi-norm (see e.g. Werner (2002, p. 1)), we call any function $d : X \times X \to [0, \infty)$ (where $X$ is some set) a semi-metric, if it has all properties of a metric except that $d(x_1, x_2) = 0$ does not necessarily imply $x_1 = x_2$ (cf. Werner (2002, p. 472)). The topology induced by a semi-metric does in general not have the Hausdorff property.

In any metric space $(X, d)$ we set, as usual, $B_\delta(x) = \{y \in X : d(x, y) < \delta\}$ for the open ball with radius $\delta > 0$ around $x \in X$.

## 2.3 Random Variables and Processes

Throughout this thesis we assume the existence of a probability space $(\Omega, \mathcal{F}, P)$, where all occurring random variables are defined on, i.e., for instance, all $\mathbb{R}^d$ valued random variates are $\mathcal{F} - \mathcal{B}^d$-measurable mappings from $\Omega$ into $\mathbb{R}^d$, where $\mathcal{B}^d$ denotes the usual Borel-$\sigma$-algebra on $\mathbb{R}^d$. Likewise random variables taking values in $M_{m,n}(\mathbb{R})$ are $\mathcal{F} - \mathcal{B}^{m,n}$-measurable mappings from $\Omega$ into $M_{m,n}(\mathbb{R})$, where $\mathcal{B}^{m,n}$ is the Borel-$\sigma$-algebra on $M_{m,n}(\mathbb{R})$ induced by any norm. Recall that all norms induce the same topologies and thus the same Borel-$\sigma$-algebras on finite dimensional spaces. Equality in law (distribution) of two random variables $X$ and $Y$ is denoted by $\overset{\mathcal{D}}{=}$ and convergence of a sequence of random variables $(X_n)_{n\in\mathbb{N}}$ in law to a random variable $X$ by $X_n \xrightarrow{\mathcal{D}} X$.

Since our random variables often take values in a product of measurable spaces, let us give a brief reminder on general product spaces. Let $I$ be an arbitrary index set and $\{(\Omega_i, \mathcal{F}_i)\}_{i\in I}$ be a family of measurable spaces. Then the product space $\prod_{i\in I} \Omega_i$ is understood to be equipped with the product $\sigma$-algebra $\prod_{i\in I} \mathcal{F}_i$, which is the smallest $\sigma$-algebra on the product space, such that all projections $\pi_n : \prod_{i\in I} \Omega_i \to \Omega_n, (\omega_i)_{i\in I} \mapsto \omega_n, n \in I$, are measurable. As is well known from basic measure theory, $\prod_{i\in I} \mathcal{F}_i$ is generated by the cylinder sets $\mathscr{Z}_{\{\mathcal{F}_i\}_{i\in I}} = \left\{(\pi_{i_1}, \ldots, \pi_{i_n})^{-1}(A) : i_1, \ldots, i_n \in I, A \in \prod_{j=1}^n \mathcal{F}_{i_j}, n \in \mathbb{N}\right\}$ (note that $\prod_{j=1}^n \mathcal{F}_{i_j} = \sigma\left(\left\{\prod_{j=1}^n C_{i_j} : C_{i_j} \in \mathcal{F}_{i_j}\right\}\right)$) as well as by the rectangular sets $\mathscr{R}_{\{\mathcal{F}_i\}_{i\in I}} = \left\{\pi_{i_1}^{-1}(A_1) \cap \ldots \cap \pi_{i_n}^{-1}(A_n) : i_1, \ldots, i_n \in I, A_1 \in \mathcal{F}_{i_1}, \ldots, A_n \in \mathcal{F}_{i_n}, n \in \mathbb{N}\right\}$ (cf. Loève (1977, part 1, 1.7 and 4.2), Shiryaev (1996, ch. II §3) or Bauer (2001, § 9)).

For some set $E$ equipped with a $\sigma$-algebra $\mathcal{E}$ an $E$-valued random process on $\mathbb{Z}$ (i.e. in discrete time) is a sequence $Z = (Z_t)_{t\in\mathbb{Z}}$ of random variables defined on a measure space $(\Omega, \mathcal{F}, P)$ assuming values in the measurable space $(E, \mathcal{E})$. $Z$ can also be regarded as a measurable mapping from $(\Omega, \mathcal{F})$ to $(E^{\mathbb{Z}}, \mathcal{E}^{\mathbb{Z}})$. An immediate consequence of the definition of the product of measurable spaces is that $Z : \Omega \to E^{\mathbb{Z}}$ is $\mathcal{F} - \mathcal{E}^{\mathbb{Z}}$-measurable, iff $Z_t : \Omega \to E$ is $\mathcal{F} - \mathcal{E}$-measurable for all $t \in \mathbb{Z}$. Furthermore, the distribution of $Z$ is determined by its finite-dimensional marginals. Thus, the two ways of looking at random sequences are equivalent. Confer any standard text on probability theory (e.g. Shiryaev (1996) or Bauer (2001)) for the details.

In time series analysis the back-shift operator $B$ is usually the map sending $X_t$ to $X_{t-1}$. However, we shall use the analytically convenient definition usually employed especially in connection with ergodic theory as a linear operator acting on the whole sequence. So, let $(E, \mathcal{E})$ be some measurable space, then we define $B$ as the right-shift operator on $E^{\mathbb{Z}}$, i.e. $B : E^{\mathbb{Z}} \to E^{\mathbb{Z}}, (z_i)_{i \in \mathbb{Z}} \mapsto (z_{i-1})_{i \in \mathbb{Z}}$. The rationale behind this definition is that we get $\pi_t(X) = X_t$ and $\pi_t(BX) = X_{t-1}$ for any $t \in \mathbb{Z}$ for an $E$-valued random sequence $X$, where $\pi_t$, $t \in \mathbb{Z}$, are the coordinate projections. It is noteworthy that $B$ is bijective and $B^{-1}$ is the left-shift operator, i.e. the forward-shift operator in a time series context. Furthermore, $B$ as well as $B^{-1}$ are obviously measurable and map cylinder sets to cylinder sets.

A random sequence $X = (X_i)_{i \in \mathbb{Z}}$ assuming values in a measurable space $(E, \mathcal{E})$ is called stationary, if the distributions of $(X_{t_1}, \ldots, X_{t_k})$ and $(X_{t_1-h}, \ldots, X_{t_k-h})$ are the same for all $t_1, \ldots, t_k, h \in \mathbb{Z}$ and $k \in \mathbb{N}$. This is obviously equivalent to $P(X \in A) = P(B^h X \in A)$ for all $h \in \mathbb{N}$ and $A \in \mathscr{Z}_{\mathcal{E}^{\mathbb{Z}}}$ or $A \in \mathscr{R}_{\mathcal{E}^{\mathbb{Z}}}$, since both the cylinder and rectangular sets are closed under intersection (cf. Brandt, Franken and Lisek (1990, A 1) who use the left-shift operator).

## 2.4 General $L^p$-spaces

The theory of the spaces of $p$-integrable real valued random variables, which can be found in most textbooks on measure and probability theory (e.g. Loève (1977), Bauer (1992), Shiryaev (1996)) or functional analysis (e.g. Werner (2002)), can be extended to random variables assuming values in an arbitrary normed space (see e.g. Dunford and Schwartz (1958)). The following brief discussion of this generalized theory presumes familiarity with the usual $L^p$ spaces of random variables assuming values in $\mathbb{R}$ at the level of Loève (1977) and the basic concepts of functional analysis to be found e.g. in Werner (2002). We only consider probability measures, but extensions to more general measures are obvious. Throughout this section $(\Omega, \mathcal{F}, P)$ is some fixed probability space.

The definitions and results below are later useful when studying the moments of solutions to a stochastic difference equation and Markov switching models. Although, we only consider $\mathbb{R}^d$ or $M_d(\mathbb{R})$ valued random variables in those sections, we consider general normed vector spaces here, since it seems worthwhile to give a rather general and comprehensive discussion, as this is rarely to be found in the literature. Moreover, most results and their proofs do not get any simpler by restricting attention to $\mathbb{R}^d$ and $M_d(\mathbb{R})$.

**Definition 2.3** *Let $(X, \|\cdot\|)$ be a normed space, $p > 0$ and $\mathcal{B}$ the Borel $\sigma$-algebra of $(X, \|\cdot\|)$.*

*Then*

$$\begin{aligned}\mathcal{L}^p(\Omega,\mathcal{F},P,X,\|\cdot\|) &:= \left\{Y:\Omega\to X:\ Y\ \mathcal{F}\text{-}\mathcal{B}\text{-}measurable,\ \int_\Omega \|Y(\omega)\|^p dP(\omega)<\infty\right\},\\ \mathcal{L}^\infty(\Omega,\mathcal{F},P,X,\|\cdot\|) &:= \{Y:\Omega\to X:\ Y\ \mathcal{F}\text{-}\mathcal{B}\text{-}measurable, \exists K<\infty\ s.t.\ \|Y\|\le K\ a.s.\}.\end{aligned}$$

*Moreover, define* $\mathcal{N}=\{Y:\Omega\to X:\ Y\ \mathcal{F}\text{-}\mathcal{B}\text{-}measurable, \|Y\|=0\ a.s.\}$.

Note that, since we only use the norm of an $X$-valued random variable in the above definition, all integrands are real valued and we thus do not have to discuss a notion of integrating functions which assume values in a general normed space (see Dunford and Schwartz (1958) for a thorough discussion on integrating Banach space valued functions).

The following lemma, which is obviously implied by the definition, allows us to deduct most properties of the above defined spaces from the properties of the well-known special case with $X=\mathbb{R}$.

**Lemma 2.4** *Let* $(X,\|\cdot\|)$ *be a normed space,* $p\in(0,\infty]$ *and* $Y$ *be an* $X$ *valued random variable. Then* $Y\in\mathcal{L}^p(\Omega,\mathcal{F},P,X,\|\cdot\|)\Leftrightarrow\|Y\|\in\mathcal{L}^p(\Omega,\mathcal{F},P,\mathbb{R})$*. Furthermore,* $Y\in\mathcal{N}\Leftrightarrow$ $Y=0$ *a.s.*

Since $\mathcal{L}^p(\Omega,\mathcal{F},P,\mathbb{R})$ is a vector space, the above lemma and the properties of a norm imply:

**Proposition 2.5** $\mathcal{L}^p(\Omega,\mathcal{F},P,X,\|\cdot\|)$ *is a linear space and* $\mathcal{N}$ *is a subspace of* $\mathcal{L}^p(\Omega,\mathcal{F},P,X,\|\cdot\|)$ *for all* $p\in(0,\infty]$.

As in the classical situation this enables us to define $L^p$ spaces:

**Definition 2.6** *Let* $(X,\|\cdot\|)$ *be a normed space. Define for* $p\in(0,\infty]$

$$L^p(\Omega,\mathcal{F},P,X,\|\cdot\|):=\mathcal{L}^p(\Omega,\mathcal{F},P,X,\|\cdot\|)/\mathcal{N}, \tag{2.1}$$

*for* $Y,Z\in\mathcal{L}^p(\Omega,\mathcal{F},P,X,\|\cdot\|)$

$$\|Y\|_{L^p}=E(\|Y\|^p)^{1/p}\quad for\quad 1\le p<\infty, \tag{2.2}$$
$$\|Y\|_{L^\infty}=\inf\{K\in\mathbb{R}^+:\|Y\|\le K\ a.s.\}\quad for\quad p=\infty, \tag{2.3}$$
$$d_{L^p}(Y,Z)=E(\|Y-Z\|^p)\quad for\quad p<1 \tag{2.4}$$

*and for equivalence classes* $[Y],[Z]\in L^p(\Omega,\mathcal{F},P,X,\|\cdot\|)$

$$\|[Y]\|_{L^p}=\|Y\|_{L^p}\quad for\quad 1\le p\le\infty, \tag{2.5}$$
$$d_{L^p}([Y],[Z])=d_{L^p}(Y,Z)\quad for\quad p<1. \tag{2.6}$$

Moreover, Lemma 2.4 shows together with the inclusion relations for the $L^p(\Omega,\mathcal{F},P,\mathbb{R},|\cdot|)$ spaces the inclusion relations for $L^p$ spaces over general normed spaces:

**Corollary 2.7** *Let* $(X,\|\cdot\|)$ *be a normed vector space. Then*

$$\begin{aligned}\mathcal{L}^s(\Omega,\mathcal{F},P,X,\|\cdot\|) &\subseteq \mathcal{L}^r(\Omega,\mathcal{F},P,X,\|\cdot\|)\\ L^s(\Omega,\mathcal{F},P,X,\|\cdot\|) &\subseteq L^r(\Omega,\mathcal{F},P,X,\|\cdot\|)\end{aligned}$$

*for* $0<r\le s\le\infty$*. In particular, if* $Y\in L^s$ *for* $s\ge 1$ *then* $\|Y\|_{L^r}\le\|Y\|_{L^s}$ *for all* $r\in[1;s]$.

Note that despite the above inclusion relation it is still possible to have $E(\|Y\|^r) > E(\|Y\|^s)$, if $Y \in L^s$, $s \in (0,1]$ and $r \in (0,s)$. So the second part of the above corollary can not be generalized to include $s < 1$.

In the following we briefly analyse the analytic properties of the above defined $L^p$ spaces.

**Proposition 2.8** $\|\cdot\|_{L^p}$ *is a semi-norm on* $\mathcal{L}^p(\Omega, \mathcal{F}, P, X, \|\cdot\|)$ *and a norm on* $L^p$ *for* $1 \leq p \leq \infty$. *In the case* $0 < p < 1$ $d_{L^p}(\cdot,\cdot)$ *is a semi-metric on* $\mathcal{L}^p$ *and a metric on* $L^p$.

As usual we will not distinguish between random variables in $\mathcal{L}^p$ and the corresponding equivalence classes in $L^p$. Hence, we will from now on only use the symbol $L^p$. However, one should bear in mind that $\mathcal{L}^p$ with the above defined semi-norm/metric is in contrast to $L^p$ not a Hausdorff topological vector space and limits are only unique up to a.s. identity.
**Proof:** Apart from the triangle (Minkowski) inequalities all properties of a (semi)-norm, resp. -metric are obvious. The triangle inequality follows from the triangle inequalities in $L^p(\Omega, \mathcal{F}, P, \mathbb{R}, |\cdot|)$: Let $V, Y, Z \in L^p(\Omega, \mathcal{F}, P, X, \|\cdot\|)$, then

$$\begin{aligned} \|Z+Y\|_{L^p} &= E(\|Z+Y\|^p)^{1/p} \leq E((\|Z\| + \|Y\|)^p)^{1/p} \leq E(\|Z\|^p)^{1/p} + E(\|Y\|^p)^{1/p} \\ &= \|Z\|_{L^p} + \|Y\|_{L^p} \end{aligned}$$

for $1 \leq p < \infty$, respectively,

$$\begin{aligned} d_{L^p}(Y,Z) &= E(\|Y-Z\|^p) \leq E((\|Y-V\| + \|V-Z\|)^p) \\ &\leq E(\|Y-V\|^p) + E(\|V-Z\|^p) = d_{L^p}(Y,V) + d_{L^p}(V,Z) \end{aligned}$$

for $0 < p < 1$. For $p = \infty$ observe that, if $Y$ is a.s. bounded by $K$ and $Z$ by $K'$, $Z + Y$ is a.s. bounded by $K + K'$. □
In the classical set-up $L^2$ is a Hilbert space. This can be reproduced in higher dimensions:

**Theorem 2.9** *Let* $(X, \langle\cdot,\cdot\rangle)$ *be an inner product space (with* $\|\cdot\|$ *denoting the canonical norm). On* $L^2(\Omega, \mathcal{F}, P, X, \|\cdot\|)$ *define a bilinear form* $\langle\cdot,\cdot\rangle_{L^2}$ *by setting* $\langle Y, Z\rangle_{L^2} := E(\langle Y, Z\rangle)$. *Then* $\langle\cdot,\cdot\rangle_{L^2}$ *is a scalar product and induces the norm* $\|\cdot\|_{L^2}$. *Moreover, for all* $Y, Z \in L^2(\Omega, \mathcal{F}, P, X, \|\cdot\|)$

$$|\langle Y, Z\rangle_{L^2}| \leq E(\|Y\|\|Z\|) \leq \|Y\|_{L^2}\|Z\|_{L^2}.$$

**Proof:** It suffices to show the last two inequalities. This shows that $\langle\cdot,\cdot\rangle_{L^2}$ is well-defined. That $\langle\cdot,\cdot\rangle_{L^2}$ has all properties of a scalar product and induces the $L^2$-norm can be seen immediately. But the left inequality is simply the Cauchy Schwarz inequality in $(X, \langle\cdot,\cdot\rangle)$ and the right inequality is the Cauchy-Schwarz (Hölder) inequality in $L^2(\Omega, \mathcal{F}, P, \mathbb{R}, |\cdot|)$, since $\|Y\|_{L^2(\Omega,\mathcal{F},P,X,\|\cdot\|)} = \|\|Y\|\|_{L^2(\Omega,\mathcal{F},P,\mathbb{R},|\cdot|)}$. □
The following two theorems give natural extensions of the Hölder inequality to higher dimensional spaces. One comes from viewing the real (complex) numbers as an algebra, the other one from identifying them with the linear automorphisms over themselves.

**Theorem 2.10** *Let* $(X, \|\cdot\|)$ *be a normed algebra and* $1 \leq p, q \leq \infty$ *such that* $\frac{1}{p} + \frac{1}{q} = 1$. *For* $Y \in L^p(\Omega, \mathcal{F}, P, X, \|\cdot\|)$ *and* $Z \in L^q(\Omega, \mathcal{F}, P, X, \|\cdot\|)$ *the Hölder inequality holds:*

$$\|YZ\|_{L^1} = E(\|YZ\|) \leq E(\|Y\|\|Z\|) \leq \|Y\|_{L^p}\|Z\|_{L^q}.$$

**Proof:** The first inequality holds, since $X$ is a normed algebra, and the second one is the usual Hölder inequality for $\mathbb{R}$-valued random variables. $\square$

The random variables occurring in the following proposition take values in different normed vector spaces. All norms are denoted by $\|\cdot\|$.

**Theorem 2.11** *Let $X$ and $V$ be two normed vector spaces and $B(X,V)$ be the space of bounded linear operators from $X$ to $V$ (equipped with the induced operator norm). Moreover, let $1 \leq p,q \leq \infty$ be such that $\frac{1}{p}+\frac{1}{q}=1$. For $Y \in L^p(\Omega,\mathcal{F},P,X,\|\cdot\|)$ and $A \in L^q(\Omega,\mathcal{F},P,B(X,V),\|\cdot\|)$ the Hölder inequality holds:*

$$\|AY\|_{L^1} = E(\|AY\|) \leq E(\|A\|\|Y\|) \leq \|Y\|_{L^p}\|A\|_{L^q},$$

*i.e. $AY \in L^1(\Omega,\mathcal{F},P,V,\|\cdot\|)$*

**Proof:** The first inequality holds due to the definition of an operator norm and the second one is the usual Hölder inequality for $\mathbb{R}$-valued random variables. $\square$

The Riesz-Fischer theorem is also extendable to random variables assuming values in a Banach space.

**Theorem 2.12 (Riesz-Fischer)** *Let $(X,\|\cdot\|)$ be a Banach space. Then $(L^p(\Omega,\mathcal{F},P,X,\|\cdot\|),\|\cdot\|_{L^p})$ is a Banach space for $p \geq 1$ and $(L^p(\Omega,\mathcal{F},P,X,\|\cdot\|),d_{L^p}(\cdot,\cdot))$ is a complete metric space for $0<p<1$.*

For $1 \leq p < \infty$ this result can be found in Dunford and Schwartz (1958, III.6.5). The proof given there is a straightforward extension of the proof for $L^p(\Omega,\mathcal{F},P,\mathbb{R},|\cdot|)$. Below we give a proof of Theorem 2.12 that is an extension of the proof given in Loève (1977, p. 163) for real valued random variables and finite $p$. We use the following auxiliary result:

**Lemma 2.13** *Let $(X,\|\cdot\|)$ be a normed space. Let $Y_n$ be a sequence in $L^p(\Omega,\mathcal{F},P,X,\|\cdot\|)$ and $Y$ be an $X$-valued random variable. If $\|Y_n - Y\|_{L^p} \to 0$, resp. $d_{L^p}(Y_n,Y) \to 0$, for some $p \in (0,\infty]$, then $Y$ is in $L^p(\Omega,\mathcal{F},P,X,\|\cdot\|)$.*

**Proof:** Note that $Y_n$ is a Cauchy sequence in the respective norm/metric. For finite $p$ we have $E(\|Y_n - Y\|^p) \to 0$. Thus the triangle inequality implies $E(|\|Y_n\| - \|Y\||^p) \to 0$. From $\|Y_n\| \in L^p(\Omega,\mathcal{F},P,\mathbb{R},|\cdot|)$ and Loève (1977, 9.4 d.) we thus have $\|Y\| \in L^p(\Omega,\mathcal{F},P,\mathbb{R},|\cdot|)$ and an application of Lemma 2.4 concludes the proof.

For $p=\infty$ one similarly obtains $\||\|Y_n\| - \|Y\||\|_{L^\infty} \to 0$. From the completeness of $L^\infty(\Omega,\mathcal{F},P,\mathbb{R},|\cdot|)$ one thus has $\|Y\| \in L^\infty(\Omega,\mathcal{F},P,\mathbb{R},|\cdot|)$, so that Lemma 2.4 again shows the claim. $\square$

**Proof of Theorem 2.12:** Only completeness remains to be shown: Let $(Y_n)_{n\in\mathbb{N}}$ be a Cauchy sequence in $L^p(\Omega,\mathcal{F},P,X,\|\cdot\|)$.

For finite $p$ this implies $E(\|Y_m - Y_n\|^p) \to 0$ with $m,n \to \infty$. From the Markov inequality we thus obtain for every $\epsilon > 0$:

$$P(\|Y_m - Y_n\| \geq \epsilon) \leq \frac{1}{\epsilon^p}E(\|Y_m - Y_n\|^p) \to 0 \text{ with } m,n \to \infty$$

Hence, $\|Y_m - Y_n\|$ converges to 0 for $m, n \to \infty$ in probability and a.s. along a subsequence $(Y_{n_k})_{k\in\mathbb{N}}$. So, since $X$ is a Banach space, there is some $Y$ such that $Y_{n_k} \to Y$ a.s. for $k \to \infty$. $E(\|Y_m - Y_{n_k}\|^p) \to 0$ with $m, k \to \infty$ now implies using Fatou's Lemma:

$$E(\|Y_m - Y\|^p) = E(\liminf_{k\to\infty} \|Y_m - Y_{n_k}\|^p) \le \liminf_{k\to\infty} E(\|Y_m - Y_{n_k}\|^p) \to 0 \text{ as } m \to \infty.$$

This shows that $(Y_m)$ converges to $Y$ in $L^p$ and the completeness employing the last lemma.

For $p = \infty$ we have from $\|Y_m - Y_n\|_{L^\infty} \to 0$ that for any $\epsilon > 0$ there is an $N \in \mathbb{N}$ such that $\|Y_m - Y_n\| < \epsilon$ $(*)$ a.s. for all $m, n \ge N$. Thus $Y_m$ is a.s. a Cauchy sequence and so there is a random variable $Y$ against which $Y_m$ converges a.s. By $(*)$ this convergence is uniform and so taking the limit for $n \to \infty$ in $(*)$ and applying the last Lemma shows that $Y_m \to Y$ in $L^\infty$ and $Y \in L^\infty$. □

In order to answer the question, whether an $\mathbb{R}^d$-valued random variable is in $L^p$ (over $\mathbb{R}^d$) provided its components are in $L^p$ (over $\mathbb{R}$) and vice versa, we give some general result on the relation between the $L^p$ spaces over isomorphic normed spaces.

**Theorem 2.14** *Let $(X, \|\cdot\|)$ and $(Y, \|\cdot\|)$ be two isomorphic normed vector spaces. Then the spaces $L^p(\Omega, \mathcal{F}, P, X, \|\cdot\|)$ and $L^p(\Omega, \mathcal{F}, P, Y, \|\cdot\|)$ with the norm $\|\cdot\|_{L^p}$, resp. metric $d_{L^p}$, are isomorphic for all $p \in (0, \infty]$. If $X$ and $Y$ are moreover isometric, so are the respective $L^p$ spaces.*

The explicit construction of the isomorphism in the proof below shows that for two equivalent norms $\|\cdot\|_1$ and $\|\cdot\|_2$ over some vector space $X$ the sets $L^p(\Omega, \mathcal{F}, P, X, \|\cdot\|_1)$ and $L^p(\Omega, \mathcal{F}, P, X, \|\cdot\|_2)$ agree.

**Proof:** Let $T : X \to Y$ be an isomorphism between $X$ and $Y$ such that $m\|x\| \le \|Tx\| \le M\|x\|\ \forall x \in X$ with some constants $0 < m \le M$, then $T^{-1} : Y \to X$ is an isomorphism from $Y$ to $X$ and $M^{-1}\|y\| \le \|T^{-1}y\| \le m^{-1}\|y\|\ \forall y \in Y$. On $L^p(\Omega, \mathcal{F}, P, X, \|\cdot\|)$ now define $S : L^p(\Omega, \mathcal{F}, P, X, \|\cdot\|) \to L^p(\Omega, \mathcal{F}, P, Y, \|\cdot\|), Z \mapsto TZ$ and $R : L^p(\Omega, \mathcal{F}, P, Y, \|\cdot\|) \to L^p(\Omega, \mathcal{F}, P, X, \|\cdot\|), Z \mapsto T^{-1}Z$. The linear operators are well defined and bounded, since for $S$ we have $m^p E(\|Z\|^p) \le E(\|TZ\|^p) \le M^p E(\|Z\|^p)$ for $p \in (0, \infty)$, resp. $m\|Z\|_{L^\infty} \le \|TZ\|_{L^\infty} \le M\|Z\|_{L^\infty}$ for $p = \infty$, for all $Z \in L^p$ and likewise results hold for $R$. The bijectivity property follows from the obvious $RS = I$ and $SR = I$. Finally, the claimed isomorphicity, resp. isometricity, are an immediate consequence of the above inequalities as well, noting that $m = M = 1$ can be chosen, if $X$ and $Y$ are isometric. □

**Corollary 2.15** *Let $X = (X_1, X_2, \ldots, X_d)^\mathsf{T}$ be an $\mathbb{R}^d$-valued random variable, $\|\cdot\|$ any norm on $\mathbb{R}^d$ and $p \in (0, \infty]$. Then $X \in L^p(\Omega, \mathcal{F}, P, \mathbb{R}^d, \|\cdot\|)$, iff $X_i \in L^p(\Omega, \mathcal{F}, P, \mathbb{R}, |\cdot|)$ for $i = 1, \ldots, d$.*

Consequently the sets of $p$-integrable $\mathbb{R}^d$-valued random variables are independent of the particular norm on $\mathbb{R}^d$.

**Proof:** "$\Rightarrow$": Assume first $\|\cdot\| = \|\cdot\|_\infty$. Then $X \in L^p$ obviously implies $X_i \in L^p$ for all $i = 1, \ldots, d$. For general $\|\cdot\|$ the identity operator on $\mathbb{R}^d$ is an isomorphism between $(\mathbb{R}^d, \|\cdot\|)$ and $(\mathbb{R}^d, \|\cdot\|_\infty)$. Using the Theorem before (and the comment thereafter), one thus has, that $X \in L^p(\Omega, \mathcal{F}, P, \mathbb{R}^d, \|\cdot\|)$ implies $X_i \in L^p$.

"$\Leftarrow$": Assume now first that $\|\cdot\| = \|\cdot\|_1$. Then, since $X_i \in L^p$ implies $\sum_{i=1}^d |X_i| \in L^p$, $X \in L^p$ is an immediate consequence of $X_i \in L^p$ for all $i$. That for a general norm $\|\cdot\|$ on $\mathbb{R}^d$ $X \in L^p(\Omega, \mathcal{F}, P, \mathbb{R}^d, \|\cdot\|)$ is implied by $X_1, \ldots, X_d \in L^p$, is now shown by applying the last Theorem on the identity operator from $(\mathbb{R}^d, \|\cdot\|)$ to $(\mathbb{R}^d, \|\cdot\|_1)$. □

## 2.5 Ergodicity

In order to prove stationarity conditions for a stochastic difference equation and consequently Markov switching ARMA models, we shall employ the theory of ergodic random sequences, a subject to be found in most books on probability theory, e.g. Loève (1978, ch. X) or Shiryaev (1996, ch. V). A comprehensive monograph on ergodicity is Krengel (1985). In the following we provide an overview of the results needed later, following mainly Ash and Gardner (1975). A somewhat similar but shorter summary is Appendix 1.2 of Brandt, Franken and Lisek (1990). Yet, unlike them we will state the results more generally than only for random sequences.

The basis of ergodicity theory is formed by some special classes of sets and functions.

**Definition 2.16 (cf. Ash and Gardner (1975, pp. 113, 117, 119))** *Let $(\Omega, \mathcal{F}, P)$ be a probability space.*

*(i) An $\mathcal{F}$-measurable map $T : \Omega \to \Omega$ is called a* measure preserving *transformation ($P$-preserving or preserves $P$), if $P(T^{-1}(A)) = P(A)\ \forall\ A \in \mathcal{F}$.*

*(ii) A set $A \in \mathcal{F}$ is said to be* invariant *under a measure preserving transformation $T$, if $A = T^{-1}(A)$, and* almost invariant*, if $P(A \Delta T^{-1}(A)) = 0$.*

*(iii) An $\mathcal{F} - \mathcal{B}(\mathbb{R})$-measurable function $g : \Omega \to \mathbb{R}$ is called* invariant *under a measure preserving transformation $T$, if $g(T\omega) = g(\omega)\ \forall\ \omega \in \Omega$, and* almost invariant*, if $g(T\omega) = g(\omega)$ holds almost surely.*

*(iv) A measure preserving transformation $T$ is said to be* ergodic*, if either $P(A) = 0$ or $P(A^c) = 0$ for every invariant set $A$.*

*(v) A measure preserving transformation $T$ is said to be* mixing*, if for all $A, C \in \mathcal{F}$*

$$\lim_{n\to\infty} P(A \cap T^{-n}(C)) = P(A) \cdot P(C).$$

Sometimes the term *metrically transitive* is used instead of 'ergodic' (cf. Shiryaev (1996, p. 407). To define ergodicity and the mixing property for a random sequence we use the back-shift operator on the image space defined in Section 2.3.

**Definition 2.17** *A sequence $X = (X_i)_{i\in\mathbb{Z}}$ of $(E, \mathcal{E})$-valued random variables defined on a probability space $(\Omega, \mathcal{F}, P)$ is called* ergodic (mixing)*, if the back-shift operator $B : E^{\mathbb{Z}} \to E^{\mathbb{Z}}$ is an ergodic (mixing) transformation on $(E^{\mathbb{Z}}, \mathcal{E}^{\mathbb{Z}}, P_X = P \circ X^{-1})$.*

In Brandt, Franken and Lisek (1990) sequences are said to be ergodic, if the forward-shift operator is an ergodic transformation on $(E^{\mathbb{Z}}, \mathcal{E}^{\mathbb{Z}}, P_X = P \circ X^{-1})$. To see that both definitions are equivalent, we need the following lemma.

**Lemma 2.18** *Let $(\Omega, \mathcal{F}, P)$ be a probability space, $T : \Omega \to \Omega$ be measurable and bijective and $\mathscr{E} = T(\mathscr{E})$ for some subset $\mathscr{E} \subset \mathcal{F}$ with $\sigma(\mathscr{E}) = \mathcal{F}$. Then $T^{-1}$ is measurable. Moreover, if $T$ is measure preserving (ergodic), $T^{-1}$ is measure preserving (ergodic).*

Note that $\mathscr{E} = T(\mathscr{E})$ is equivalent to $\mathscr{E} = T^{-1}(\mathscr{E})$.
**Proof:** Let $A \in \mathscr{E}$, then $(T^{-1})^{-1}(A) = T(A) \in \mathscr{E}$ and this shows the measureability of $T^{-1}$ (cf. Bauer (1992, Th. 7.2)).

Let now $T$ be measure preserving and $A \in \mathcal{F}$, then

$$P\left((T^{-1})^{-1}(A)\right) = P\left(T^{-1}\left((T^{-1})^{-1}(A)\right)\right) = P(A),$$

so $T^{-1}$ is measure preserving. An under $T$ invariant $A \in \mathcal{F}$ is invariant under $T^{-1}$ as well and vice versa. Hence, $T^{-1}$ is ergodic, provided $T$ is ergodic. □

Since the back-shift operator $B$ and the forward-shift operator $B^{-1}$ are measurable and map the cylinder sets of $\mathcal{E}^{\mathbb{Z}}$ onto themselves, the above lemma implies the equivalence of the definitions.

It is also possible to characterize stationarity via a measure preserving property of the back-shift operator.

**Proposition 2.19 (cf. Ash and Gardner (1975, pp. 114f))** *A random sequence $X : (\Omega, \mathcal{F}) \to (E^{\mathbb{Z}}, \mathcal{E}^{\mathbb{Z}})$ is stationary, iff the back-shift operator $B : E^{\mathbb{Z}} \to E^{\mathbb{Z}}$ preserves $P_X$.*

**Proof (adapted from Ash and Gardner (1975, pp. 114f)):** Let $B$ preserve $P_X$ and be $A \in \mathscr{Z}_{\mathcal{E}^{\mathbb{Z}}}$. Then $P(X \in A) = P_X(A) = P_X(B^{-1}A) = \ldots = P_X(B^{-k}A) = P(B^k X \in A)$ for all natural $k$ and this implies stationarity of $X$ as noted in Section 2.3.

Conversely, the above equation shows that $P_X(A) = P_X(B^{-1}A)$ for all $A \in \mathscr{Z}_{\mathcal{E}^{\mathbb{Z}}}$, if $X$ is stationary. Since $\mathcal{G} := \left\{A \in \mathcal{E}^{\mathbb{Z}} : P_X(A) = P_X(B^{-1}A)\right\}$ is also a Dynkin system (note that $E^{\mathbb{Z}}$ is a cylinder, $B^{-1}(A^c) = (B^{-1}A)^c$ and let $A_i \in \mathcal{E}^{\mathbb{Z}}, i \in \mathbb{N}$, be disjoint, then we have that $B^{-1}A_i, i \in \mathbb{N}$ are disjoint), one obtains $\mathcal{G} = \mathcal{E}^{\mathbb{Z}}$ (cf. Bauer (1992, Th. 2.4)), because $\mathscr{Z}_{\mathcal{E}^{\mathbb{Z}}}$ is closed under intersections, and thus $B$ preserves $P_X$. □

The following lemmata and theorems summarize the properties of ergodic transformations and sequences. Their proofs can be found in Ash and Gardner (1975).

**Lemma 2.20 (Ash and Gardner (1975, Lemmata 3.2.2, 3.2.3, 3.2.4))**
*Let $(\Omega, \mathcal{F}, P)$ be a probability space.*

*(i) For every almost invariant set $A \in \mathcal{F}$ there exists an invariant $B \in \mathcal{F}$ with $P(A \Delta B) = 0$.*

*(ii) A measure preserving transformation $T$ is ergodic, iff $P(A) = 0$ or $P(A^c) = 0$ for every almost invariant set $A \in \mathcal{F}$.*

*(iii) Let $T$ be measure preserving. Then the following are equivalent:*

*(a) $T$ is ergodic.*

*(b) Every almost invariant function is a.s. constant.*

*(c) Every invariant function is a.s. constant.*

*(iv) Let $T$ be measure preserving and $A \in \mathcal{F}$. Then the following are equivalent:*

*(a) $A$ is almost invariant.*

*(b) $P(T^{-1}(A)\backslash A) = 0$.*

*(c) $P(A\backslash T^{-1}(A)) = 0$.*

Now we show that the mixing property implies ergodicity.

**Theorem 2.21 (Ash and Gardner (1975, Th. 3.2.6))** *Let $T$ be a mixing transformation on a probability space $(\Omega, \mathcal{F}, P)$. Then $T$ is ergodic.*

**Proof (Ash and Gardner (1975, p. 120)):**
Let $A \in \mathcal{F}$ be invariant. Since $T^{-n}(A) = A$, we obtain $P(A) = P(A \cap A) = P(A \cap T^{-n}(A))$. Letting $n \to \infty$ and employing the mixing property of $T$, one gets $P(A) = P(A)^2$. Thus $P(A) = 0$ or $P(A^c) = 0$ showing the ergodicity of $T$. □

To show that a measure preserving transformation is mixing, it is only necessary to verify the mixing condition on a generating algebra as the next theorem points out.

**Theorem 2.22 (Ash and Gardner (1975, Th. 3.2.7))** *Let $T$ be measure preserving on $(\Omega, \mathcal{F}, P)$ and $\mathscr{F} \subset \mathcal{F}$ an algebra such that $\sigma(\mathscr{F}) = \mathcal{F}$. If $\lim_{n\to\infty} P(A \cap T^{-n}C) = P(A)P(C)$ for all $A, C \in \mathscr{F}$, then $T$ is mixing.*

A simple example of mixing random sequences are i.i.d. sequences. The following is a generalization of the result given in Ash and Gardner (1975, p. 123).

**Proposition 2.23** *Let $X = (X_i)_{i\in\mathbb{Z}}$ be a sequence of i.i.d. random variables into a measurable space $(E, \mathcal{E})$ having common distribution $P$, then $X$ is mixing.*

**Proof:** Since $X$ is obviously stationary, the back-shift operator $B$ is a measure preserving transformation on $(E^{\mathbb{Z}}, \mathcal{E}^{\mathbb{Z}}, \bigotimes_{i\in\mathbb{Z}} P)$ (cf. Prop. 2.19). For some $A, C \in \mathscr{Z}_{\mathcal{E}^{\mathbb{Z}}}$, we have that $A = (\pi_{i_1}, \ldots, \pi_{i_m})^{-1}(A')$ for some $i_1, \ldots, i_m \in \mathbb{Z}$, $m \in \mathbb{N}$ and $A' \in \mathcal{F}^m$ and likewise $C = (\pi_{j_1}, \ldots, \pi_{j_{m'}})^{-1}(C')$ for some $j_1, \ldots, j_{m'} \in \mathbb{Z}$, $m' \in \mathbb{N}$ and $B' \in \mathcal{F}^{m'}$. This gives for all $n \in \mathbb{N}, n > \max\{j_1, \ldots, j_{m'}\} - \min\{i_1, \ldots, i_m\}$ that $P_X(A \cap B^{-n}C) = P(X \in A, X \in B^{-n}C) = P(X \in A, B^n X \in C) = P((X_{i_1}, \ldots, X_{i_m}) \in A', (X_{j_1-n}, \ldots, X_{j_{m'}-n}) \in C') = P((X_{i_1}, \ldots, X_{i_m}) \in A')P((X_{j_1-n}, \ldots, X_{j_{m'}-n}) \in C') = P(X \in A)P(B^n X \in C) = P_X(A)P_X(C)$. This shows that the mixing condition holds for all cylinder sets, which implies via Theorem 2.22 that $B$ is a mixing transformation on $(E^{\mathbb{Z}}, \mathcal{E}^{\mathbb{Z}}, \bigotimes_{i\in\mathbb{Z}} P)$, because the cylinder sets clearly form an algebra. □

Mixing sequences are especially important, since they can be combined with an ergodic sequence to build an ergodic sequence in the product space. A proof of the following result is to be found in Brown (1976).

**Theorem 2.24 (Brown (1976, Prop. 1.6), Brandt, Franken and Lisek (1990, Th. A 1.2.6))** *Let $Y_1 = (Y_{1,n})_{n\in\mathbb{Z}}$ and $Y_2 = (Y_{2,n})_{n\in\mathbb{Z}}$ be two independent stationary random sequences assuming values in measurable spaces $(E_1, \mathcal{E}_1)$, $(E_2, \mathcal{E}_2)$ respectively. If $Y_1$ is ergodic and $Y_2$ is mixing, then $Y = (Y_n)_{n\in\mathbb{Z}} = (Y_{1,n}, Y_{2,n})_{n\in\mathbb{Z}}$ is a stationary and ergodic sequence of random variables in $(E_1 \times E_2, \mathcal{E}_1 \times \mathcal{E}_2)$.*

Another way to obtain ergodic sequences from other ergodic sequences is given in the following lemma.

**Lemma 2.25 (cf. Brandt, Franken and Lisek (1990, Lemma A 1.2.7))** *Let* $X = (X_i)_{i\in\mathbb{Z}}$ *be a stationary and ergodic sequence of* $(E,\mathcal{E})$*-valued random variables and* $g_n : (E^{\mathbb{Z}},\mathcal{E}^{\mathbb{Z}}) \to (F,\mathcal{F})$, $n \in \mathbb{Z}$, *be a sequence of measurable functions such that* $g_{n-1} = g_n \circ B \,\forall\, n \in \mathbb{Z}$, *where* $B$ *is the back-shift operator on* $E^{\mathbb{Z}}$. *Then* $Y = (g_n(X))_{n\in\mathbb{Z}}$ *is also a stationary and ergodic sequence.*

**Proof:** (adapted from Brandt, Franken and Lisek (1990, p. 303))
$B$ will in the following denote the back-shift operator on both $E^{\mathbb{Z}}$ and $F^{\mathbb{Z}}$. Choose some $h \in \mathbb{N}$ and $A = (\pi_{i_1},\dots,\pi_{i_m})^{-1}(A')$ for some $i_1,\dots,i_m \in \mathbb{Z}$, $m \in \mathbb{N}$ and $A' \in \mathcal{F}^m$. Then $P(B^hY \in A) = P(B^h(g_n(X))_{n\in\mathbb{Z}} \in A) = P((g_{n-h}(X))_{n\in\mathbb{Z}} \in A) = P((g_n(B^hX))_{n\in\mathbb{Z}} \in A) = P((g_{i_1}(B^hX),\dots,g_{i_m}(B^hX)) \in A') = P(B^hX \in (g_{i_1},\dots,g_{i_m})^{-1}(A')) \overset{X\ stationary}{=} P(X \in (g_{i_1},\dots,g_{i_m})^{-1}(A')) = \dots = P(Y \in A)$ and so $Y$ is stationary.

Let now $A \in \mathcal{F}^{\mathbb{Z}}$ be invariant under $B$. Then we obtain $B^{-1}(\{x \in E : (g_n(x))_{n\in\mathbb{Z}} \in A\}) = \{x \in E : (g_n(Bx))_{n\in\mathbb{Z}} \in A\} = \{x \in E : (g_{n-1}(x))_{n\in\mathbb{Z}} \in A\} = \{x \in E : (g_n(x))_{n\in\mathbb{Z}} \in B^{-1}A\} \overset{A\ invariant}{=} \{x \in E : (g_n(x))_{n\in\mathbb{Z}} \in A\}$, hence the pre-image of $A$, i.e. $\{x \in E : (g_n(x))_{n\in\mathbb{Z}} \in A\} \in \mathcal{E}^{\mathbb{Z}}$, is invariant. Thus the ergodicity of $X$ implies via $P_Y(A) = P_X(\{x \in E : (g_n(x))_{n\in\mathbb{Z}} \in A\})$ the one of $Y$. □

The importance of measure preserving or ergodic transformations is due to the fact that the (conditional) expectation of a real valued functional $f$ on the probability space can be calculated by averaging over $(f \circ T^n(\omega))_{n\in\mathbb{N}}$. In terms of random sequences this means that ergodicity implies that a strong law of large numbers holds. These results and their implications are the subject of the following theorems.

**Theorem 2.26 (Birkhoff's ergodic theorem, see e.g. Ash and Gardner (1975, Theorems 3.3.6, 3.3.7))** *Let* $T$ *be a measure preserving transformation on a probability space* $(\Omega,\mathcal{F},P)$ *and* $f \in L^1(\Omega,\mathcal{F},P,\mathbb{R})$. *Then there exists an* $\hat{f} \in L^1$ *such that*

$$\frac{1}{n}\sum_{k=0}^{n-1} f(T^k\omega) \overset{n\to\infty}{\longrightarrow} \hat{f}(\omega)$$

*almost surely and in* $L^1$.

*If for some* $1 < p < \infty$ $f$ *is also in* $L^p$, *then* $\hat{f} \in L^p$ *and the above convergence holds also in* $L^p$.

It is immediate to see that the (almost) invariant sets form $\sigma$-algebras (cf. Ash and Gardner (1975, p. 117), Shiryaev (1996, p. 407)).

**Theorem 2.27 (Ash and Gardner (1975, Lemma 3.3.8, Th. 3.3.9,3.3.10))** *Let the map* $T$ *be a measure preserving transformation on a probability space* $(\Omega,\mathcal{F},P)$, $f \in L^1(\Omega,\mathcal{F},P,\mathbb{R})$ *and be* $\hat{f}$ *the function from Theorem 2.26. Then* $\hat{f}$ *is almost invariant and* $\mathcal{G}-\mathcal{B}(\mathbb{R})-$ *measurable, where* $\mathcal{G}$ *denotes the* $\sigma$*-algebra of all almost invariant elements in* $\mathcal{F}$. *Moreover, we have* $\hat{f} = E(f|\mathcal{G})$.

*If* $T$ *is even ergodic,* $\hat{f} = E(f)$ *holds a.s.*

**Proposition 2.28 (cf. Ash and Gardner (1975, p. 135))** *Let the transformation $T$ be measure preserving on a probability space $(\Omega, \mathcal{F}, P)$ and be $\mathscr{F} \subset \mathcal{F}$ an algebra such that $\sigma(\mathscr{F}) = \mathcal{F}$. Then the following are equivalent:*

*(i)* $T$ *is ergodic.*

*(ii)* $\frac{1}{n}\sum_{k=0}^{n-1} f(T^k\omega) \overset{\text{a.s.}}{\to} E(f)$ *for all* $f \in L^1(\Omega, \mathcal{F}, P, \mathbb{R})$

*(iii)* $\frac{1}{n}\sum_{k=0}^{n-1} I_A(T^k\omega) \overset{\text{a.s.}}{\to} P(A)$ *for all* $A \in \mathcal{F}$

*(iv)* $\frac{1}{n}\sum_{k=0}^{n-1} I_A(T^k\omega) \overset{\text{a.s.}}{\to} P(A)$ *for all* $A \in \mathscr{F}$

*(v)* $\frac{1}{n}\sum_{k=0}^{n-1} P(A \cap T^{-k}C) \to P(A)P(C)$ *for all* $A, C \in \mathscr{F}$

*(vi)* $\frac{1}{n}\sum_{k=0}^{n-1} P(A \cap T^{-k}C) \to P(A)P(C)$ *for all* $A, C \in \mathcal{F}$

For later reference we repeat the above results for the special case of random sequences. Note that stationarity of a sequence implies that $B$ is measure preserving on the image space as shown before.

**Theorem 2.29** *Let $X$ be a stationary $(E, \mathcal{E})$-valued random sequence and $f \in L^1(E^{\mathbb{Z}}, \mathcal{E}^{\mathbb{Z}}, P_X, \mathbb{R})$. Then there exists an $\hat{f} \in L^1$ such that*

$$\frac{1}{n}\sum_{k=0}^{n-1} f(B^k X) \overset{n\to\infty}{\longrightarrow} \hat{f}(X)$$

*almost surely and in $L^1$.*

*If for some $1 < p < \infty$ $f$ is also in $L^p$, then $\hat{f} \in L^p$ and the above convergence holds also in $L^p$.*

**Theorem 2.30** *Let $X$ be a stationary $(E, \mathcal{E})$-valued random sequence and $f \in L^1(E^{\mathbb{Z}}, \mathcal{E}^{\mathbb{Z}}, P_X, \mathbb{R})$ and be $\hat{f}$ the function from Theorem 2.29. Then $\hat{f}$ is almost invariant and $\mathcal{G} - \mathcal{B}(\mathbb{R})-$ measurable, where $\mathcal{G}$ denotes the $\sigma$-algebra of all almost invariant elements in $\mathcal{E}^{\mathbb{Z}}$. Moreover, we have $\hat{f}(X) = E(f(X)|\mathcal{G})$.*

*If $T$ is even ergodic, $\hat{f}(X) = E(f(X))$ holds a.s.*

**Proposition 2.31 (cf. Brandt, Franken and Lisek (1990, Th. A 1.2.2))** *Let $X$ be a stationary random sequence assuming values in $(E, \mathcal{E})$ and be $\mathscr{E} \subset \mathcal{E}^{\mathbb{Z}}$ an algebra such that $\sigma(\mathscr{E}) = \mathcal{E}^{\mathbb{Z}}$. Then the following are equivalent:*

*(i)* $X$ *is ergodic.*

*(ii)* *The back-shift operator* $B$ *is ergodic on* $(E^{\mathbb{Z}}, \mathcal{E}^{\mathbb{Z}}, P_X)$.

*(iii)* $\frac{1}{n}\sum_{k=0}^{n-1} f(B^k X) \stackrel{\text{a.s.}}{\to} E(f(X))$ *for all* $f \in L^1(E^{\mathbb{Z}}, \mathcal{E}^{\mathbb{Z}}, P_X, \mathbb{R})$

*(iv)* $\frac{1}{n}\sum_{k=0}^{n-1} I_A(B^k X) \stackrel{\text{a.s.}}{\to} P_X(A)$ *for all* $A \in \mathcal{E}^{\mathbb{Z}}$

*(v)* $\frac{1}{n}\sum_{k=0}^{n-1} I_A(B^k X) \stackrel{\text{a.s.}}{\to} P_X(A)$ *for all* $A \in \mathscr{E}$

*(vi)* $\frac{1}{n}\sum_{k=0}^{n-1} P_X(A \cap B^{-k}C) \to P_X(A)P_X(C)$ *for all* $A, C \in \mathscr{E}$

*(vii)* $\frac{1}{n}\sum_{k=0}^{n-1} P_X(A \cap B^{-k}C) \to P_X(A)P_X(C)$ *for all* $A, C \in \mathcal{E}^{\mathbb{Z}}$

Note that the cylinder sets of $\mathcal{E}^{\mathbb{Z}}$ can be taken as $\mathscr{E}$ above and that the back-shift operator $B$ can be replaced by the forward-shift operator $B^{-1}$. Furthermore, the ergodicity of the forward-shift operator $B^{-1}$ can be added to the equivalences.

To conclude this introduction to ergodicity, we state briefly the ergodicity criteria for Markov chains with at most countable state space. Note that initial distributions are not an issue in our set-up, since we deal with doubly-infinite sequences and thus any of the Markov chains considered are necessarily assumed to be stationary. Usually, positive recurrent, irreducible and aperiodic Markov Chains are called “ergodic” in the literature on countable state space Markov Chains (see e.g. Resnick (1992), Asmussen (2003)) irrespective of, whether the chain is stationary or not, since these texts deal with chains starting at time zero with some initial distribution. For a countable state space Markov chain to be ergodic in the sense defined by us, it needs to be stationary, irreducible and positive recurrent (cf. Asmussen (2003, p. 19), Ash and Gardner (1975, section 3.5)), as shall briefly be shown below. So, aperiodicity is not required of a Markov chain to be ergodic in our set-up. Furthermore, on a finite state space any recurrent chain is automatically positive recurrent (cf. Resnick (1992), Asmussen (2003)).

In Ash and Gardner (1975, section 3.5) ergodicity criteria are unfortunately only proved for Markov chains on a countable state space starting at time zero using the forward shift operator $B^{-1}$. We will now show that a doubly-infinite stationary, irreducible and positive recurrent Markov chain is ergodic using the following result of Brémaud (1999), which is an immediate generalization of a result to be found in all standard texts (e.g. Resnick (1992, Proposition 2.12.4)). Brémaud (1999) also works solely with Markov chains starting at some time “zero”, but it is obvious that this does not affect the validity of the result.

**Lemma 2.32 (cf. Brémaud (1999, Corollary 4.1))** *Let $\Delta = (\Delta_t)_{t\in\mathbb{Z}}$ be a stationary, irreducible and positive recurrent Markov chain with countable state space $E$ (equipped with $\sigma$-algebra $\mathcal{E}$). Let $L \in \mathbb{N}_0$ and be $g : E^{L+1} \to \mathbb{R}$ a function in $L^1(E^{L+1}, \mathcal{E}^{L+1}, P_{(\Delta_0,\Delta_1,\dots,\Delta_L)}, \mathbb{R}, |\cdot|)$. Then for any fixed $t_0 \in \mathbb{Z}$*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} g(X_{k+t_0}, X_{k+1+t_0}, \dots, X_{k+L+t_0}) = E(g) \ \textit{a.s.} \tag{2.7}$$

A proof is to be found in Brémaud (1999) and note that we can view $(\Delta_t)_{t\geq t_0}$ as a Markov chain starting with its stationary distribution at time “zero” ($t_0$).

**Theorem 2.33** *Let $\Delta = (\Delta_t)_{t\in\mathbb{Z}}$ be a stationary, irreducible and positive recurrent Markov chain with countable state space $E$ (equipped with $\sigma$-algebra $\mathcal{E}$). Then $\Delta$ is ergodic.*

**Proof:** According to Proposition 2.31 and the remarks thereafter it suffices to show that for any cylinder set $A$ in $\mathcal{E}^{\mathbb{Z}}$

$$\frac{1}{n}\sum_{k=0}^{n-1} I_A(B^{-k}\Delta) \overset{\text{a.s.}}{\to} P_\Delta(A)$$

holds.

Let $A$ be in $\mathscr{Z}_{\mathcal{E}^{\mathbb{Z}}}$ then there are $t_0, m \in \mathbb{N}_0$ and $A' \in \mathcal{E}^{m+1}$ such that

$$A = (\pi_{t_0+1}, \dots, \pi_{t_0+m+1})^{-1}(A').$$

Hence, $I_A(\cdot) = I_{A'}((\pi_{t_0+1}, \dots, \pi_{t_0+m+1})(\cdot))$ and so using the last lemma

$$\frac{1}{n}\sum_{k=0}^{n-1} I_A(B^{-k}\Delta) = \frac{1}{n}\sum_{k=1}^{n} I_{A'}(\Delta_{t_0+k}, \dots, \Delta_{k+m+t_0}) \overset{n\to\infty}{\longrightarrow} P_{(\Delta_0,\dots,\Delta_m)}(A') \text{ a.s.}$$

But the stationarity of $\Delta$ implies $P_{(\Delta_0,\dots,\Delta_m)}(A') = P_{(\Delta_{t_0+1},\dots,\Delta_{t_0+m+1})}(A') = P_\Delta(A)$ and thus $\Delta$ is ergodic. $\square$

## 2.6 Geometric Ergodicity, Strong Mixing and Extreme Values

The aim of this section is to give a brief outline of the concepts of geometric ergodicity of Markov chains and of strong mixing of random sequences. Furthermore, we review briefly the basic results from extreme value theory for stationary random sequences presuming knowledge of the basics of extreme value theory for i.i.d. sequences. In the following the necessary terminology from general Markov chain theory is introduced, but without any detailed discussions. For more details the monograph Meyn and Tweedie (1993), for instance, could be consulted. Throughout we presume the considered Markov chains to be homogeneous and to have a subset of $\mathbb{R}^d$ with appropriate $d$ or of a topologically isomorphic vector space as state space. Moreover, the $\sigma$-algebra over the state space is assumed to be the restriction of the natural Borel $\sigma$-algebra to the state space. As usual, $P^n(\cdot,\cdot) : E\times\mathcal{E} \to [0,1]$ denotes the $n$-step transition kernel. Furthermore, we shall employ the *total variation norm*, $\|\cdot\|_{TV}$. Recall that for two measures $\mu_1$ and $\mu_2$ on a measurable space $(E,\mathcal{E})$ this is defined as

$$\|\mu_1 - \mu_2\|_{TV} := \sup_{P\in\mathcal{P}} \sum_{j=1}^{\mathrm{card}(P)} |\mu_1(A_{j,P}) - \mu_2(A_{j,P})| \, ,$$

where $\mathcal{P}$ is the set of all countable measurable partitions $P = \{A_{1,P}, A_{2,P}, \ldots, A_{\mathrm{card}(P),P}\}$ of $E$.

The properties for Markov chains put forward in the following definition are mainly of interest when dealing with Markov chains that have not begun in the infinite past or started with their stationary distribution, but with some arbitrary initial distribution at time zero. Yet, it will turn out that these criteria are very helpful in order to study the concept of strong mixing for Markov-switching processes.

**Definition 2.34 (cf. Feigin and Tweedie (1985) or Basrak (2000))** *Let $X = (X_t)$ be a Markov chain with state space $E$ that is equipped with $\sigma$-algebra $\mathcal{E}$.*

*(i) $X$ is said to be a* weak Feller chain*, if $E\left(g(X_1)|X_0 = y\right)$ is a continuous function in $y \in E$ for all bounded and continuous functions $g : E \to \mathbb{R}$.*

*(ii) If $\mu$ is some nondegenerate measure on $(E,\mathcal{E})$ and for all $x \in E$ the following implication holds for every $A \in \mathcal{E}$*

$$\mu(A) > 0 \;\Rightarrow\; \sum_{n=1}^{\infty} P^n(x, A) > 0,$$

*then $X$ is called $\mu$-irreducible.*

*(iii) Provided there is a probability measure $\pi$ on $(E,\mathcal{E})$ such that for all $x \in E$*

$$\|P^n(x,\cdot) - \pi(\cdot)\|_{TV} \to 0$$

*as $n \to \infty$, $X$ is said to be* Harris ergodic*. If there is even a $\rho \in (0,1)$ such that*

$$\rho^{-n}\|P^n(x,\cdot) - \pi(\cdot)\|_{TV} \to 0$$

*as $n \to \infty$, $X$ is referred to as being* geometrically ergodic*.*

Note that Harris ergodicity in particular implies ergodicity in the sense of the previous section when dealing with stationary versions of a Markov chain. The following theorem is often used to show geometric ergodicity of Markov chains.

**Theorem 2.35 (Feigin and Tweedie (1985, Th. 1))** *Let $X = (X_t)$ be a weak Feller chain on $(E,\mathcal{E})$ and assume the existence of a measure $\mu$ and a compact set $K \in \mathcal{E}$ with $\mu(K) > 0$ such that* *(i)* *$X$ is $\mu$-irreducible and*

*(ii)* *there exists a non-negative continuous $g : E \to \mathbb{R}$ satisfying*

$$g(x) \geq 1 \ \forall\, x \in K$$

*and*

$$E\left(g(X_1)|X_0 = y\right) \leq (1-\delta)g(y) \ \ \forall y \in E \backslash K$$

*with some $\delta > 0$.*

*Then $X$ is geometrically ergodic.*

A particular concept formalizing the idea of weak dependence for a stochastic process is strong mixing. For a detailed discussion of this concept and related ones confer, for instance, Doukhan (1994).

**Definition 2.36 (cf. Leadbetter, Lindgren and Rootzén (1983, p. 52), Basrak (2000, Def. 2.2.1) and Basrak, Davis and Mikosch (2002b))** *A discrete time stationary stochastic process $X = (X_n)_{n\in\mathbb{Z}}$ is called* strongly mixing*, if*

$$\alpha_l := \sup\left\{|P(A\cap B) - P(A)P(B)| : A \in \mathcal{F}_{-\infty}^0, B \in \mathcal{F}_l^\infty\right\} \to 0$$

*as $l \to \infty$, where $\mathcal{F}_{-\infty}^0 := \sigma\left(\ldots, X_{-2}, X_{-1}, X_0\right)$ and $\mathcal{F}_l^\infty = \sigma\left(X_l, X_{l+1}, X_{l+2}, \ldots\right)$. The values $\alpha_l$ are called mixing coefficients.*

*If there are constants $C \in \mathbb{R}^+$ and $a \in (0,1)$ such that $\alpha_l \leq Ca^l$, $X$ is said to be* strongly mixing with geometric rate*.*

Any strongly mixing process is in particular mixing, as an application of Theorem 2.22 to the cylinder sets shows. The following observation is immediate from the definition, but later turns out to be essential.

**Proposition 2.37** *Let $(X_n, Y_n)_{n\in\mathbb{Z}}$ be a bivariate stochastic process that is strongly mixing (with geometric rate). Then both the univariate processes $(X_n)_{n\in\mathbb{Z}}$ and $(Y_n)_{n\in\mathbb{Z}}$ are strongly mixing (with geometric rate).*

The following relation between geometric ergodicity and strong mixing for Markov chains is very well-known. Details on the necessary arguments are to be found in Basrak, Davis and Mikosch (2002b, Section 2), for instance.

**Proposition 2.38** *Let $X = (X_n)_{n\in\mathbb{Z}}$ be a geometrically ergodic and stationary Markov chain, then $X$ is in particular strongly mixing with geometric rate.*

The strong mixing property of a random process has important implications when studying its extremal behaviour, since it implies the so called condition $D$ and thus $D(u_n)$ for any sequence $u_n$ (cf. Leadbetter (1974), Leadbetter, Lindgren and Rootzén (1983)). Under these conditions the Fisher-Tippett theorem, viz. that any non-degenerate limiting distribution of the linearly transformed maxima can only be of Gumbel, Fréchet or Weibull type, is valid for stationary random sequences, see Leadbetter, Lindgren and Rootzén (1983, Th. 3.3.3) or Embrechts, Klüppelberg and Mikosch (1997, Th. 4.4.1).

Moreover, these conditions are essential when employing the concept of extremal indices (cf. in particular Leadbetter, Lindgren and Rootzén (1983, Section 3.7)). Actually, Loynes (1965) showed the validity of the Fisher-Tippett theorem for stationary random sequences under strong mixing originally. For an up-to-date overview of extreme value theory for stationary random sequences see Embrechts, Klüppelberg and Mikosch (1997).

# Chapter 3

# Vague Convergence and Regular Variation

In this section we first give a brief summary on vague convergence of measures and then apply this concept to define and study regular variation of random variables in $\mathbb{R}^d$ for arbitrary $d \in \mathbb{N}$. The later is often also referred to as *multivariate regular variation*, as opposed to *univariate* regular variation, i.e. regular variation on $\mathbb{R}$.

## 3.1 Vague Convergence of Measures

Vague convergence is a mode of convergence for measures intrinsically linked to, but generally weaker than the more familiar weak convergence (convergence in distribution). Apart from the book by Resnick (1987) a highly readable account on this concept can be found in Bauer (1992). The following brief introduction is based on these two references. For an alternative introduction we refer to Lindskog (2004).

$E$ shall in the following denote any space that is endowed with a topology $\mathfrak{E}$ such that $(E, \mathfrak{E})$ is a locally compact polish (i.e. complete, separable and metrizable) space. By $\rho(\cdot, \cdot)$ we denote a metric inducing $\mathfrak{E}$ and $\mathcal{E}$ is the Borel $\sigma$-algebra (w.r.t. $\mathfrak{E}$). Furthermore, $M_+(E)$ denotes the set of all non-negative Radon measures (i.e. the locally finite measures that are regular from within and defined on the Borel $\sigma$-algebra) on $(E, \mathcal{E})$. We define $C_c(E)$ to be the set of all real valued continuous functions with compact support, i.e. $C_c(E) = \{f \in C(E) : supp(f) = \overline{\{x \in E : f(x) \neq 0\}} \text{ compact}\}$. Now we topologise $M_+(E)$ by defining $\mathfrak{M}_+(E)$ to be the weakest topology that makes the maps $\mu \mapsto \mu(f) = \int_E f d\mu$ from $M_+(E)$ to $\mathbb{R}$ continuous for all $f \in C_c(E)$. Thus a fundamental system of neighbourhoods in $\mathfrak{M}_+(E)$ is given by the system of sets

$$U_{f_1,\cdots,f_n;\epsilon}(\mu_0) := \{\mu \in M_+(E) : |\mu(f_i) - \mu_0(f_i)| < \epsilon, i = 1, \ldots, n\}$$

for $n \in \mathbb{N}$, $\epsilon > 0$ and $f_1, \ldots, f_n \in C_c(E)$. We call $\mathfrak{M}_+(E)$ the *vague topology* on $M_+(E)$. Note that, since for fixed $f \in C_c(E)$ the mapping $\mu \mapsto \mu(f)$ behaves very similar to a linear functional on $M_+(E)$, this concept is strongly related to a general weak topology as discussed e.g. in Werner (2002, Section VII.3). The main difference is that $M_+(E)$ is not a vector space, but only a convex cone. A simple consequence of Riesz' representation Theorem (cf. Bauer (1992, § 29)) is that $(M_+(E), \mathfrak{M}_+(E))$ is a Hausdorff space.

This is also part of the following much more powerful result, which also shows that in $(M_+(E), \mathfrak{M}_+(E))$ only sequences rather than nets need to be studied.

**Proposition 3.1 (Resnick (1987, Proposition 3.17))** *$(M_+(E), \mathfrak{M}_+(E))$ is a polish space.*

The above result is also part of Bauer (1992, § 31.5). At this stage it is necessary to state that the definitions of the vague topology we use (based on Bauer (1992)) and the one in Resnick (1987), where $C_c^+(E)$ (i.e. the non-negative functions in $C_c(E)$) is used instead of $C_c(E)$, are equivalent. To see this one just has to note that obviously a function $f : E \to \mathbb{R}$ is in $C_c(E)$, iff both $f_+$ and $f_-$ are in $C_c^+(E)$, where $f_+$ is given by $f_+(x) = \max\{f(x), 0\}$ and $f_- = (-f)_+$.

We are now in a position to define the needed concept of convergence of measures:

**Definition 3.2** *Let $(E, \mathfrak{E})$ be a locally compact polish space. A sequence $(\mu_n)_{n\in\mathbb{N}}$ of non-negative Radon measures on $E$ is said to be* vaguely convergent *to a Radon measure $\mu$, denoted by $\mu_n \xrightarrow{v} \mu$, if $\mu_n \xrightarrow{n\to\infty} \mu$ in the vague topology.*

From the above discussion we have:

**Lemma 3.3 (cf. Resnick (1987, p. 140), Bauer (1992, Def. 30.1))** *Vague limits are unique and $\mu_n \xrightarrow{v} \mu$, iff*

$$\lim_{n\to\infty} \int_E f d\mu_n = \int_E f d\mu \,\forall\, f \in C_c(E).$$

Two important characterizations of vague convergence, which we shall employ later, are given in the following proposition.

**Proposition 3.4 (Resnick (1987, Prop. 3.12), Bauer (1992, Satz 30.2))** *Let $\mu \in M_+(E)$ and $(\mu_n)_{n\in\mathbb{N}}$ be a sequence in $M_+(E)$. Then the following are equivalent:*

*(i)* $\mu_n \xrightarrow{v} \mu$ *for* $n \to \infty$.

*(ii)* $\mu_n(B) \to \mu(B)$ *for* $n \to \infty$ *for all relatively compact $B$ in the Borel-$\sigma$-algebra $\mathcal{E}$ that are $\mu$-boundaryless, i.e. $\mu(\partial B) = 0$.*

*(iii)* $\limsup_{n\to\infty} \mu_n(K) \le \mu(K)$ *and* $\liminf_{n\to\infty} \mu_n(G) \ge \mu(G)$ *for all compact $K \in \mathcal{E}$ and all open relatively compact $G$.*

The $\mu$-boundaryless sets are henceforth denoted by $\mathscr{B}_\mu$, i.e. $\mathscr{B}_\mu := \{B \in \mathcal{E} : \mu(\partial B) = 0\}$.

To end this brief introduction into vague convergence we show that a continuous map from one locally compact polish space under some additional assumptions induces a map between the Radon measures on these spaces that is continuous w.r.t. the vague topologies.

**Theorem 3.5 (Resnick (1987, Proposition 3.18))** *Let $(E_1, \mathfrak{E}_1)$ and $(E_2, \mathfrak{E}_2)$ be locally compact polish spaces and $T : E_1 \to E_2$ be a continuous mapping such that $T^{-1}(K)$ is compact for every compact $K \subseteq E_2$. Define $\hat{T} : M_+(E_1) \to M_+(E_2)$ by $\hat{T}(\mu) = \mu \circ T^{-1}$. Then $\hat{T}$ is continuous with respect to the vague topologies. In particular,*

$$\mu_n \xrightarrow{v} \mu \;\Rightarrow\; \hat{T}(\mu_n) \xrightarrow{v} \hat{T}(\mu). \tag{3.1}$$

We include a proof of this result, since, as we shall show by a counterexample below, one topological assertion used in the proof given in Resnick (1987) does not hold in general. However, apart from correcting for this we follow the lines of proof given there.
**Proof:** Due to Proposition 3.1 it suffices to establish (3.1). Let thus $(\mu_n)_{n\in\mathbb{N}}$ be a sequence of Radon measures on $E_1$ converging vaguely to a Radon measure $\mu$. For any $f \in C_c(E_2)$ and $\nu \in M_+(E_1)$ we have by the transformation theorem of elementary integration theory that

$$\int_{E_2} f d\hat{T}(\nu) = \int_{E_2} f d(\nu \circ T^{-1}) = \int_{E_1} f \circ T d\nu. \tag{3.2}$$

It is immediate that $\{x \in E_1 : f \circ T(x) \neq 0\} = T^{-1}\left(\{y \in E_2 : f(y) \neq 0\}\right)$ and thus

$$\begin{aligned} supp(f \circ T) &= \overline{\{x \in E_1 : f \circ T(x) \neq 0\}} = \overline{T^{-1}\left(\{y \in E_2 : f(y) \neq 0\}\right)} \\ &\subseteq T^{-1}\left(\overline{\{y \in E_2 : f(y) \neq 0\}}\right) = T^{-1}\left(supp(f)\right). \end{aligned}$$

The inclusion relation is shown as follows: For each $x \in \overline{T^{-1}\left(\{y \in E_2 : f(y) \neq 0\}\right)}$ there is a sequence $x_n \in T^{-1}\left(\{y \in E_2 : f(y) \neq 0\}\right)$ such that $x_n \to x$. From the continuity of $T$ one obtains that the sequence $T(x_n) \in \{y \in E_2 : f(y) \neq 0\}$ converges to $T(x) \in \overline{\{y \in E_2 : f(y) \neq 0\}}$ and, hence, $x \in T^{-1}(T(x)) \subseteq T^{-1}\left(\overline{\{y \in E_2 : f(y) \neq 0\}}\right)$.

From the above inclusion, the property that pre-images of compact sets under $T$ are compact and the fact that $supp(f \circ T)$ is closed by definition, we obtain that $supp(f \circ T)$ is compact using the well known fact that closed subsets of compacts are compact (see e.g. Rudin (1976, Th. 2.34)). Thus $f \circ T \in C_c(E_1)$. Using Lemma 3.3 and (3.2) this shows

$$\lim_{n\to\infty} \int_{E_2} f d\hat{T}(\mu_n) = \lim_{n\to\infty} \int_{E_1} f \circ T d\mu_n \to \int_{E_1} f \circ T d\mu = \int_{E_2} f d\hat{T}(\mu)$$

and again from Lemma 3.3 it follows that $\hat{T}(\mu_n) \xrightarrow{v} \hat{T}(\mu)$. □
In the proof to be found in Resnick (1987) the identity $supp(f \circ T) = T^{-1}(supp(f))$ is used. This equality does, however, not hold in general as the following counterexample shows. $\mathbb{R}$ with the usual topology is a locally compact polish space. Define $f : \mathbb{R} \to \mathbb{R}$ by

$$f(x) = \begin{cases} \sin(x) & \text{for } x \in (0, \pi) \\ 0 & \text{otherwise} \end{cases}.$$

Obviously $f \in C_c(\mathbb{R})$ and even $f \in C_c^+(\mathbb{R})$. Moreover, consider the map $T : \mathbb{R} \to \mathbb{R}$ given by

$$T(x) = \begin{cases} x & \text{for } x \in (0, \infty) \\ 0 & \text{for } x \in [-1, 0] \\ x+1 & \text{for } x \in (-\infty, -1) \end{cases}.$$

Then $T$ is continuous and thus $T^{-1}(A)$ is closed for all closed sets $A$. One sees immediately that moreover $T^{-1}(B)$ is bounded for all bounded sets $B$. Hence, $T^{-1}(K)$ is compact for all compact sets $K$. But we have:

$$\begin{aligned} \{y \in \mathbb{R} : f(y) \neq 0\} &= (0, \pi), \\ supp(f) &= [0, \pi], \end{aligned}$$

$$\begin{aligned}
T^{-1}\left((0,\pi)\right) &= (0,\pi), \\
T^{-1}\left([0,\pi]\right) &= [-1,\pi], \\
supp(f\circ T) &= \overline{T^{-1}\left(\{y\in\mathbb{R}: f(y)\neq 0\}\right)} = [0,\pi], \\
T^{-1}(supp(f)) &= [-1,\pi].
\end{aligned}$$

Thus $supp(f\circ T)\neq T^{-1}(supp(f))$, so the inclusion $supp(f\circ T)\subseteq T^{-1}(supp(f))$ employed in the above proof is strict for this choice of $E_1, E_2, f$ and $T$.

## 3.2 Multivariate Regular Variation

To extend the well-established concept of regular variation from the classical univariate framework to a multivariate one, vague convergence is employed. In the following we briefly review the concept of multivariate regular variation of random vectors, mainly based on Basrak (2000) and Mikosch (2003) (see Resnick (1987, Section 5.4.2) for an earlier account and Lindskog (2004) for a very thorough introduction emphasizing especially the geometry of the employed spaces). An extension of the concept to stochastic processes is Hult and Lindskog (2004). We presume familiarity with the basics of univariate regular variation (see Bingham, Goldie and Teugels (1989), Resnick (1987) or Embrechts, Klüppelberg and Mikosch (1997)). Recall in particular that for a function $f:\mathbb{R}^+\to\mathbb{R}$ univariate regular variation at infinity means that for large $x\in\mathbb{R}$ the function behaves roughly like $x^\alpha$ for some $\alpha\in\mathbb{R}$, which is called the index of regular variation. Formally we have the following definition:

**Definition 3.6 (Regular variation on $\mathbb{R}^+$)** *A Lebesgue measurable function $f:\mathbb{R}^+\to\mathbb{R}^+$ is said to be* regularly varying *at infinity with index $\alpha$, if for all $x>0$*

$$\lim_{t\to\infty}\frac{f(tx)}{f(t)} = x^\alpha. \tag{3.3}$$

*In the case $\alpha=0$, we speak of* slow variation.

*An $\mathbb{R}$-valued random variable $X$ with distribution function $F$ is said to be regularly varying with index $\alpha>0$, if $\bar{F}(x):=1-F(x)$ is regularly varying with index $-\alpha$ at infinity.*

In the following we will employ the space $\overline{\mathbb{R}^d}\backslash\{0\}$. For $d=1$ this is obtained as follows: Take the space $\mathbb{R}$ with the usual topology and form the usual two point compactification by setting $\overline{\mathbb{R}}=\mathbb{R}\cup\{\infty,-\infty\}$ and adding the neighbourhoods of $\pm\infty$, i.e. the sets $[-\infty,a)$ and $(a,\infty]$ with $a\in\mathbb{R}$, to the basic open sets. Then take $\overline{\mathbb{R}}\backslash\{0\}$ and remove the open neighbourhoods of 0 from the topology. This is also referred to as one point uncompactification. For the $d$-dimensional case one takes the two point compactification $\overline{\mathbb{R}^d}$, which is simply the $d$-fold product of $\overline{\mathbb{R}}$, and the product topology. Then one removes the point 0 from $\overline{\mathbb{R}^d}$ and the open neighbourhoods of 0 from the topology. One can interpret this procedure as interchanging the roles of zero and infinity. In $\overline{\mathbb{R}^d}\backslash\{0\}$ compact sets can by characterized by being closed (in the usual sense) and bounded away from zero. By this procedure we obtain a locally compact polish space, actually a possible metric on $\overline{\mathbb{R}}\backslash\{0\}$

is given by $d(x, y) := |x^{-1} - y^{-1}|$ (cf. Resnick (1987, p. 225f)). For the construction of a possible metric on $\overline{\mathbb{R}^d}\backslash\{0\}$ see Lindskog (2004, Th. 1.5), for instance.

Now we can define multivariate regular variation following Basrak (2000, p. 27ff) with a slight, but necessary, modification due to Lindskog (2004, Th. 1.21), viz. to demand nondegeneracy of the limiting measure. The reason, why one needs to exclude $\mu_X = 0$ is that otherwise Proposition 3.9 (ii) is violated for the zero measure, as an inspection of the arguments given in Basrak (2000) shows.

**Definition 3.7 (Regular variation on $\mathbb{R}^d$)**

*(i) Let $X$ be an $\mathbb{R}^d$-valued random variable. If there exists a non-zero $\mu_X$ in $M_+(\overline{\mathbb{R}^d}\backslash\{0\})$ with $\mu_X(\overline{\mathbb{R}^d}\backslash\mathbb{R}^d) = 0$, a relatively compact set $E$ in $\mathcal{B}(\overline{\mathbb{R}^d})$, i.e. a Borel set, and a dense subset $T \subset (0, \infty)$, such that $tE \in \mathscr{B}_{\mu_X}\ \forall\, t \in T$ and*

$$\mu_{X,t}(\cdot) := \frac{P(X \in t\cdot)}{P(X \in tE)} \xrightarrow{v} \mu_X(\cdot)$$

*in $M_+(\overline{\mathbb{R}^d}\backslash\{0\})$ for $t \to \infty$, then $X$ is said to be* (multivariate) regularly varying.

*(ii) A random sequence $(X_n)_{n\in\mathbb{Z}}$ is said to be* (multivariate) regularly varying*, if all its finite dimensional distributions are regularly varying.*

It is important to note that not any Radon measure $\mu_X$ can appear, when $X$ is regularly varying:

**Proposition 3.8 (cf.Basrak (2000, Th. 2.1.4))** *If $X$ is a regularly varying random variable, then there exists an $\alpha > 0$ such that $\mu_X(uS) = u^{-\alpha}\mu_X(S)$ for all $S \in \mathscr{B}_{\mu_X}$ and $u > 0$. Moreover, $\partial B_\delta(0) \in \mathscr{B}_{\mu_X}$ for all $\delta > 0$. In particular, $\mu_X$ has no atoms.*

That $\alpha$ has to be strictly positive rather than only non-negative as demanded in Basrak (2000) is observed in Lindskog (2004), since $\alpha = 0$ and a non-zero $\mu_X$ would result in a contradiction to $\mu_X(\overline{\mathbb{R}^d}\backslash\mathbb{R}^d) = 0$. This applies also to all other theorems taken from Basrak (2000).

Several equivalent ways can be used to define multivariate regular variation, for a first account see Mikosch (1999). The following theorem combines the results given in Basrak (2000, Th. 2.1.8, p. 31) and Lindskog (2004, Th. 1.8, 1.14, 1.15 and 1.21). As usual $\|\cdot\|$ below denotes an arbitrary, fixed norm on $\mathbb{R}^d$ and $\mathbb{S}^{d-1}$ the unit sphere in $\mathbb{R}^d$, i.e. $\mathbb{S}^{d-1} = \partial B_1(0)$, w.r.t. to this norm $\|\cdot\|$.

**Theorem 3.9** *Let $X$ be an $\mathbb{R}^d$-valued random variable. Then the following are equivalent:*

*(i) $X$ is regularly varying.*

*(ii) There exists an $\mathbb{S}^{d-1}$-valued random variable $\theta$ such that for some $\alpha > 0$ and every $u > 0$*

$$\frac{P\left(\|X\| > tu, \frac{X}{\|X\|} \in \cdot\right)}{P(\|X\| > t)} \xrightarrow{v} u^{-\alpha}P(\theta \in \cdot)$$

*in $M_+(\mathbb{S}^{d-1})$ for $t \to \infty$.*

*(iii) There exists an $\mathbb{S}^{d-1}$-valued random variable $\theta$ and a positive sequence $(a_n)_{n\in\mathbb{N}}$, $a_n \to \infty$, such that for some $\alpha > 0$ and every $u > 0$*

$$nP\left(\|X\| > ua_n, \frac{X}{\|X\|} \in \cdot\right) \xrightarrow{v} u^{-\alpha} P(\theta \in \cdot)$$

*in $M_+(\mathbb{S}^{d-1})$ for $n \to \infty$.*

*(iv) There exists a positive sequence $(a_n)_{n\in\mathbb{N}}$, $a_n \to \infty$, and a non-zero $\nu_X \in M_+\left(\overline{\mathbb{R}^d}\backslash\{0\}\right)$ with $\nu_X\left(\overline{\mathbb{R}^d}\backslash\mathbb{R}^d\right) = 0$ such that*

$$nP(X \in a_n \cdot) \xrightarrow{v} \nu_X(\cdot)$$

*in $M_+\left(\overline{\mathbb{R}^d}\backslash\{0\}\right)$ for $n \to \infty$.*

*If (iv) holds, then there exists an $\alpha > 0$ such that $\nu_X(tA) = t^{-\alpha}\nu(A)$ for all Borel sets $A$ and $\partial B_\delta(0) \in \mathscr{B}_{\nu_X}$ for all $\delta > 0$. In particular, $\nu_X$ has no atoms.*

*$\alpha$ is called the* index of regular variation, *$P(\theta \in \cdot) \in M_+(\mathbb{S}^{d-1})$ the* spectral measure of regular variation *of $X$ and $\nu_X$ the* measure of regular variation *of $X$.*

Note that our definition of regular variation and (*iv*) is norm-free. This implies that one can take any norm in (*ii*) and (*iii*). However, the spectral measures and, of course, the unit spheres are different for different norms, see Hult and Lindskog (2002) for a discussion of the implications of using different norms.

Moreover, one deducts from (*ii*) that for a regularly varying random variable $X$ and any norm $\|X\|$ is regularly varying with the same index. In particular, one has for every $u > 0$ that $\lim_{t\to\infty} \frac{P(\|X\|>tu)}{P(\|X\|>t)} = u^{-\alpha}$.

**Proof:** The equivalence of $(i) - (iii)$ is shown in Basrak (2000) and that (*iv*) holds for regularly varying $X$ can be found in Mikosch (2003), indeed, it is intuitively rather obvious from (*iii*). The equivalence of (*ii*) and (*iv*), as well as the implications of (*iv*) for $\nu_X$ are shown in Lindskog (2004). Below we give an alternative proof that (*iv*) implies (*ii*) that we developed in temporary ignorance of the work by Lindskog (2004).

Note that due to the scaling relation $\nu_X(tA) = t^{-\alpha}\nu(A)$ and the nondegeneracy of $\nu_X$, we have, if (*iv*) holds, that there exists a relatively compact $K \in \mathcal{B}\left(\overline{\mathbb{R}^d}\backslash\{0\}\right)$ with $\nu_X(K) > 0$ or equivalently that $\nu_X\left((1,\infty]\mathbb{S}^{d-1}\right) > 0$. (Recall the definition of the product of two sets given in Section 2.1.1.)

We will now briefly show $(iv) \Rightarrow (ii)$.

For all sufficiently large $t \in \mathbb{R}$ there is an $n \in \mathbb{N}$ such that $a_n \le t \le a_{n+1}$ (w.l.o.g. one may assume $a_n$ is strictly increasing). For any set $S \in \mathcal{B}(\mathbb{S}^{d-1})$ with $(1,\infty]S \in \mathscr{B}_{\nu_X}$ (see Figure 3.1 for an example of such a set) we have

$$\begin{aligned}\frac{n(n+1)P\left(\|X\| > ua_{n+1}, X/\|X\| \in S\right)}{(n+1)nP(\|X\| > a_n)} &\le \frac{P\left(\|X\| > ut, X/\|X\| \in S\right)}{P(\|X\| > t)} \\ &\le \frac{(n+1)nP\left(\|X\| > ua_n, X/\|X\| \in S\right)}{n(n+1)P(\|X\| > a_{n+1})}.\end{aligned}$$

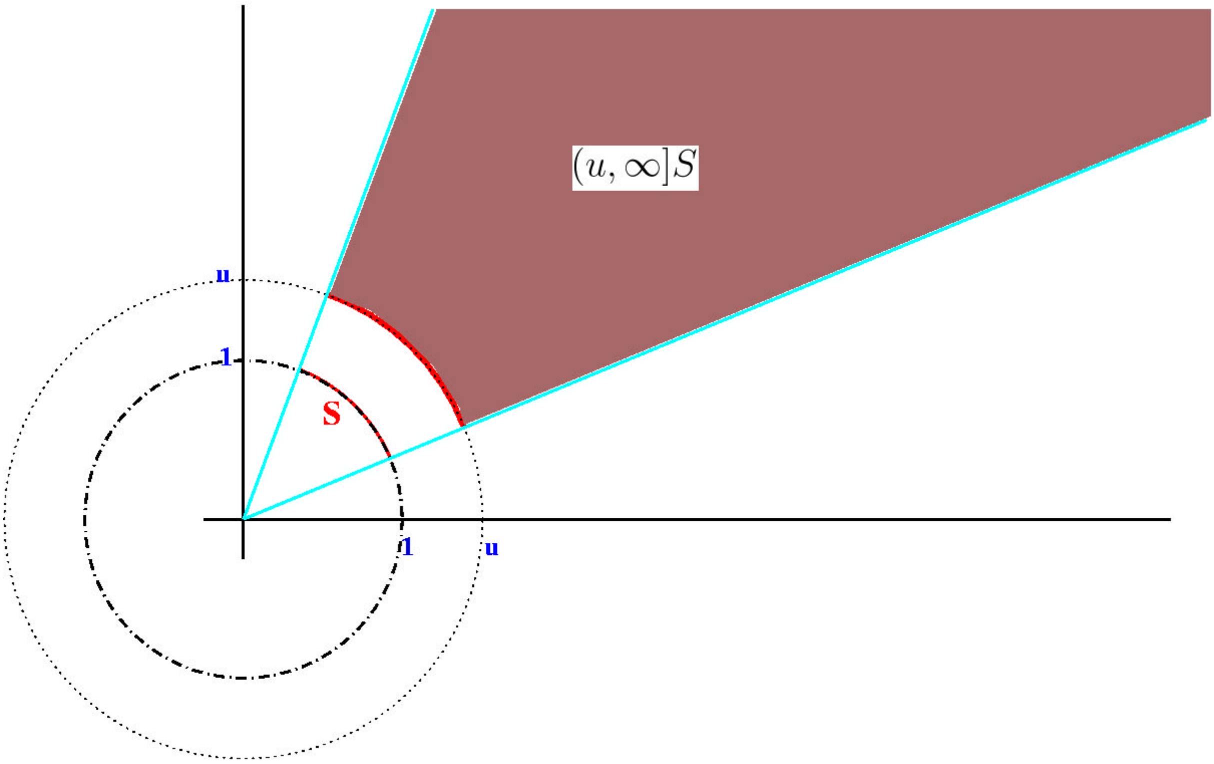


Figure 3.1: A set of the form $(u,\infty]S$ with $S \in \mathcal{B}(\mathbb{S}^1)$ and $u \in \mathbb{R}^+$.

For any $r \in \mathbb{R}^+$ obviously $(r,\infty]S \in \mathscr{B}_{\nu_X}$ and, moreover, $(r,\infty]\mathbb{S}^{d-1} \in \mathscr{B}_{\nu_X}$, since

$$\partial\left((r,\infty]\mathbb{S}^{d-1}\right) = r\mathbb{S}^{d-1} = \partial B_r(0).$$

From $(iv)$ and Proposition 3.4 it follows that for $t \to \infty$ (implies $n \to \infty$)

$$\begin{aligned}
(n+1)P\left(\|X\| > ua_{n+1}, X/\|X\| \in S\right) &\to \nu_X\left((u,\infty]S\right)\\
nP(\|X\| > a_n) &\to \nu_X\left((1,\infty]\mathbb{S}^{d-1}\right)\\
nP\left(\|X\| > ua_n, X/\|X\| \in S\right) &\to \nu_X\left((u,\infty]S\right)\\
(n+1)P(\|X\| > a_{n+1}) &\to \nu_X\left((1,\infty]\mathbb{S}^{d-1}\right).
\end{aligned}$$

Hence for $t \to \infty$:

$$\frac{P\left(\|X\| > ut, X/\|X\| \in S\right)}{P(\|X\| > t)} \to \frac{\nu_X\left((u,\infty]S\right)}{\nu_X\left((1,\infty]\mathbb{S}^{d-1}\right)} = \frac{u^{-\alpha}\nu_X\left((1,\infty]S\right)}{\nu_X\left((1,\infty]\mathbb{S}^{d-1}\right)}.$$

Setting

$$\tilde{\mu}(A) = \frac{\nu_X\left((1,\infty]A\right)}{\nu_X\left((1,\infty]\mathbb{S}^{d-1}\right)}$$

for all $A \in \mathcal{B}(\mathbb{S}^{d-1})$ defines a (Radon) probability measure $\tilde{\mu}$ on $\mathbb{S}^{d-1}$, i.e. $\tilde{\mu}(\cdot) = P(\theta \in \cdot)$ for some $\mathbb{S}^{d-1}$-valued random vector $\theta$, since by $(iv)$ the denominator does not vanish. Using the obvious $S \in \mathscr{B}_{\tilde{\mu}} \Leftrightarrow (1,\infty]S \in \mathscr{B}_{\nu_X}$ ($(1,\infty]\partial S \cup \mathring{S} = \partial((1,\infty]S)$ and $\nu_X(S) = 0$)

Proposition 3.4 gives (*ii*). $\square$

In the following we will mainly rely on the last characterization of regular variation given in the above theorem. To compare some of our results to previous ones we need to compare our notion of regular variation to another one which dates back to the work of Kesten (1973) and has frequently been used when studying the tail behaviour of random recurrence equations (e.g. LePage (1983), Klüppelberg and Pergamenchtchikov (2004) and Saporta (2004a)). The basic idea of this notion of multivariate regular variation is to demand that all linear functionals of an $\mathbb{R}^d$-valued random variable are univariate regularly varying. Formally this means:

**Definition 3.10 (cf. Basrak, Davis and Mikosch (2002a))** *An $\mathbb{R}^d$-valued random variable $X$ is called* regularly varying in the sense of Kesten*, if there exists an $\alpha > 0$ and a slowly varying $L : (0, \infty) \to \mathbb{R}^+$ such that for all $x \in \mathbb{R}^d$*

$$\lim_{u\to\infty} \frac{P(\langle x, X\rangle > u)}{u^{-\alpha}L(u)} = w(x) \tag{3.4}$$

*exists and there is one $x_0 \neq 0$ with $w(x_0) > 0$.*

$\langle\cdot,\cdot\rangle$ denotes the Euclidean scalar product and $\alpha$ is again called index of regular variation. In most cases our notion of regular variation is equivalent to the one of Kesten. Problems occur, however, when $\alpha$ is an integer. The best known result linking the different notions is given in Basrak, Davis and Mikosch (2002a).

**Theorem 3.11 (Basrak, Davis and Mikosch (2002a, Th. 1.1))** *Let $X$ be an $\mathbb{R}^d$-valued random variable.*

*(i) If $X$ is regularly varying with index $\alpha > 0$, then it is regularly varying in the sense of Kesten with the same index.*

*(ii) If $X$ is regularly varying in the sense of Kesten with a noninteger index $\alpha > 0$, then $X$ is regularly varying with the same index and the spectral measure is uniquely determined.*

*(iii) If $X$ assumes values in $[0, \infty)^d$ and satisfies (3.4) for all $x \in [0, \infty)^d\backslash\{0\}$ with a noninteger $\alpha > 0$ and $w(x_0) > 0$ for some $x_0 \in [0, \infty)^d\backslash\{0\}$, then $X$ is regularly varying with index $\alpha$ and the spectral measure is uniquely determined.*

*(iv) If $X$ assumes values in $[0, \infty)^d$ and is regularly varying in the sense of Kesten with index $\alpha$ that is an odd integer, then $X$ is regularly varying with the same index and the spectral measure is uniquely determined.*

From the remark after Theorem 3.9 and the well-known results on univariate regularly varying random variables (cf. e.g. Embrechts, Klüppelberg and Mikosch (1997)) we immediately infer:

**Proposition 3.12** *Let $X$ be an $\mathbb{R}^d$-valued multivariate regularly varying random variable with index $\alpha$. Then:*

$$\begin{array}{ll} X \in L^\beta & \forall\, 0 < \beta < \alpha \\ X \notin L^\beta & \forall\, \beta > \alpha \end{array}$$

Observe that $X$ may or may not be in $L^\alpha$. We shall briefly show this for the case $d = 1$: Let $X$ be an $\mathbb{R}^+$-valued random variable with regularly varying tail of index $\alpha > 0$. There exists a slowly varying function $l(x)$ such that $P(X > x) = l(x)x^{-\alpha}$ and

$$\begin{aligned} E(|X|^\alpha) &= \int_0^\infty P(|X|^\alpha > x)\, dx \\ &= \int_0^\infty P(|X| > x^{1/\alpha})\, dx \\ &= \int_0^\infty l(x^{1/\alpha})\frac{1}{x}\, dx. \end{aligned}$$

Thus it suffices to show that the very last integral can be finite as well as infinite. Consider

$$l(x) := (\log(x^\alpha))^{-2}, \quad x \geq 2,$$

then $l(\cdot)$ is slowly varying, since l'Hospital's rule gives

$$\lim_{y\to\infty} \frac{\log(y)}{\log(yx^\alpha)} = \lim_{y\to\infty} \frac{1/y}{x^\alpha/(yx^\alpha)} = 1$$

and thus

$$\lim_{t\to\infty} \frac{(\log(t^\alpha x^\alpha))^{-2}}{(\log(t^\alpha))^{-2}} \stackrel{y:=t^\alpha}{=} \lim_{y\to\infty} \left(\frac{\log(y)}{\log(yx^\alpha)}\right)^2 = 1$$

for any $x > 0$. One obtains

$$\int_2^\infty \frac{1}{x(\log x)^2}\, dx = \int_{\log 2}^\infty \frac{1}{e^y y^2} e^y\, dy = \int_{\log 2}^\infty \frac{1}{y^2} dy < \infty.$$

Thus it is possible for the above integral to be finite and thus $E|X|^\alpha < \infty$. That the integral can as well be infinite is obvious using e.g. $l(x) = 1$.

## 3.3 Transformations of Regularly Varying Random Variables

Now we summarize results on the regular variation of combinations and transformations of regularly varying random variables. Finally, we will give a new theorem on the behaviour of series of linearly transformed regularly varying i.i.d. random variables, which is a considerable, but (at least in our set-up) straightforward extension of Resnick and Willekens (1991, Th 2.1).

We first note some well-known result on the combination of i.i.d. multivariate regularly varying random variates. As usual $\epsilon_0$ denotes the Dirac measure w.r.t. 0.

**Lemma 3.13** *Let $X_1, X_2, \ldots, X_k$ with $k \in \mathbb{N}$ be i.i.d. regularly varying $\mathbb{R}^d$-valued random variables with index $\alpha$, measure $\nu_X$ and normalizing sequence $(a_n)$ such that (iv) in Theorem 3.9 holds. Then $X = (X_1^\mathsf{T}, \ldots, X_k^\mathsf{T})^\mathsf{T}$ is a regularly varying $\mathbb{R}^{kd}$-valued random variable with index $\alpha$ and measure*

$$\begin{aligned}\nu(dx_1, \ldots, dx_k) &= \nu_X(dx_1)\epsilon_0(dx_2, dx_3, \ldots, dx_k) + \nu_X(dx_2)\epsilon_0(dx_1, dx_3, dx_4, \ldots, dx_k) \\ &\quad + \ldots + \nu_X(dx_k)\epsilon_0(dx_1, dx_2, \ldots, dx_{k-1})\end{aligned}$$

*(where $\nu_X$ is assumed to be extended to $\overline{\mathbb{R}^d}$ by setting $\nu_X(\{0\}) = 0$ and the usual convention $0 \cdot \infty = 0$ is employed).*

Recall that any relatively compact set is bounded away from zero (in $\mathbb{R}^{kd}$) and thus for a relatively compact set of the form $A_1 \times \ldots \times A_k$ with $A_i \in \mathcal{B}(\overline{\mathbb{R}^d})$, i.e. a "rectangle", there is at least one index $j$ such that $A_j$ is bounded away from zero (in $\mathbb{R}^d$) and, thus, assuming $j = 1$, one has $\nu(A_1 \times \ldots \times A_k) = \nu(A_1)\epsilon_0(A_2 \times \ldots \times A_k)$. In particular, if there is also an $l \in 2, \ldots, k$ such that $A_l$ does not contain 0, then $\nu(A_1 \times \ldots \times A_k) = 0$. This shows that $\nu$ concentrates on the "axes". For $d > 1$ this means that it concentrates on $\left(\mathbb{R}^d \times 0_{\mathbb{R}^d} \times 0_{\mathbb{R}^d} \times \ldots \times 0_{\mathbb{R}^d}\right) \cup \left(0_{\mathbb{R}^d} \times \mathbb{R}^d \times 0_{\mathbb{R}^d} \times \ldots \times 0_{\mathbb{R}^d}\right) \cup \ldots \cup \left(0_{\mathbb{R}^d} \times \ldots \times 0_{\mathbb{R}^d} \times \mathbb{R}^d\right)$. Figure 3.2 exemplifies this fact that for two independent real-valued regularly varying random variables (in this example symmetric Cauchy ones, i.e. the index of regular variation is one) the measure of regular variation concentrates on $\mathbb{R} \times \{0\}$ and $\{0\} \times \mathbb{R}$, which means that the two random variables are never both "large" at the same time.
Unfortunately a proof in full rigour seems to be lacking in the literature (the necessary arguments are, however, briefly stated in the proof of Lindskog (2004, Th. 1.28) and for the case of $\mathbb{R}^+$-valued random variables in Resnick (1987, p. 227)). We need the following auxiliary result:

**Lemma 3.14 (cf. Lindskog (2004, Lemma 1.10 and remark thereafter))** *Let $\mu, \mu_1, \mu_2, \ldots$ be in $M_+(\overline{\mathbb{R}^d} \backslash \{0\})$ such that $\mu(\overline{\mathbb{R}^d} \backslash \mathbb{R}^d) = 0$ and $\mu(uB) = u^{-\alpha}\mu(B)$ holds for all $u > 0$ and Borel sets $A$ with some $\alpha > 0$. Assume that*

$$\mu_n\left([a_1, b_1) \times [a_2, b_2) \times \ldots \times [a_d, b_d)\right) \to \mu\left([a_1, b_1) \times [a_2, b_2) \times \ldots \times [a_d, b_d)\right)$$

*as $n \to \infty$ for all $a_1, a_2, \ldots, a_d, b_1, b_2, \ldots b_d \in \mathbb{R}$ with $a_i < b_i$ such that $[a_1, b_1) \times [a_2, b_2) \times \ldots \times [a_d, b_d)$ is bounded away from zero (in $\mathbb{R}^d$) and $\mu$-boundaryless. Then $\mu_n \xrightarrow{v} \mu$.*

Lindskog (2004, Lemma 1.10) states the above result assuming only that $(a_1, \ldots, a_b)^\mathsf{T}$ and $(b_1, \ldots, b_d)^\mathsf{T}$ are non-zero instead of demanding that the rectangles be $\mu$-boundaryless. This is due to the fact that he concludes in the arguments leading to the above Lemma that for any $\mu$ satisfying the above conditions all rectangles bounded away from zero are necessarily $\mu$-boundaryless. Unfortunately, this does not hold in general. A counterexample can be given using the set-up of Lemma 3.13. Take $d = 1, k = 2$ and $[a_1, b_1) \subset \mathbb{R}$ bounded away from zero and such that $\nu_X([a_1, b_1)) > 0$. Note that $nP((X_1, X_2)^\mathsf{T} \in \cdot)$ plays the role of $\mu_n$, and $\nu$ as defined in Lemma 3.13 satisfies all conditions on $\mu$ in the above Lemma. Take now $a_2 = 0$ and $b_2 > 0$. Then $[a_1, b_1) \times [a_2, b_2)$ is bounded away from zero in $\mathbb{R}^2$ and we have $\partial\left([a_1, b_1) \times [a_2, b_2)\right) = \{a_1\} \times [a_2, b_2] \cup \{b_1\} \times [a_2, b_2] \cup [a_1, b_1] \times \{a_2\} \cup [a_1, b_1] \times \{b_2\}$. Noting that $\nu\left([a_1, b_1] \times \{a_2\}\right) = 1 \cdot \nu_X([a_1, b_1)] > 0$, we see that $[a_1, b_1) \times [a_2, b_2)$ is not

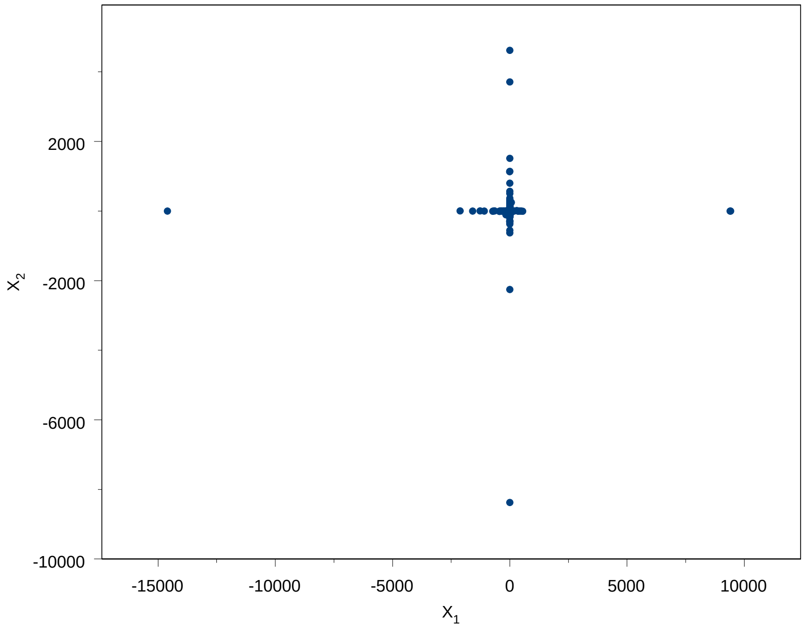


Figure 3.2: Scatter plot of 10.000 simulations from a pair $(X_1, X_2)$ of independent standard Cauchy random variables

$\nu$-boundaryless. By the way, the same result is obtained when choosing $b_2 = 0$ and $a_2 < 0$. The crucial point is that $\{0\} \times [a_1, b_1]$ is not contained in any appropriately scaled unit sphere (w.r.t. $\|\cdot\|_\infty$).

**Proof of Lemma 3.13:** Note that by the definition of $\nu$ we have that $\nu(\overline{\mathbb{R}^d}\backslash\mathbb{R}^d) = 0$ and the scaling property $\nu(u\cdot) = u^{-\alpha}\nu(\cdot)$ holds with $\alpha$ being the index of regular variation of the random variables $X_i$. Hence, we can apply the above Lemma. For some set $A$ of the form $[a_{11}, b_{11}) \times [a_{12}, b_{12}) \times \ldots \times [a_{1d}, b_{1d}) \times [a_{21}, b_{21}) \times [a_{22}, b_{22}) \times \ldots \times [a_{2d}, b_{2d}) \times \ldots \times [a_{k1}, b_{k1}) \times [a_{k2}, b_{k2}) \times \ldots \times [a_{kd}, b_{kd})$ which is bounded away from 0 in $\mathbb{R}^{kd}$ there is at least one index $j$ such that $A_j := [a_{j1}, b_{j1}) \times [a_{j2}, b_{j2}) \times \ldots \times [a_{jd}, b_{jd})$ is bounded away from 0 in $\mathbb{R}^d$. For the ease of notation assume w.l.o.g. $j = 1$. Now only two cases may occur. Either there is an $l \in 2, \ldots, k$ such that $A_l$ is bounded away from 0 and so we have

$$nP(X \in a_n A) \leq \frac{nP(X_1 \in a_n A_1) nP(X_l \in a_n A_l)}{n} \to 0$$

as $n \to \infty$, since $nP(X_1 \in a_n A_1) \to \nu_X(A_1) < \infty$ and $nP(X_l \in a_n A_l) \to \nu_X(A_l) < \infty$. This gives $nP(X \in a_n A) \to \nu(A)$.

Otherwise $\overline{A_2 \times \ldots \times A_k}$ contains the zero in $\mathbb{R}^{(k-1)d}$. Now we again distinguish three cases. If 0 is in the interior (w.r.t. $\mathbb{R}^{(k-1)d}$) of $A_2 \times \ldots \times A_k$, then

$$\lim_{n\to\infty} P((X_2^\mathsf{T}, \ldots, X_k^\mathsf{T})^\mathsf{T} \in a_n(A_2 \times \ldots \times A_k)) = 1$$

and so

$$nP(X \in a_n A) = nP(X_1 \in a_n A_1) P((X_2^\mathsf{T}, \ldots, X_k^\mathsf{T})^\mathsf{T} \in a_n(A_2 \times \ldots \times A_k)) \to \nu_X(A_1) = \nu(A).$$

If $\nu_X(A_1) = 0$, we have that

$$nP(X \in a_n A) \leq nP(X_1 \in a_n A_1) \to 0$$

and thus $nP(X \in a_n A) \to \nu(A)$. The case remaining to consider is that $\nu_X(A_1) > 0$ and 0 is on the boundary (w.r.t. $\mathbb{R}^{(k-1)d}$) of $A_2 \times \ldots \times A_k$. Then $A_1 \times \{0_{\mathbb{R}^{(k-1)d}}\} \subseteq \partial A$, but $\nu(A_1 \times \{0_{\mathbb{R}^{(k-1)d}}\}) = \nu_X(A_1)$ and thereby $A$ is not $\nu$-boundaryless, so the behaviour of $nP(X \in a_n A)$ does not matter when considering vague convergence to $\nu$.

A straightforward application of the foregoing Lemma now concludes the proof. □

The following extension can be shown using basically the same arguments, but a slightly more tedious notation.

**Lemma 3.15** *Let $X_1, X_2, \ldots, X_k$ with $k \in \mathbb{N}$ be independent regularly varying $\mathbb{R}^d$-valued random variables with common index $\alpha$, measures $\nu_{X_1}, \ldots, \nu_{X_k}$ and a common normalizing sequence $(a_n)$ such that (iv) in Theorem 3.9 holds. Then $X = (X_1^\mathsf{T}, \ldots, X_k^\mathsf{T})^\mathsf{T}$ is a regularly varying $\mathbb{R}^{kd}$-valued random variable with index $\alpha$ and measure*

$$\begin{aligned}\nu(dx_1, \ldots, dx_k) &= \nu_{X_1}(dx_1)\epsilon_0(dx_2, dx_3, \ldots, dx_k) + \nu_{X_2}(dx_2)\epsilon_0(dx_1, dx_3, dx_4 \ldots, dx_k) \\ &\quad + \ldots + \nu_{X_k}(dx_k)\epsilon_0(dx_1, dx_2, \ldots, dx_{k-1})\end{aligned}$$

*(where $\nu_{X_i}$ is assumed to be extended to $\overline{\mathbb{R}^d}$ by setting $\nu_{X_i}(\{0\}) = 0$ and the usual convention $0 \cdot \infty = 0$ is employed).*

It should also be obvious that the above result can be extended to the case where $X_i$ assumes values in $\mathbb{R}^{d_i}$ for possibly different $d_i$. Since we do not at all need such a result later, we refrain from stating it in its details.

Furthermore, it can be shown that comparatively light-tailed random linear transformations preserve regular variation. This extension of a result from Breiman (1965) to the multivariate set-up is due to Basrak, Davis and Mikosch (2002b) (see also Basrak (2000, Prop. 2.1.18, Cor. 2.1.19)).

**Theorem 3.16 (cf. Basrak, Davis and Mikosch (2002b, Prop. A1), Basrak (2000, Cor. 2.1.19))** *Let $X$ be an $\mathbb{R}^d$-valued random variable that is regularly varying with index $\alpha$, measure $\nu$ and normalizing sequence $(a_n)$, i.e. $nP(X \in a_n \cdot) \xrightarrow{v} \nu(\cdot)$. Assume that $A$ is an $M_{qd}(\mathbb{R})$-valued random variable independent of $X$ and $A \in L^\gamma$ for some $\gamma > \alpha$. Then*

$$nP(AX \in a_n \cdot) \xrightarrow{v} \tilde{\nu}(\cdot) := E(\nu \circ A^{-1}(\cdot))$$

*in $M_+\left(\overline{\mathbb{R}^q} \backslash \{0\}\right)$.*

*In particular, provided there is a relatively compact $K \in \mathcal{B}\left(\overline{\mathbb{R}^q} \backslash \{0\}\right)$ such that*

$$E\left(\nu\left(A^{-1}(K)\right)\right) > 0,$$

*$AX$ is regularly varying with index $\alpha$, measure $\tilde{\nu}$ and normalizing sequence $(a_n)$.*

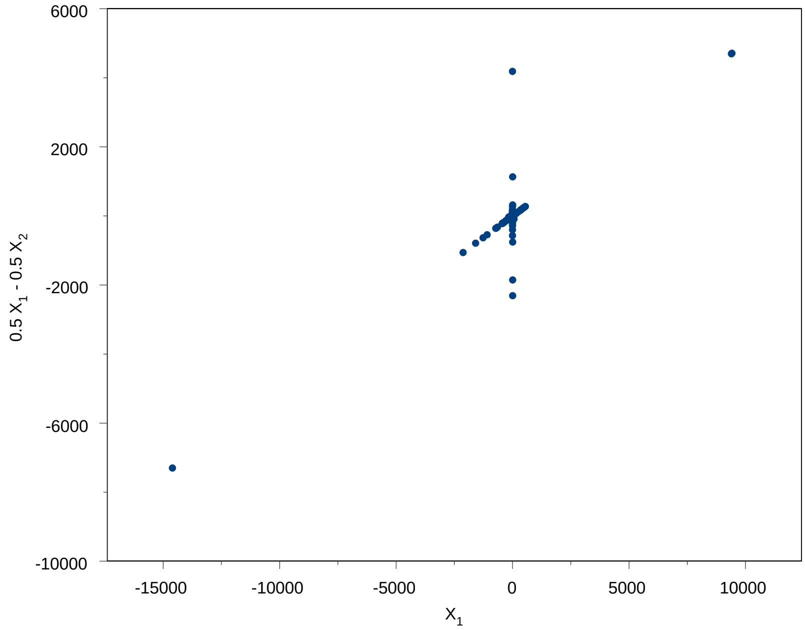


Figure 3.3: Scatter plot of $(X_1, X_1/2 - X_2/2)$ using the 10.000 simulations from a pair $(X_1, X_2)$ of independent standard Cauchy random variables from Figure 3.2

Note that $A^{-1}$ does not mean the inverse of the matrix $A$, but that the pre-image under the linear map $A$ is taken. $A$ does not have to be invertible or of full rank. Note, moreover, that the pre-image of a set not containing 0 under a continuous linear mapping never contains zero and pre-images of sets bounded away from zero are bounded away from zero.

To exemplify this result, Figure 3.3 depicts the independent Cauchy random variables simulated for Figure 3.2 after applying the deterministic linear map $A = \begin{pmatrix} 1 & 0 \\ 1/2 & -1/2 \end{pmatrix}$. A straightforward calculation shows that the measure of regular variation of the transformed pair $(X_1, X_1/2 - X_2/2)$ concentrates on $(1, 1/2)^{\mathsf{T}}\mathbb{R}$ and $\{0\} \times \mathbb{R}$.

**Proof:** For a proof of the first, i.e. the vague convergence, part we refer to Basrak, Davis and Mikosch (2002b).

The second part is to be found in Basrak (2000). However, as the proof is rather short and adds some important insights, we include it here. Note that $\tilde{\nu}\left(\overline{\mathbb{R}^q}\backslash\mathbb{R}^q\right) = 0$ is implied by the fact that $\nu(\overline{\mathbb{R}^d}\backslash\mathbb{R}^d) = 0$ and every element of $M_{qd}(\mathbb{R})$ maps $\mathbb{R}^d$ to $\mathbb{R}^q$. The assumption of the existence of a relatively compact $K \in \mathcal{B}\left(\overline{\mathbb{R}^q}\backslash\{0\}\right)$ such that $E\left(\nu\left(A^{-1}(K)\right)\right) > 0$ gives that $\tilde{\nu}$ is non-zero. $\tilde{\nu}$ is a Radon, i.e. locally finite, measure, since any relatively compact $K$ is contained in a set of the form $(r, \infty]\mathbb{S}^{q-1}$ for an appropriate $r > 0$ and

$$\begin{aligned}
\tilde{\nu}((r,\infty]\mathbb{S}^{q-1}) &= E\left(\nu\left(A^{-1}((r,\infty]\mathbb{S}^{q-1})\right)\right) = E\left(\nu\left(A^{-1}((r,\infty]\mathbb{S}^{q-1})\right) I_{M_{qd}(\mathbb{R})\backslash\{0\}}(A)\right) \\
&\leq E\left(\nu\left(\|A\|^{-1}(r,\infty]\mathbb{S}^{q-1}\right) I_{M_{qd}(\mathbb{R})\backslash\{0\}}(A)\right) \\
&= E\left(\|A\|^{\alpha}\nu\left((r,\infty]\mathbb{S}^{q-1}\right) I_{M_{qd}\backslash\{0\}}(A)\right) = \nu\left((r,\infty]\mathbb{S}^{q-1}\right) E\left(\|A\|^{\alpha}\right) < \infty
\end{aligned}$$

using that $A^{-1}((r,\infty]\mathbb{S}^{q-1}) = \emptyset$, if $A = 0$, and $A^{-1}((r,\infty]\mathbb{S}^{q-1}) \subseteq \|A\|^{-1}(r,\infty]\mathbb{S}^{q-1}$ otherwise. Combining these facts shows that $(iv)$ in Theorem 3.9 holds and so $AX$ is regularly varying with measure $\tilde{\nu}$ and normalizing sequence $(a_n)$. Moreover, for any $u > 0$ and Borel set $C$ we have $\tilde{\nu}(uC) = E(\nu(A^{-1}(uC))) = E(\nu(uA^{-1}(C))) = E(u^{-\alpha}\nu(A^{-1}(C)))$ and thus the index of regular variation is $\alpha$. □

By combining the previous results we can now show that finite sums of appropriate random linear transformations of independent multivariate regularly varying random variables are again regularly varying. This is an extension of Resnick and Willekens (1991, Eq. (2.4)) and can also be interpreted as a generalization of Davis and Resnick (1996, Lemma 2.1) (cf. also Resnick (1987, p. 225)) to the multivariate random case, a similar one is given in Konstantinides and Mikosch (2004, p. 10). Whereas the previous results were formulated for random variables in $(\mathbb{R}^+)^d$, respectively $\mathbb{R}^+$, only, we consider general $\mathbb{R}^d$-valued random variables. However, the reasoning in the proof given below is basically the same as for the previous results except that using Theorem 3.16 rather than the more general Theorem 3.5 shortens the argumentation.

**Theorem 3.17** *Let $X_1, X_2, \ldots, X_k$ with $k \in \mathbb{N}$ be independent regularly varying $\mathbb{R}^d$-valued random variables with common index $\alpha$, measures $\nu_{X_1}, \ldots, \nu_{X_k}$ and a common normalizing sequence $(a_n)$ such that $(iv)$ in Theorem 3.9 holds. Assume, moreover, that $A_1, \ldots, A_k$ are $M_{qd}(\mathbb{R})$-valued random variables independent of $X = (X_1^\intercal, \ldots, X_k^\intercal)^\intercal$ and $A_i \in L^\gamma\ \forall i \in \{1, \ldots, k\}$ for some $\gamma > \alpha$. Then*

$$nP\left(\sum_{i=1}^{k} A_i X_i \in a_n \cdot\right) \xrightarrow{v} \nu(\cdot) := \sum_{i=1}^{k} E\left(\nu_{X_i} \circ A_i^{-1}(\cdot)\right)$$

*in $M_+\left(\overline{\mathbb{R}^q}\backslash\{0\}\right)$.*

*In particular, provided there is a relatively compact $K \in \mathcal{B}\left(\overline{\mathbb{R}^q}\backslash\{0\}\right)$ and an index $j \in \{1, \ldots, k\}$ such that $E\left(\nu_{X_j}\left(A_j^{-1}(K)\right)\right) > 0$, $Y = \sum_{i=1}^k A_i X_i$ is regularly varying with index $\alpha$, measure $\nu$ and normalizing sequence $(a_n)$.*

One possible immediate extension is that the random variables $X_i$ are $\mathbb{R}^{d_i}$-valued with possibly different $d_i$. Furthermore, note that the matrices $A_1, \ldots, A_k$ are not assumed to be independent, but may have *any* dependence structure as long as $\{A_i\}_{i=1,\ldots,k}$ and $\{X_i\}_{i=1,\ldots,k}$ are independent. This actually is the crucial fact, why this theorem will turn out to be most helpful in later sections.

**Proof:** By Lemma 3.15 we have that $X = (X_1^\intercal, \ldots, X_k^\intercal)^\intercal$ is a regularly varying $\mathbb{R}^{kd}$-valued random variable with index $\alpha$, normalizing sequence $(a_n)$ and measure

$$\begin{aligned}\bar{\nu}(dx_1, \ldots, dx_k) &= \nu_{X_1}(dx_1)\epsilon_0(dx_2, dx_3, \ldots, dx_k) + \nu_{X_2}(dx_2)\epsilon_0(dx_1, dx_3, dx_4 \ldots, dx_k) \\ &\quad + \ldots + \nu_{X_k}(dx_k)\epsilon_0(dx_1, dx_2, \ldots, dx_{k-1}).\end{aligned}$$

By setting

$$A = \begin{pmatrix} A_1 & A_2 & \ldots & A_k \end{pmatrix}$$

we obtain an $M_{q,kd}(\mathbb{R})$-valued random variable that is independent of $X$ and in $L^\gamma$ (using Th. 2.14 and Cor. 2.15). Thus Theorem 3.16 gives that

$$nP(AX \in a_n \cdot) = nP\left(\sum_{i=1}^{k} A_i X_i \in a_n \cdot\right) \xrightarrow{v} \nu(\cdot) := E(\bar{\nu} \circ A^{-1}(\cdot)).$$

Now we only need to analyse $\nu$. Let $C$ be any measurable subset of $\overline{\mathbb{R}^q}\backslash\{0\}$. It is immediate that $A^{-1}(C) = \left(A_1^{-1}(C) \times 0_{\mathbb{R}^{(k-1)d}}\right) \cup \left(0_{\mathbb{R}^d} \times A_2^{-1}(C) \times 0_{\mathbb{R}^{(k-2)d}}\right) \cup \ldots \cup \left(0_{\mathbb{R}^{(k-1)d}} \times A_k^{-1}(C)\right) \cup \bigcup_{i=1}^{k} \bigcup_{j=i+1}^{k} C_{ij}$ for appropriately chosen sets $C_{ij} \subseteq \mathbb{R}^{kd}$ such that $C_{ij} \cap \mathbb{R}^{(i-1)d} \times 0_{\mathbb{R}^d} \times \mathbb{R}^{(k-i)d} = \emptyset$ as well as $C_{ij} \cap \mathbb{R}^{(j-1)d} \times 0_{\mathbb{R}^d} \times \mathbb{R}^{(k-j)d} = \emptyset$ holds. In other words $C_{ij}$ is the part of $A^{-1}(C)$ where both the $i$-th and $j$-th ("$d$-dimensional") coordinates are different from zero. Note that necessarily $0 \notin A_i^{-1}(C)$ and thus the sets $A_1^{-1}(C) \times 0_{\mathbb{R}^{(k-1)d}}$, $0_{\mathbb{R}^d} \times A_2^{-1}(C) \times 0_{\mathbb{R}^{(k-2)d}}, \ldots, 0_{\mathbb{R}^{(k-1)d}} \times A_k^{-1}(C)$ are pairwise disjoint. Furthermore, it follows from the way we have selected the $C_{ij}$ that $\bar{\nu}(C_{ij}) = 0$. Hence, we obtain

$$\bar{\nu} \circ A^{-1}(C) = \sum_{i=1}^{k} \nu_{X_i}(A_i^{-1}(C)) = \sum_{i=1}^{k} \nu_{X_i} \circ A_i^{-1}(C).$$

Since $C$ was arbitrary, taking the expectation gives

$$nP\left(\sum_{i=1}^{k} A_i X_i \in a_n \cdot\right) \xrightarrow{v} \nu(\cdot) = E\left(\sum_{i=1}^{k} \nu_{X_i} \circ A_i^{-1}(\cdot)\right).$$

The other claims are now shown as in the proof of Theorem 3.16. □

We now aim at generalizing the last theorem to series of linearly transformed i.i.d. random variables that satisfy an appropriate summability condition. It will turn out that a straightforward extension of Resnick and Willekens (1991, Th. 2.1), who consider $(\mathbb{R}^+)^d$-valued random variables, to general $\mathbb{R}^d$-valued ones is possible. Basically all their arguments carry through in our set-up. Below we shall give a proof for the general case, which basically imitates the second part of the proof in Resnick and Willekens (1991) and uses Resnick and Willekens (1991, Th. 2.1) for $\mathbb{R}^+$-valued random variables. Thereby we avoid repeating the highly technical first part of Resnick and Willekens' proof employing Pratt's lemma. Let us thus first recall the result on $(\mathbb{R}^+)^d$. $\|\cdot\|$ denotes any fixed norms on $\mathbb{R}^d$ and $\mathbb{R}^q$ and their induced operator norm. By $M_{qd}((\mathbb{R}^+)^d)$ we denote the real $q \times d$ matrices that have only non-negative entries.

**Theorem 3.18 (cf. Resnick and Willekens (1991, Th. 2.1))** *Let $X = (X_k)_{k\in\mathbb{N}_0}$ be a sequence of i.i.d. regularly varying $(\mathbb{R}^+)^d$-valued random variables with index $\alpha$, measure $\nu$ and normalizing sequence $(a_n)$ such that (iv) in Theorem 3.9 holds. Assume, moreover, that $A = (A_k)_{k\in\mathbb{N}_0}$ is a sequence of $M_{qd}(\mathbb{R}^+)$-valued random variables independent of $X$.*

*If $\alpha < 1$, assume that there is an $0 < \eta < \alpha$ with $\alpha + \eta < 1$ such that $A_k \in L^{\alpha+\eta}$ for all $k \in \mathbb{N}_0$ and*

$$\sum_{k=0}^{\infty} E\left(\|A_k\|^{\alpha+\eta}\right) < \infty \text{ as well as } \sum_{k=0}^{\infty} E\left(\|A_k\|^{\alpha-\eta}\right) < \infty. \tag{3.5}$$

*If $\alpha \geq 1$, assume that there is an $0 < \eta < \alpha$ such that $A_k \in L^{\alpha+\eta}$ for all $k \in \mathbb{N}_0$ and*

$$\sum_{k=0}^{\infty} E\left(\|A_k\|^{\alpha+\eta}\right)^{1/(\alpha+\eta)} < \infty, \qquad \sum_{k=0}^{\infty} E\left(\|A_k\|^{\alpha-\eta}\right)^{1/(\alpha+\eta)} < \infty. \tag{3.6}$$

*Then the tail behaviour of $Y = \sum_{k=0}^{\infty} A_k X_k$ is given by*

$$nP\left(\sum_{k=0}^{\infty} A_k X_k \in a_n \cdot\right) \xrightarrow{v} \tilde{\nu}(\cdot) := \sum_{k=0}^{\infty} E\left(\nu \circ A_k^{-1}(\cdot)\right)$$

*in $M_+\left(\overline{(\mathbb{R}^+)^q}\backslash\{0\}\right)$. When taking pre-images the linear operators $A_k$ are regarded as mappings $A_k : (\mathbb{R}^+)^d \to (\mathbb{R}^+)^d$.*

*In particular, provided there is a relatively compact $K \in \mathcal{B}(\overline{(\mathbb{R}^+)^q}\backslash\{0\})$ and an index $j \in \mathbb{N}_0$ such that $E\left(\nu\left(A_j^{-1}(K)\right)\right) > 0$, $Y = \sum_{k=0}^{\infty} A_k X_k$ is regularly varying with index $\alpha$, measure $\tilde{\nu}$ and normalizing sequence $(a_n)$.*

For a proof of the vague convergence part we refer to Resnick and Willekens (1991). In the case $q = d = 1$ we can regard the matrices $A_k$ as mappings from $\mathbb{R}$ to $\mathbb{R}$. Note that the equivalence of all norms over finite dimensional linear spaces ensures that condition (3.5), resp. (3.6), is independent of the actually employed norm.

**Proof:** We shall only show that $\tilde{\nu}$ is locally finite, since the remainder is obvious in view of the arguments given for Theorem 3.16.

First we note that for all $k \in \mathbb{N}_0$

$$\|A_k\|^{\alpha} \leq \max\{\|A_k\|^{\alpha+\eta}, \|A_k\|^{\alpha-\eta}\} \leq \|A_k\|^{\alpha+\eta} + \|A_k\|^{\alpha-\eta}$$

which implies

$$E(\|A_k\|^{\alpha}) \leq E(\|A_k\|^{\alpha+\eta} + \|A_k\|^{\alpha-\eta}).$$

For $\alpha < 1$ one thus immediately obtains

$$\sum_{k=0}^{\infty} E\left(\|A_k\|^{\alpha}\right) < \infty$$

from (3.5). For $\alpha \geq 1$ the same follows using (3.6) after noting that for large enough $k$ one has that $E(\|A_k\|^{\alpha+\eta} + \|A_k\|^{\alpha-\eta}) < 1$ and so

$$\begin{aligned} E(\|A_k\|^{\alpha}) &\leq E(\|A_k\|^{\alpha+\eta} + \|A_k\|^{\alpha-\eta}) \\ &\leq E(\|A_k\|^{\alpha+\eta} + \|A_k\|^{\alpha-\eta})^{1/(\alpha+\eta)} \\ &\leq E(\|A_k\|^{\alpha+\eta})^{1/(\alpha+\eta)} + E(\|A_k\|^{\alpha-\eta})^{1/(\alpha+\eta)}, \end{aligned}$$

because $|a+b|^r \leq |a|^r + |b|^r$ for all $a, b \in \mathbb{R}$ and $0 < r \leq 1$ (see e.g. Loève (1977, p. 157)).

As in the proof of Theorem 3.16 it suffices to consider the sets $(r, \infty]\mathbb{S}_+^{q-1}$ for $r > 0$, where $\mathbb{S}_+^{q-1} := \mathbb{S}^{q-1} \cap (\mathbb{R}^+)^q$ is the "unit sphere" in $(\mathbb{R}^+)^q$. From the above result we obtain

$$\tilde{\nu}((r, \infty]\mathbb{S}_+^{q-1}) = \sum_{k=0}^{\infty} E\left(\nu\left(A_k^{-1}((r, \infty]\mathbb{S}_+^{q-1})\right)\right)$$

$$
\begin{aligned}
&= \sum_{k=0}^{\infty} E\left(\nu\left(A_k^{-1}((r,\infty]\mathbb{S}_+^{q-1})\right) I_{M_{qd}(\mathbb{R})\setminus\{0\}}(A_k)\right) \\
&\leq \sum_{k=0}^{\infty} E\left(\nu\left(\|A_k\|^{-1}(r,\infty]\mathbb{S}_+^{q-1}\right) I_{M_{qd}(\mathbb{R})\setminus\{0\}}(A_k)\right) \\
&= \sum_{k=0}^{\infty} E\left(\|A_k\|^{\alpha}\nu\left((r,\infty]\mathbb{S}_+^{q-1}\right) I_{M_{qd}\setminus\{0\}}(A_k)\right) \\
&= \nu\left((r,\infty]\mathbb{S}_+^{q-1}\right)\sum_{k=0}^{\infty} E\left(\|A_k\|^{\alpha}\right) < \infty
\end{aligned}
$$

using that $A_k^{-1}((r,\infty]\mathbb{S}_+^{q-1}) = \emptyset$, if $A_k = 0$, and $A_k^{-1}((r,\infty]\mathbb{S}_+^{q-1}) \subseteq \|A_k\|^{-1}(r,\infty]\mathbb{S}_+^{q-1}$ otherwise. □

Getting rid of the positivity restrictions on the matrices $A_k$ and i.i.d. $X_k$ we obtain:

**Theorem 3.19** *Let $X = (X_k)_{k\in\mathbb{N}_0}$ be a sequence of i.i.d. regularly varying $\mathbb{R}^d$-valued random variables with index $\alpha$, measure $\nu$ and normalizing sequence $(a_n)$ such that (iv) in Theorem 3.9 holds. Assume, moreover, that $A = (A_k)_{k\in\mathbb{N}_0}$ is a sequence of $M_{qd}(\mathbb{R})$-valued random variables independent of $X$.*

*If $\alpha < 1$, assume that there is an $0 < \eta < \alpha$ with $\alpha + \eta < 1$ such that $A_k \in L^{\alpha+\eta}$ for all $k \in \mathbb{N}_0$ and*

$$\sum_{k=0}^{\infty} E\left(\|A_k\|^{\alpha+\eta}\right) < \infty \text{ as well as } \sum_{k=0}^{\infty} E\left(\|A_k\|^{\alpha-\eta}\right) < \infty. \tag{3.7}$$

*If $\alpha \geq 1$, assume that there is an $0 < \eta < \alpha$ such that $A_k \in L^{\alpha+\eta}$ for all $k \in \mathbb{N}_0$ and*

$$\sum_{k=0}^{\infty} E\left(\|A_k\|^{\alpha+\eta}\right)^{1/(\alpha+\eta)} < \infty, \qquad \sum_{k=0}^{\infty} E\left(\|A_k\|^{\alpha-\eta}\right)^{1/(\alpha+\eta)} < \infty. \tag{3.8}$$

*Then the tail behaviour of $Y = \sum_{k=0}^{\infty} A_k X_k$ is given by*

$$nP\left(\sum_{k=0}^{\infty} A_k X_k \in a_n \cdot\right) \xrightarrow{v} \tilde{\nu}(\cdot) := \sum_{k=0}^{\infty} E\left(\nu \circ A_k^{-1}(\cdot)\right) \tag{3.9}$$

*in $M_+\left(\overline{\mathbb{R}^q}\setminus\{0\}\right)$.*

*In particular, provided there is a relatively compact $K \in \mathcal{B}\left(\overline{\mathbb{R}^q}\setminus\{0\}\right)$ and an index $j \in \mathbb{N}_0$ such that $E\left(\nu\left(A_j^{-1}(K)\right)\right) > 0$, $Y = \sum_{k=0}^{\infty} A_k X_k$ is regularly varying with index $\alpha$, measure $\tilde{\nu}$ and normalizing sequence $(a_n)$.*

Again condition (3.7), resp. (3.8), is independent of the norm used. The proof below indicates that under appropriately adapted summability conditions one should be able to extend the result to independent $(X_k)$ that are regularly varying with common index and normalizing sequence, but different measures $\nu_{X_k}$, and further to the case of $\mathbb{R}^{d_k}$-valued $X_i$ and $M_{qd_k}$-valued $A_k$ with possibly different $d_i$. Since such extensions are of no relevance in our later studies, we do not pursue them.

**Proof:** We only prove (3.9). The other assertions are shown as in Theorems 3.16 and 3.18.

To prove (3.9) we employ Proposition 3.4. Let $K$ be a compact subset of $\overline{\mathbb{R}^d}\backslash\{0\}$. Let $\epsilon := dist(K,0)$ denote the distance of the set $K$ from 0 w.r.t. $\|\cdot\|$. For $0<\delta<\epsilon$ the set $K_\delta = \overline{\{x\in\mathbb{R}^d : dist(K,x)\leq\delta\}}\cup K$ is compact in $\overline{\mathbb{R}^d}\backslash\{0\}$ (the union with $K$ is taken to avoid notational ambiguities for $K\cap\overline{\mathbb{R}^d}\backslash\mathbb{R}^d\neq\emptyset$). The event $\sum_{k=0}^\infty A_kX_k\in a_nK$ is part of the event that $\sum_{k=0}^L A_kX_k\in a_nK_\delta$ or $a_n^{-1}\left\|\sum_{k=L+1}^\infty A_kX_k\right\|>\delta$ for any integer $L$. From the triangle inequality it is immediate that $a_n^{-1}\left\|\sum_{k=L+1}^\infty A_kX_k\right\|>\delta$ is in turn part of the event $\sum_{k=L+1}^\infty \|A_k\|\|X_k\|>a_n\delta$. Thus:

$$nP\left(\sum_{k=0}^\infty A_kX_k\in a_nK\right)\leq nP\left(\sum_{k=0}^L A_kX_k\in a_nK_\delta\right)+nP\left(\sum_{k=L+1}^\infty \|A_k\|\|X_k\|>a_n\delta\right)$$

Letting $n\to\infty$ we can apply Theorem 3.17 in connection with Proposition 3.4 (iii) on the first and Theorem 3.18 (with $d=1$) in connection with Proposition 3.4 (ii) to the second term:

$$\begin{aligned}\limsup_{n\to\infty} nP\left(\sum_{k=0}^\infty A_kX_k\in a_nK\right) &\leq \limsup_{n\to\infty} nP\left(\sum_{k=0}^L A_kX_k\in a_nK_\delta\right)\\ &+ \limsup_{n\to\infty} nP\left(\sum_{k=L+1}^\infty \|A_k\|\|X_k\|>a_n\delta\right)\\ &\leq \sum_{k=0}^L E\left(\nu\circ A_k^{-1}(K_\delta)\right)+\delta^{-\alpha}\nu((1,\infty])\sum_{k=L+1}^\infty E(\|A_k\|^\alpha).\end{aligned}$$

As in Theorem 3.18 it is shown that $\sum_{k=0}^\infty E(\|A_k\|^\alpha)<\infty$ and so the second term vanishes for $L\to\infty$. Since Radon measures on polish spaces are regular (cf. Bauer (1992, Def. 25.3 and Korollar 26.4), it holds that $E\left(\nu\circ A_k^{-1}(K_\delta)\right)\to E\left(\nu\circ A_k^{-1}(K)\right)$ for $\delta\to 0$. Moreover, for all $\delta<\epsilon$ it is immediate to see that $\sum_{k=0}^\infty E\left(\nu\circ A_k^{-1}(K_\delta)\right)$ is finite and monotonically increasing in $\delta$. Thus letting first $L\to\infty$ and afterwards $\delta\to 0$:

$$\limsup_{n\to\infty} nP\left(\sum_{k=0}^\infty A_kX_k\in a_nK\right)\ \leq\ \sum_{k=0}^\infty E\left(\nu\circ A_k^{-1}(K)\right).$$

Consider now any open relatively compact subset $G$ of $\overline{\mathbb{R}^d}\backslash\{0\}$. There exists a sequence $(G_m)$ of open relatively compact sets of $G$ such that $G_m\subset\overline{G_m}\subset G_{m+1}\nearrow G$ (strict inclusions). This implies $dist(G_m,G^c)>0$ and so for fixed integers $L,m$ there exists an $\epsilon>0$ which only depends upon $m$ such that, if

$$\sum_{k=0}^L A_kX_k\in a_nG_m \text{ and } \left\|\sum_{k=L+1}^\infty A_kX_k\right\|\leq a_n\epsilon,$$

then

$$\sum_{k=0}^\infty A_kX_k\in a_nG$$

holds.

Therefore

$$\begin{aligned} nP\left(\sum_{k=0}^{\infty} A_k X_k \in a_n G\right) &\geq nP\left(\sum_{k=0}^{L} A_k X_k \in a_n G_m, \left\|\sum_{k=L+1}^{\infty} A_k X_k\right\| \leq a_n \epsilon\right) \\ &\geq nP\left(\sum_{k=0}^{L} A_k X_k \in a_n G_m\right) - nP\left(\left\|\sum_{k=L+1}^{\infty} A_k X_k\right\| > a_n \epsilon\right) \\ &\geq nP\left(\sum_{k=0}^{L} A_k X_k \in a_n G_m\right) - nP\left(\sum_{k=L+1}^{\infty} \|A_k\| \|X_k\| > a_n \epsilon\right), \end{aligned}$$

where we used $P(A \cap B) = P(A \backslash B^c) \geq P(A) - P(B^c)$ in the second and the triangle inequality in the last step. Letting $n \to \infty$ we can apply Theorem 3.17 in connection with Proposition 3.4 (iii) on the first and Theorem 3.18 in connection with Proposition 3.4 (ii) to the second term and obtain:

$$\begin{aligned} \liminf_{n\to\infty} nP\left(\sum_{k=0}^{\infty} A_k X_k \in a_n G\right) &\geq \liminf_{n\to\infty} nP\left(\sum_{k=0}^{L} A_k X_k \in a_n G_m\right) \\ &- \liminf_{n\to\infty} nP\left(\sum_{k=L+1}^{\infty} \|A_k\| \|X_k\| > a_n \epsilon\right) \\ &\geq \sum_{k=0}^{L} E\left(\nu \circ A_k^{-1}(G_m)\right) - \epsilon^{-\alpha} \nu((1,\infty]) \sum_{k=L+1}^{\infty} E(\|A_k\|^\alpha). \end{aligned}$$

For $L \to \infty$ the second term vanishes as before. Since $\tilde{\nu}$ is a Radon measure, one obtains that $\sum_{k=0}^{\infty} E\left(\nu \circ A_k^{-1}(G)\right)$ is finite and due to the regularity $E\left(\nu \circ A_k^{-1}(G_m)\right) \to E\left(\nu \circ A_k^{-1}(G)\right)$ as $m \to \infty$ monotonically. A bounded convergence argument thus establishes $\sum_{k=0}^{\infty} E\left(\nu \circ A_k^{-1}(G_m)\right) \to \sum_{k=0}^{\infty} E\left(\nu \circ A_k^{-1}(G)\right)$ as $m \to \infty$. Thus letting first $L \to \infty$ and afterwards $m \to \infty$:

$$\liminf_{n\to\infty} nP\left(\sum_{k=0}^{\infty} A_k X_k \in a_n G\right) \geq \sum_{k=0}^{\infty} E\left(\nu \circ A_k^{-1}(G)\right).$$

Combining the above results for compact and open relatively compact sets gives (3.9) using Proposition 3.4 (iii). □

The criteria involving the finiteness of series in the foregoing two theorems are more or less motivated by the lines of proof, they are minimally necessary for the used arguments to work. However, one will in general prefer more straightforward sufficient criteria. A first step is the following.

**Lemma 3.20** *Assume that $(A_k)_{k\in\mathbb{N}_0}$ is a sequence of $M_{qd}(\mathbb{R})$-valued random variables and that there is an $\beta > 0$ such that $A_k \in L^\beta$ for all $k \in \mathbb{N}_0$. If, moreover,*

$$\limsup_{k\to\infty} E\left(\|A_k\|^\beta\right)^{1/k} < 1, \tag{3.10}$$

*then it holds for all* $0 < \gamma \le \beta$ *and* $\tau > 0$ *that*

$$\sum_{k=0}^{\infty} E\left(\|A_k\|^\gamma\right) < \infty$$

*and*

$$\sum_{k=0}^{\infty} E\left(\|A_k\|^\gamma\right)^{1/\tau} < \infty.$$

Note that condition (3.10) is independent of the employed norm. Similar conditions are used heavily throughout the thesis.

**Proof:** Jensen's inequality applied to the concave map $x \mapsto x^{\frac{\gamma}{\beta}}$ gives $E(\|A_k\|^\gamma) \le E(\|A_k\|^\beta)^{\frac{\gamma}{\beta}}$. Noting that for any positive sequence $(c_k)$ and arbitrary $r > 0$ we have that

$$\limsup_{k\to\infty} c_k < 1 \;\Leftrightarrow\; \limsup_{k\to\infty} c_k^r < 1,$$

all claims immediately follow from the root criterion of standard analysis. $\square$

**Lemma 3.21** *Consider the basic set-up of Theorem 3.19. If, moreover, for some* $\beta > \alpha$

$$\limsup_{k\to\infty} E\left(\|A_k\|^\beta\right)^{1/k} < 1, \tag{3.11}$$

*then the condition (3.7), resp. (3.8), is satisfied for all admissible* $\eta$ *with* $\eta \le \beta - \alpha$.

# Chapter 4

# A First Order Stochastic Difference Equation

In this section we first give a general theorem on the existence of a unique stationary solution to a general stochastic difference equation of the form $Y_n = A_n Y_{n-1} + C_n$ due to Brandt (1986) (cf. also Brandt, Franken and Lisek (1990, Section 9.1)). The version we give below is a straightforward extension of Brandt's result and has first been noted in Bougerol and Picard (1992b). Later on the existence of moments is studied.

## 4.1 Stationary Solutions to $Y_n = A_n Y_{n-1} + C_n$

Our aim is to establish a general theorem for a stochastic difference equation of the type $Y_n = A_n Y_{n-1} + C_n$ to have a stationary and ergodic solution:

**Theorem 4.1 (cf. Brandt (1986, Th. 1), Bougerol and Picard (1992b, Th. 1.1))** *Let $(A_n, C_n) \in M_d(\mathbb{R}) \times \mathbb{R}^d$, $n \in \mathbb{Z}$, be a stationary ergodic process with finite $E(\log^+ \|A_0\|)$ and $E(\log^+ \|C_0\|)$ (where $\log^+(x) := \max(0, \log(x))$). Assume furthermore that*

$$\gamma := \inf_{n \in \mathbb{N}_0} \left( \frac{1}{n+1} E\left(\log \|A_0 A_{-1} \cdots A_{-n}\|\right) \right) < 0. \tag{4.1}$$

*Then the stochastic process $X = (X_n)_{n \in \mathbb{Z}}$ defined by*

$$X_n = \sum_{k=0}^{\infty} A_n A_{n-1} \cdots A_{n-k+1} C_{n-k} = C_n + \sum_{k=1}^{\infty} A_n A_{n-1} \cdots A_{n-k+1} C_{n-k} \tag{4.2}$$

*is the unique stationary solution of the stochastic difference equation $Y_n = A_n Y_{n-1} + C_n$. Moreover, $X$ is ergodic and the series in equation (4.2) converges almost surely absolutely.*

*Let $V_0$ be an arbitrary $\mathbb{R}^d$-valued random variable defined on the same probability space as $(A_n, C_n)_{n \in \mathbb{Z}}$ and define $(V_n)_{n \in \mathbb{N}}$ recursively via $V_n = A_n V_{n-1} + C_n$ then*

$$\|X_n - V_n\| \xrightarrow{\text{a.s.}} 0 \text{ with } n \to \infty \tag{4.3}$$

*and, in particular,*

$$V_n \xrightarrow{\mathscr{D}} X_0 \text{ with } n \to \infty, \tag{4.4}$$

*i.e. the distribution of $V_n$ converges to the stationary distribution of $X_n$.*

$\|\cdot\|$ denotes any norm on $\mathbb{R}^d$ and the induced operator norm on $M_d(\mathbb{R})$. Furthermore, $\gamma$ as defined in (4.1) is usually called *(top) Lyapunov coefficient* or *exponent* (in the special case $d=1$ we have $\gamma = \inf_{n\in\mathbb{N}_0}((1/(n+1))\sum_{k=0}^n E(\log|A_{-k}|)) = E(\log|A_0|))$. Generally a simple condition ensuring $\gamma < 0$ is $E(\log\|A_0\|) < 0$. Unfortunately, this condition is too restrictive to be of use in most interesting cases with $d$ being greater than one. To be able to prove the above result we need the following lemma on the behaviour of ergodic sequences of random matrices, which is Theorem 1 of Furstenberg and Kesten (1960) slightly adapted to our set-up. An inspection of the proof in Furstenberg and Kesten (1960) immediately gives (4.5) and shows that their result holds for any algebra norm and not only for the column-sum norm they employ.

**Lemma 4.2 (cf. Furstenberg and Kesten (1960, Th. 1))** *Let $A = (A_n)_{n\in\mathbb{Z}}$ be an $M_d(\mathbb{R})$-valued stationary random sequence and let $\|\cdot\|$ denote any algebra norm on $M_d(\mathbb{R})$. Then*

$$\lim_{n\to\infty}\frac{1}{n+1}E(\log\|A_0A_{-1}\cdots A_{-n}\|) = \inf_{n\in\mathbb{N}_0}\left(\frac{1}{n+1}E(\log\|A_0A_{-1}\cdots A_{-n}\|)\right) =: \gamma \quad (4.5)$$

*and, provided $A$ is also ergodic and $E(\log^+\|A_0\|) < \infty$,*

$$\limsup_{n\to\infty}\frac{1}{n+1}\log\|A_m\cdots A_{m-n}\| \le \gamma \ \text{ a.s. } \forall m\in\mathbb{Z} \quad (4.6)$$

*and*

$$\limsup_{n\to\infty}\frac{1}{n+1}\log\|A_{m+n}\cdots A_m\| \le \gamma \ \text{ a.s. } \forall m\in\mathbb{Z}. \quad (4.7)$$

*Moreover, $\gamma$ is independent of the algebra norm used and $E(\log^+\|A_0\|)$ is finite for all norms, if it is finite for only one.*

We prove below only the very last claim. The proof of the other results proceeds totally along the lines of the proof of Theorem 1 in Furstenberg and Kesten (1960) and is thus omitted. The first assertion (4.5) is a consequence of the submultiplicativity of $\|\cdot\|$ and a lemma on subadditive sequences to be found e.g. in Hille and Phillips (1957):

**Lemma 4.3 (Hille and Phillips (1957, Lemma 4.7.1))** *Let $(a_n)_{n\in\mathbb{N}}$ be a subadditive sequence of real numbers, i.e. $a_{n+m}\le a_n + a_m$ for all $n,m\in\mathbb{N}$, then $\lim_{n\to\infty}\frac{a_n}{n} = \inf_{n\in\mathbb{N}}\frac{a_n}{n}$.*

The second assertion (4.6) then follows from the first one by using the properties of ergodic sequences given in Section 2.5.

Actually, it can be shown that in (4.6) and (4.7) "lim sup" can be replaced by "lim" and that equality holds (see Furstenberg and Kesten (1960, Th. 2)), but the proof of this result is rather involved and for our purposes (4.6) suffices. Note, however, that (4.6) with equality is also an immediate consequence of the results on general subadditive processes reported in Kingman (1968) and Kingman (1973).

**Proof of Th. 4.2:** As said before, only the fact that $\gamma$ is independent of the algebra norm and that finiteness of $E(\log^+\|A_0\|)$ for one norm implies finiteness for all norms are to be shown.

Let $\|\cdot\|$ and $\|\cdot\|_*$ be two norms and $E(\log^+ \|A_0\|) < \infty$. There exists a finite $M > 0$ with $\|\cdot\|_* \leq M\|\cdot\|$. Thus $E(\max(0, \log \|A_0\|_*)) \leq E(\max(0, \log \|A_0\| + \log M)) \leq E(\max(0, \log \|A_0\|)) + |\log M| < \infty$.

Assume now that $\|\cdot\|$ and $\|\cdot\|_*$ are algebra norms. There exists also a $0 < m \leq M$ such that $\|\cdot\|_* \geq m\|\cdot\|$. This gives for all natural $n$

$$\frac{1}{n+1}\log\|A_0\cdots A_{-n}\| + \frac{\log m}{n+1} \leq \frac{1}{n+1}\log\|A_0\cdots A_{-n}\|_* \leq \frac{1}{n+1}\log\|A_0\cdots A_{-n}\| + \frac{\log M}{n+1}.$$

Hence,

$$\lim_{n\to\infty}\frac{1}{n+1}E(\log\|A_0\cdots A_{-n}\|) = \lim_{n\to\infty}\frac{1}{n+1}E(\log\|A_0\cdots A_{-n}\|_*)$$

and so $\gamma$ is independent of the particular algebra norm used. □

**Proof of Theorem 4.1: (adapted from Brandt (1986) and Bougerol and Picard (1992b))**

*Step 1:* Convergence of the series in (4.2).

$E(\log^+ \|C_0\|) < \infty$ implies via

$$E(\log^+\|C_0\|) = \int_0^\infty P(\log^+\|C_0\| > x)dx = -\frac{\gamma}{2}\int_0^\infty P(\log^+\|C_0\| > -\frac{\gamma}{2}s)ds$$

the stationarity of $(C_n)$ and the integral comparison criterion from standard analysis that

$$\sum_{k=0}^{\infty} P\left(\log^+\|C_{n-k}\| > -\frac{k\gamma}{2}\right) < \infty.$$

Since

$$\limsup_{k\to\infty}\left\{\omega\in\Omega : \frac{\log^+\|C_{n-k}(\omega)\|}{k} > \frac{-\gamma}{2}\right\} \supseteq \left\{\omega\in\Omega : \limsup_{k\to\infty}\frac{\log^+\|C_{n-k}(\omega)\|}{k} > \frac{-\gamma}{2}\right\},$$

the Borel-Cantelli lemma implies that

$$\limsup_{k\to\infty}\frac{1}{k}\log^+\|C_{n-k}\| \leq -\frac{\gamma}{2} \text{ a.s.}$$

Together with Lemma 4.2 and the stationarity of $(A_n)_{n\in\mathbb{Z}}$ we thus obtain

$$\limsup_{k\to\infty}\frac{1}{k}\log\|A_n\cdots A_{n-k+1}C_{n-k}\| \leq \limsup_{k\to\infty}\frac{1}{k}\left(\log\|A_n\cdots A_{n-k+1}\| + \log\|C_{n-k}\|\right) \leq \frac{\gamma}{2}$$

a.s. This yields

$$\limsup_{k\to\infty}\|A_n\cdots A_{n-k+1}C_{n-k}\|^{1/k} \leq \exp\left(\frac{\gamma}{2}\right) < 1 \text{ a.s.}$$

and so by the root criterion from standard analysis the series $\sum_{k=0}^{\infty} A_n A_{n-1}\cdots A_{n-k+1}C_{n-k}$ converges almost surely absolutely.

*Step 2:* $X$ is a solution of $Y_n = A_n Y_{n-1} + C_n$.

$$\begin{aligned} A_n X_{n-1} + C_n &= A_n \left( C_{n-1} + \sum_{k=1}^{\infty} A_{n-1} \cdots A_{n-k} C_{n-k-1} \right) + C_n \\ &= C_n + A_n C_{n-1} + \sum_{k=2}^{\infty} A_n \cdots A_{n-k+1} C_{n-k} = X_n \end{aligned}$$

*Step 3:* Stationarity and Ergodicity of $X$
On the range $D \subset (M_d(\mathbb{R}) \times \mathbb{R}^d)^{\mathbb{Z}}$ of $(A_n, C_n)_{n\in\mathbb{Z}}$ define for $n \in \mathbb{Z}$ the functions $g_n$ as $g_n((a_i, c_i)_{i\in\mathbb{Z}}) = \sum_{k=0}^{\infty} a_n \cdots a_{n-k+1} c_{n-k}$. The sequence $(g_n(A_i, C_i)_{i\in\mathbb{Z}})_{n\in\mathbb{Z}}$ is according to step 1 a.s. a sequence with all elements being finite, since $\mathbb{Z}$ is countable, and we have $X = (g_n((A_i, C_i)_{i\in\mathbb{Z}}))_{n\in\mathbb{Z}}$. Moreover, each $g_n$ is obviously a limit of measurable functions and thus measurable. A trivial calculation shows that $g_{n-1} = g_n \circ B$ with $B$ being the back-shift operator on $(M_d(\mathbb{R}) \times \mathbb{R}^d)^{\mathbb{Z}}$. Now Lemma 2.25 shows that $X$ is stationary and ergodic.
*Step 4:* Uniqueness of the solution.
Let $X = (X_n)$ and $Z = (Z_n)$ be two solutions of $Y_n = A_n Y_{n-1} + C_n$. For $k \in \mathbb{N}$ we have

$$\|X_n - Z_n\| = \|A_n \cdots A_{n-k}(X_{n-k-1} - Z_{n-k-1})\| \le \|A_n \cdots A_{n-k}\| \cdot (\|X_{n-k-1}\| + \|Y_{n-k-1}\|)$$

and from $0 > \gamma \ge \limsup_{k\to\infty} \log(\|A_n \cdots A_{n-k}\|)/k$ we get

$$\|A_n \cdots A_{n-k}\| = \exp\left(\log(\|A_n \cdots A_{n-k}\|)/k\right)^k \overset{k\to\infty}{\longrightarrow} 0 \text{ a.s.}$$

This gives that $\|A_n \cdots A_{n-k}\| \cdot (\|X_{n-k-1}\| + \|Y_{n-k-1}\|)$ converges to zero in probability (use e.g. Brockwell and Davis (1991, Prop. 6.1.1 (ii))) and thus a.s. along a subsequence. Hence, $\|X_n - Z_n\|$ is necessarily a.s. zero, so $X \overset{a.s.}{=} Z$.
*Step 5:* Convergence to the stationary solution for arbitrary starting values $V_0$.

The recursive definition of $(V_n)_{n\in\mathbb{N}}$ immediately gives for all $n \in \mathbb{N}$:

$$V_n = C_n + \sum_{k=1}^{n-1} A_n A_{n-1} \cdots A_{n-k+1} C_{n-k} + A_n \cdots A_1 V_0. \tag{4.8}$$

Thus

$$\begin{aligned} \|X_n - V_n\| &= \left\| \sum_{k=n}^{\infty} A_n \cdots A_{n-k+1} C_{n-k} - A_n \cdots A_1 V_0 \right\| \\ &= \left\| A_n \cdots A_1 \left( C_0 + \sum_{k=n+1}^{\infty} A_0 \cdots A_{n-k+1} C_{n-k} - V_0 \right) \right\| \\ &= \|A_n \cdots A_1 (X_0 - V_0)\| \\ &\le \|A_n \cdots A_1\| (\|X_0 + Y_0\|) \overset{\text{a.s.}}{\to} 0 \text{ with } n \to \infty, \end{aligned}$$

since $\|A_n \cdots A_1\| = \exp\left(\log(\|A_n \cdots A_1\|)/n\right)^n \overset{n\to\infty}{\longrightarrow} 0$ a.s. as argued above. (4.4) now is implied by the standard theory on convergence in distribution (see e.g. Brockwell and

Davis (1991, Prop. 6.3.3)), since $X_n \stackrel{\mathscr{D}}{=} X_0$ for all $n \in \mathbb{Z}$. □

The following result on the Lyapunov exponent is sometimes useful to show that strict negativity of $\gamma$ is a necessary condition for some particular stochastic recurrence equation to have a solution.

**Lemma 4.4 (Bougerol and Picard (1992b, Lemma 3.4))** *Let $(A_n)_{n\in\mathbb{Z}}$ be an ergodic and stationary sequence of random matrices in $M_d(\mathbb{R})$. If $E(\log^+\|A_0\|)$ is finite and $\lim_{n\to\infty}\|A_0A_{-1}\cdots A_{-n}\| = 0$ a.s., then the top Lyapunov exponent*

$$\gamma = \inf_{n\in\mathbb{N}_0}\left(\frac{1}{n+1}E\left(\log\|A_0A_{-1}\cdots A_{-n}\|\right)\right)$$

*is strictly negative.*

## 4.2 Existence of Moments

In the following we give conditions ensuring that the solution of $Y_n = A_nY_{n-1} + C_n$ with stationary and ergodic$(A_n, C_n)$ is in $L^p$ for some $p > 0$. For a stochastic difference equation with $(A_n, C_n)$ being a sequence in $\mathbb{R}^2$ some results are already to be found in Karlsen (1990b). Saporta (2004a) briefly mentions conditions for the existence of moments, if $(A_n)$ and $(C_n)$ are assumed to be independent, and the one dimensional case is further shortly studied in Saporta (2004b). First we note that instead of considering the limit behaviour of the logarithm of the norm of a product of random matrices (i.e. the Lyapunov coefficient) one can look at any positive power.

**Lemma 4.5** *Let $A = (A_t)_{t\in\mathbb{Z}}$ be an $M_d(\mathbb{R})$ valued stationary random sequence and $\|\cdot\|$ be any algebra norm over $M_d(\mathbb{R})$. If for some $s > 0$*

$$\limsup_{n\to\infty} E(\|A_0\cdots A_{-n}\|^s)^{1/(n+1)} =: \gamma_s < 1$$

*or*

$$\limsup_{n\to\infty} \|A_0\cdots A_{-n}\|_{L^\infty}^{1/(n+1)} =: \gamma_\infty < 1,$$

*then*

$$\lim_{n\to\infty}\frac{1}{n+1}E(\log\|A_0\cdots A_{-n}\|) < 0.$$

*All limits above are independent of the algebra norm used.*

Obviously, one has $\limsup_{n\to\infty} E(\|A_0\cdots A_{-n}\|^s)^{1/(n+1)} < 1$, if and only if $\limsup_{n\to\infty}\|A_0 \cdots A_{-n}\|_{L^s}^{1/(n+1)} < 1$, in the case $1 \le s < \infty$.

**Proof:** The independence from the algebra norm employed is again an immediate consequence of the equivalence of any two norms and the fact that we have $\lim_{n\to\infty} m^{1/(n+1)} = 1$ for any $m > 0$.

Assume now that the $L^\infty$-condition above holds. Then for some $\epsilon$ such that $\gamma_\infty < \epsilon < 1$ there is an $N \in \mathbb{N}$ such that $\|A_0\cdots A_{-n}\|^{1/(n+1)} < \epsilon < 1$ a.s. $\forall n \ge N$. Taking the logarithm, the expectation and the limit for $n \to \infty$ (which exists by Lemma 4.2) concludes the proof.

If the asymptotic moment condition holds for some $s > 0$, we get from the concavity of the logarithm, Jensen's inequality and again Lemma 4.2

$$\begin{aligned} 0 &> \frac{1}{s} \limsup_{n\to\infty} \log\left(E\left(\|A_0 \cdots A_{-n}\|^s\right)^{1/(n+1)}\right) \\ &\geq \limsup_{n\to\infty} \frac{1}{s(n+1)} E\left(\log \|A_0 \cdots A_{-n}\|^s\right) = \lim_{n\to\infty} \frac{1}{n+1} E(\log \|A_0 \cdots A_{-n}\|). \end{aligned}$$

□

In some cases it is possible to replace the limes superior by a limit in the Lemma above.

**Lemma 4.6** *Let $A = (A_t)_{t\in\mathbb{Z}}$ be an $M_d(\mathbb{R})$ valued stationary random sequence and $\|\cdot\|$ be any algebra norm over $M_d(\mathbb{R})$.*

*(i) If $A_0 \in L^\infty$,*

$$\gamma_\infty = \lim_{n\to\infty} \|A_0 \cdots A_{-n+1}\|_{L^\infty}^{1/n} = \inf_{n\in\mathbb{N}} \|A_0 \cdots A_{-n+1}\|^{1/n}.$$

*(ii) If $(A_t)$ is an i.i.d. sequence and $A_0 \in L^s$ for some $s \in \mathbb{R}^+$, it holds that*

$$\gamma_s = \lim_{n\to\infty} E(\|A_0 \cdots A_{-n+1}\|^s)^{1/n} = \inf_{n\in\mathbb{N}} E(\|A_0 \cdots A_{-n+1}\|^s)^{1/n}.$$

Note that for notational ease we often writelim sup generally in the following, even if the above Lemma is applicable.
**Proof:** In the case (i) we have

$$\begin{aligned} \log \|A_0 \cdots A_{-(n+m)+1}\|_{L^\infty} &\leq \log \|A_0 \cdots A_{-n+1}\|_{L^\infty} + \log \|A_{-n} \cdots A_{-n-m+1}\|_{L^\infty} \\ &= \log \|A_0 \cdots A_{-n+1}\|_{L^\infty} + \log \|A_0 \cdots A_{-m+1}\|_{L^\infty} \end{aligned}$$

and this gives via Lemma 4.3

$$\lim_{n\to\infty} \frac{1}{n} \log \|A_0 \cdots A_{-n+1}\|_{L^\infty} = \lim_{n\to\infty} \log \|A_0 \cdots A_{-n+1}\|_{L^\infty}^{1/n} = \inf_{n\in\mathbb{N}} \log \|A_0 \cdots A_{-n+1}\|_{L^\infty}^{1/n}.$$

The continuity and strict monotonicity of the exponential function now shows (i). For (ii) one observes

$$\begin{aligned} \log E(\|A_0 \cdots A_{-(n+m)+1}\|^s) &\leq \log E(\|A_0 \cdots A_{-n+1}\|^s) + \log E(\|A_{-n} \cdots A_{-n-m+1}\|^s) \\ &= \log E(\|A_0 \cdots A_{-n+1}\|^s) + \log E(\|A_0 \cdots A_{-m+1}\|^s) \end{aligned}$$

and proceeds analogously. □
The next Theorem gives an extension of Theorem 4.1 by considering moments of the solution of the difference equation and is a straightforward extension of Karlsen (1990b, Th. 3.1) to the multidimensional set-up.

**Theorem 4.7** *Let* $(A_n, C_n) \in M_d(\mathbb{R}) \times \mathbb{R}^d$, $n \in \mathbb{Z}$, *be a stationary ergodic process with finite* $E(\log^+ \|A_0\|)$ *and* $E(\log^+ \|C_0\|)$ *(where* $\log^+(x) := \max(0, \log(x))$*). Assume furthermore that*

$$\gamma := \inf_{n\in\mathbb{N}_0} \left( \frac{1}{n+1} E\left(\log \|A_0 A_{-1} \cdots A_{-n}\|\right) \right) < 0, \tag{4.9}$$

*that* $A_0 \cdots A_{-k+1} C_{-k} \in L^p$ *holds for all* $k \in \mathbb{N}_0$ *and some* $p > 0$ *and that for* $1 \le p \le \infty$

$$\sum_{k=0}^{\infty} \|A_0 \cdots A_{-k+1} C_{-k}\|_{L^p}, \tag{4.10}$$

*respectively for* $0 < p < 1$

$$\sum_{k=0}^{\infty} E(\|A_0 \cdots A_{-k+1} C_{-k}\|^p) = \sum_{k=0}^{\infty} d_{L^p}(A_0 \cdots A_{-k+1} C_{-k}, 0), \tag{4.11}$$

*converges. Then the stochastic process* $X = (X_n)_{n\in\mathbb{Z}}$ *defined by*

$$X_n = \sum_{k=0}^{\infty} A_n A_{n-1} \cdots A_{n-k+1} C_{n-k} = C_n + \sum_{k=1}^{\infty} A_n A_{n-1} \cdots A_{n-k+1} C_{n-k} \tag{4.12}$$

*is the unique stationary solution of the stochastic difference equation* $Y_n = A_n Y_{n-1} + C_n$. *Moreover,* $X$ *is ergodic and in* $L^p$ *and the series in equation (4.12) converges almost surely absolutely and in* $L^p$.

It is clear that, if (4.10) holds for some $p \ge 1$, it holds for all $r \in [1, p]$ as well. However, if (4.11) holds for some $p \in (0, 1)$ this does not imply that it holds for all smaller values of $p$ as well. Yet, of course, if $X \in L^p$ for some $p > 0$, then $X \in L^r$ for all $r \in (0, p]$. Note, however, that the asymptotic conditions given in the next lemmata are much better behaved. If there is one $p \in (0, \infty]$ that fulfils the asymptotic condition, then the asymptotic conditions for all $s \in (0, p]$ are satisfied as well (use Jensen's inequality as in the proof of Lemma 3.20).

**Proof:** Only the convergence of the series in $L^p$ needs to be shown, since this implies that $(X_t)$ is a sequence of random variables in $L^p$. Furthermore, due to the stationarity only $X_0$ needs to be considered. If $1 \le p \le \infty$, the assumed absolute convergence (in $L^p$) (4.10) of the series immediately gives the convergence of the series defining $X_0$ in $L^p$, since $L^p$ is a Banach space for such $p$.

Consider now the case $0 < p < 1$. The following properties of $d_{L^p}$ are obvious from the definition:

$$\begin{aligned} d_{L^p}(X, Y) &= d_{L^p}(X - Y, 0)\, \forall\, X, Y \in L^p, & (4.13)\\ d_{L^p}(X + Y, 0) &= d_{L^p}(X, -Y) \le d_{L^p}(X, 0) + d_{L^p}(Y, 0)\, \forall\, X, Y \in L^p. & (4.14) \end{aligned}$$

So one obtains for $m, n \in \mathbb{N}$, $m > n$,

$$\begin{aligned} d_{L^p}\left( \sum_{k=0}^{m} A_0 \cdots A_{-k+1} C_{-k}, \sum_{k=0}^{n} A_0 \cdots A_{-k+1} C_{-k} \right) &= d_{L^p}\left( \sum_{k=n+1}^{m} A_0 \cdots A_{-k+1} C_{-k}, 0 \right) \\ &\le \sum_{k=n+1}^{m} d_{L^p}\left(A_0 \cdots A_{-k+1} C_{-k}, 0\right). \end{aligned}$$

Thus the assumptions imply that $\left(\sum_{k=0}^{m} A_0 \cdots A_{-k+1}C_{-k}\right)_{m\in\mathbb{N}}$ is a Cauchy sequence in $L^p$ and hence convergent, since by Theorem 2.12 $L^p$ is complete. □

The next theorems are concerned with conditions ensuring (4.10), resp. (4.11). A simple consequence of the root criterion of standard analysis is:

**Lemma 4.8** *Let $1 \le p \le \infty$, resp. $0 < p < 1$, and assume that*

$$\limsup_{k\to\infty} \|A_0 \cdots A_{-k+1}C_{-k}\|_{L^p}^{1/k} < 1,$$

*resp.*

$$\limsup_{k\to\infty} \left(E(\|A_0 \cdots A_{-k+1}C_{-k}\|^p)\right)^{1/k} < 1,$$

*holds. Then (4.10), resp. (4.11), is fulfilled.*

Note that for finite $p \ge 1$ one has

$$\limsup_{k\to\infty} \|A_0 \cdots A_{-k+1}C_{-k}\|_{L^p}^{1/k} < 1 \Leftrightarrow \limsup_{k\to\infty} \left(E(\|A_0 \cdots A_{-k+1}C_{-k}\|^p)\right)^{1/k} < 1.$$

Using the Hölder inequality one can give a criterion to check (4.10), resp. (4.11), that also ensures the negativity of the Lyapunov coefficient. For the one-dimensional case a similar result is Karlsen (1990b, Corollary 3.1).

**Proposition 4.9** *Let $p \in (0,\infty)$. If there exist $r, s \ge 1$ with $1/r + 1/s = 1$, such that $A_0 \cdots A_{-k+1} \in L^{pr}$ $\forall k \in \mathbb{N}$, $\limsup_{k\to\infty} E(\|A_0 \cdots A_{-k+1}\|^{pr})^{1/k} < 1$ for $0 < pr < \infty$, resp. $\lim_{k\to\infty} \|A_0 \cdots A_{-k+1}\|_{L^\infty}^{1/k} < 1$ for $pr = \infty$, and $C_0 \in L^{ps}$, then (4.9) and (4.10) for $p \ge 1$, resp. (4.11) for $0 < p < 1$, hold.*

**Proof:** (4.9) is implied by Lemma 4.5. The Hölder inequality gives

$$E(\|A_0 \cdots A_{-k+1}C_{-k}\|^p) \le \|A_0 \cdots A_{-k+1}\|_{L^{pr}}^p \|C_{-k}\|_{L^{ps}}^p,$$

which implies

$$\limsup_{k\to\infty} E(\|A_0 \cdots A_{-k+1}C_{-k}\|^p)^{1/k} \le \limsup_{k\to\infty} \left(\|A_0 \cdots A_{-k+1}\|_{L^{pr}}^{p/k} \|C_{-k}\|_{L^{ps}}^{p/k}\right) < 1.$$

Thus the assumptions of Lemma 4.8 are fulfilled and this gives that (4.10), resp. (4.11), holds. □

An important consequence is that, provided $A_0 \cdots A_{-k+1} \in L^\infty$ for all natural $k$, which especially is the case, if the stationary distribution of $A_n$ has bounded or finite support, and $\lim_{k\to\infty} \|A_0 \cdots A_{-k+1}\|_{L^\infty}^{1/k} < 1$, the solution $(X_n)$ to the stochastic difference equation has a $p$-th moment, if the noise sequence $(C_n)$ has one.

It remains to give a criterion for $X_n$ to be in $L^\infty$:

**Proposition 4.10** *If $A_0 \in L^\infty$, $\lim_{k\to\infty} \|A_0 \cdots A_{-k+1}\|_{L^\infty}^{1/k} < 1$ and $C_0 \in L^\infty$, then (4.9) and (4.10) with $p = \infty$ hold.*

**Proof:** $A_0 \in L^\infty$ gives $A_0 \cdots A_{-k+1} \in L^\infty$ for all $k \in \mathbb{N}$ and (4.9) is implied by Lemma 4.5. Furthermore, we have

$$\limsup_{k\to\infty} \|A_0 \cdots A_{-k+1} C_{-k}\|_{L^\infty}^{1/k} \le \limsup_{k\to\infty} \left( \|A_0 \cdots A_{-k+1}\|_{L^\infty}^{1/k} \|C_{-k}\|_{L^\infty}^{1/k} \right) < 1.$$

Thus the assumptions of Lemma 4.8 are fulfilled with $p = \infty$ and this gives that (4.10) holds. □

It is straightforward to obtain a very simple condition ensuring

$$\lim_{k\to\infty} \|A_0 \cdots A_{-k+1}\|_{L^\infty}^{1/k} < 1.$$

**Lemma 4.11** *If there is a $c < 1$ such that $\|A_0\| < c$ a.s., then $\lim_{k\to\infty} \|A_0 \cdots A_{-k+1}\|_{L^\infty}^{1/k} < 1$ holds.*

The proof is trivial and unfortunately the above lemma is too restrictive to be (at least straightforwardly) applicable in most interesting cases. Note that $\|A_0\| < c$ a.s. with $c < 1$ is equivalent to $\|A_0\|_{L^\infty} < 1$. Furthermore, it may at a first glance look strange to demand $\|A_0\| < c$ a.s. for some $c < 1$ instead of $\|A_0\| < 1$ a.s.. But in the later case it is possible to have $\|A_0\|_{L^\infty} = 1$ and even $\|A_0 \cdots A_{-k+1}\|_{L^\infty}^{1/k} = 1$ for all natural $k$ (a simple example is that $(A_t)$ are i.i.d. random variables drawn from the uniform distribution on $(0; 1)$).

# 4.3 Regularly Varying Tails

The aim of this section is to study the tail behaviour of the stationary solution of the stochastic difference equation $Y_n = A_n Y_{n-1} + C_n$ with stationary and ergodic input $(A_n, C_n) \in M_d(\mathbb{R}) \times \mathbb{R}^d$. We restrict ourselves to two important and essentially different cases when regularly varying tail behaviour shows up. In both cases we cannot work with a general stationary and ergodic input $(A_n, C_n)$ but need rather heavy independence assumptions.

## 4.3.1 Regularly Varying Noise

For the remainder of this section we assume $(C_n)_{n\in\mathbb{Z}}$ to be an i.i.d. regularly varying noise sequence and further that the sequences $(A_n)_{n\in\mathbb{Z}}$ and $(C_n)_{n\in\mathbb{Z}}$ are independent of each other. However, no restrictions whatsoever are imposed upon the dependence structure of $(A_n)$. Our results obtained in the following are a generalization of Resnick and Willekens (1991, Sec. 3), who demand that also $(A_n)$ be i.i.d. They are also similar to those of Grey (1994) and Konstantinides and Mikosch (2004), where the one dimensional case with the joint sequence $(A_n, C_n)_{n\in\mathbb{Z}}$ being assumed to be i.i.d., but possible dependence between $A_n$ and $C_n$ is studied. Another related paper is Davis and Resnick (1996) who studied bilinear processes.

Using Theorem 4.1 it follows rather immediately from Theorem 3.19 that under appropriate technical conditions the stationary distribution of the solution $X_0$ to the stochastic difference equation is regularly varying with the same index and normalizing sequence as the noise $C_0$. For $d = 1$ this means, in particular, that the distributions of $X_0$ and $\epsilon_0$ are tail equivalent and, provided the upper tail is nondegenerate, both belong to the maximum domain of attraction of the Fréchet distribution $\Phi_\alpha$, where $\alpha$ is the common index of regular variation (see Leadbetter, Lindgren and Rootzén (1983), Resnick (1987, Sec. 1.2, 1.5) or Embrechts, Klüppelberg and Mikosch (1997, Sec. 3.3.1)). Note that at least one of the two tails is always non-degenerately regularly varying.

**Theorem 4.12** *Let $A_n \in M_d(\mathbb{R})$, $n \in \mathbb{Z}$, be a stationary ergodic process and $C_n$ a sequence of i.i.d. regularly varying $\mathbb{R}^d$-valued random variables with index $\alpha > 0$, measure $\nu$ and normalizing sequence $(a_n)$ such that (iv) in Theorem 3.9 holds. Assume that the sequences $(A_n)_{n\in\mathbb{Z}}$ and $(C_n)_{n\in\mathbb{Z}}$ are independent of one another and, furthermore, that*

$$\gamma := \inf_{n\in\mathbb{N}_0} \left( \frac{1}{n+1} E\left(\log \|A_0 A_{-1} \cdots A_{-n}\|\right) \right) < 0. \tag{4.15}$$

*If $\alpha < 1$, assume there is an $0 < \eta < \alpha$ with $\alpha + \eta < 1$ such that $A_0 \cdots A_{-k+1} \in L^{\alpha+\eta}$ for all $k \in \mathbb{N}_0$ and that*

$$\sum_{k=0}^{\infty} E\left(\|A_0 \cdots A_{-k+1}\|^{\alpha+\eta}\right) < \infty, \ \sum_{k=0}^{\infty} E\left(\|A_0 \cdots A_{-k+1}\|^{\alpha-\eta}\right) < \infty. \tag{4.16}$$

*If $\alpha \geq 1$, assume there is $0 < \eta < \alpha$ such that $A_0 \cdots A_{-k+1} \in L^{\alpha+\eta}$ for all $k \in \mathbb{N}_0$ and*

$$\sum_{k=0}^{\infty} E\left(\|A_0 \cdots A_{-k+1}\|^{\alpha+\eta}\right)^{1/(\alpha+\eta)} < \infty, \ \sum_{k=0}^{\infty} E\left(\|A_0 \cdots A_{-k+1}\|^{\alpha-\eta}\right)^{1/(\alpha+\eta)} < \infty. \tag{4.17}$$

*Then the assertions of Theorem 4.1 hold. Moreover, $X_0$ (and therefore the "one"-dimensional marginal distribution of $X = (X_n)_{n\in\mathbb{Z}}$, i.e. the unique stationary solution of the stochastic difference equation $Y_n = A_n Y_{n-1} + C_n$) is multivariate regularly varying with index $\alpha$, normalizing sequence $(a_n)$ and measure*

$$\tilde{\nu}(\cdot) = \sum_{k=0}^{\infty} E\left(\nu \circ (A_0 A_{-1} \cdots A_{-k+1})^{-1}(\cdot)\right) = \nu(\cdot) + \sum_{k=1}^{\infty} E\left(\nu \circ (A_0 A_{-1} \cdots A_{-k+1})^{-1}(\cdot)\right). \tag{4.18}$$

*Moreover, $X_0$ is in $L^\beta$ for all $0 < \beta < \alpha$. If $C_0 \in L^\alpha$, then also $X_0 \in L^\alpha$.*

**Proof:** It mainly remains to show that all conditions of Theorems 4.1 and 3.19 are satisfied under the assumptions made. Proposition 2.23 and Theorem 2.24 give that $(A_n, C_n)_{n\in\mathbb{Z}}$ is a stationary and ergodic sequence. Since $A_0 \in L^{\alpha+\eta}$, $E(\log^+ \|A_0\|)$ is finite (because $\log(x)$ is for large $x$ ultimately bounded by any positive power of $x$). As $C_0$ is regularly varying with index $\alpha$, $C_0 \in L^\beta$ for all $0 < \beta < \alpha$ (cf. Proposition 3.12) and thus $E(\log^+ \|C_0\|) < \infty$. From (4.18) it is obvious that $\tilde{\nu} \geq \nu$. So the nondegeneracy of $\nu$ ensures that there is a relatively compact $K$ with $\tilde{\nu}(K) > 0$.

$X_0 \in \beta$ for all $0 < \beta < \alpha$ is an immediate consequence of Proposition 3.12. This could alternatively be shown using Theorem 4.7.

Assume now $C_0 \in L^\alpha$. For $\alpha \geq 1$, (4.17) implies immediately that

$$\sum_{k=0}^{\infty} \|A_0 \cdots A_{-k+1}\|_{L^\alpha} < \infty.$$

Using the independence of $(A_k)$ and $(C_k)$ this gives

$$\sum_{k=0}^{\infty} \|A_0 \cdots A_{-k+1} C_{-k}\|_{L^\alpha} \leq \|C_0\|_{L^\alpha} \sum_{k=0}^{\infty} \|A_0 \cdots A_{-k+1}\|_{L^\alpha} < \infty.$$

For $0 < \alpha < 1$, (4.16) implies immediately that

$$\sum_{k=0}^{\infty} E\left(\|A_0 \cdots A_{-k+1}\|^\alpha\right) < \infty.$$

Using the independence of $(A_k)$ and $(C_k)$ this gives

$$\sum_{k=0}^{\infty} E\left(\|A_0 \cdots A_{-k+1} C_{-k}\|\right) \leq E\left(\|C_0\|^\alpha\right) \sum_{k=0}^{\infty} E\left(\|A_0 \cdots A_{-k+1}\|^\alpha\right) < \infty.$$

So Theorem 4.7 gives that $X_0 \in L^\alpha$ and that the series defining $X_n$ converges in $L^\alpha$. □

The above theorem naturally raises the question, whether the regular variation of the "one"-dimensional marginal distribution of the series $X = (X_n)_{n\in\mathbb{Z}}$ is all that can be shown or whether one can also establish regular variation of the whole sequence $X$. For the later one needs to show that all finite dimensional distributions of $X$ are regularly varying. Unfortunately, a short proof along the lines of Basrak, Davis and Mikosch (2002b, Cor 2.7) cannot be given in our case, since $(A_i)_{i>k}$ does not have to be independent of $X_k$. However, one can still give a rather straightforward but notationally tedious proof using Theorem 3.19.

**Theorem 4.13** *If the conditions of Theorem 4.12 are satisfied, the solution $X = (X_n)_{n\in\mathbb{Z}}$ is even regularly varying as a sequence with index $\alpha$.*

**Proof:** It remains only to show that all finite dimensional distributions of $X = (X_n)$ are regularly varying. We restrict ourselves to show that the "two"-dimensional marginals are again regularly varying. It should be obvious that the very same arguments can be used for all higher dimensional marginals, but the general case is notationally most burdensome.

W.l.o.g. we only consider the joint distribution of $X_0$ and $X_h$ for $h \in \mathbb{N}$. From the series representations of $X_0 = C_0 + \sum_{k=1}^{\infty} A_0 A_{-1} \cdots A_{-k+1} C_{-k}$ and $X_h = C_h + \sum_{k=1}^{\infty} A_h A_{h-1} \cdots A_{h-k+1} C_{h-k}$ we can construct a series representation of $(X_0^\mathsf{T}, X_h^\mathsf{T})^\mathsf{T}$ as follows. Set

$$\mathbf{A}_h = \begin{pmatrix} 0_{M_d(\mathbb{R})} \\ I_d \end{pmatrix}$$

$$\mathbf{A}_{h-k} = \begin{pmatrix} 0_{M_d(\mathbb{R})} \\ A_h A_{h-1} \cdots A_{h-k+1} \end{pmatrix} \qquad \text{for } k = 1, 2, \ldots, h-1$$

$$\mathbf{A}_0 = \begin{pmatrix} I_d \\ A_h A_{h-1} \cdots A_1 \end{pmatrix}$$

$$\mathbf{A}_{h-k} = \begin{pmatrix} A_0 A_{-1} \cdots A_{h-k+1} \\ A_h A_{h-1} \cdots A_{h-k+1} \end{pmatrix} \qquad \text{for } k = h+1, h+2, \ldots.$$

Obviously one has $(X_0^\intercal, X_h^\intercal)^\intercal = \sum_{k=0}^\infty \mathbf{A}_{h-k} C_{h-k}$ and that the sequences $(\mathbf{A}_{h-k})_{k\in\mathbb{N}_0}$ and $(C_{h-k})_{k\in\mathbb{N}_0}$ are independent of each other. On $\mathbb{R}^{2d}$ now consider the norm $\|\cdot\|_*$ defined via the norm $\|\cdot\|$ used on $\mathbb{R}^d$ by $\|(x_1^\intercal, x_2^\intercal)^\intercal\| = \max\{\|x_1\|, \|x_2\|\}$. For any matrix $A \in M_{2d,d}(\mathbb{R})$ with $A = (A_1^\intercal, A_2^\intercal)^\intercal$ where $A_1, A_2 \in M_d(\mathbb{R})$ it holds that $\|A\|_* \leq \max\{\|A_1\|, \|A_2\|\} \leq \|A_1\| + \|A_2\|$. Using (4.16), resp. (4.17), we thus obtain from the definition of $\mathbf{A}_{h-i}$ that $\mathbf{A}_{h-i} \in L^{\alpha+\eta}$ for all $i \in \mathbb{N}_0$ and

$$\sum_{k=0}^\infty E\left(\|\mathbf{A}_{h-k}\|^{\alpha+\eta}\right) < \infty, \sum_{k=0}^\infty E\left(\|\mathbf{A}_{h-k}\|^{\alpha-\eta}\right) < \infty,$$

if $\alpha < 1$, respectively

$$\sum_{k=0}^\infty E\left(\|\mathbf{A}_{h-k}\|^{\alpha+\eta}\right)^{1/(\alpha+\eta)} < \infty, \sum_{k=0}^\infty E\left(\|\mathbf{A}_{h-k}\|^{\alpha-\eta}\right)^{1/(\alpha+\eta)} < \infty,$$

if $\alpha \geq 1$. So Theorem 3.19 gives that

$$nP\left((X_0^\intercal, X_h^\intercal)^\intercal \in a_n \cdot\right) \xrightarrow{v} \bar{\nu}(\cdot) := \sum_{k=0}^\infty E\left(\nu \circ \mathbf{A}_{h-k}^{-1}(\cdot)\right)$$

as $n \to \infty$. Since $\mathbf{A}_h^{-1}\left(0_{\mathbb{R}^d} \times K\right) = K$ for all $K \subseteq \mathbb{R}^d$, the nondegeneracy of $\nu$ ensures the nondegeneracy of $\bar{\nu}$ and, in particular, that $(X_0^\intercal, X_h^\intercal)^\intercal$ is multivariate regularly varying with index $\alpha$, measure $\bar{\nu}$ and normalizing sequence $(a_n)$. $\square$

Employing the Lemmata 3.21 and 4.5 we obtain the following result from the above two theorems which will usually suffice to deal with almost all situations one actually encounters.

**Corollary 4.14** *Let $A_n \in M_d(\mathbb{R})$, $n \in \mathbb{Z}$, be a stationary ergodic process and $C_n$ a sequence of i.i.d. regularly varying $\mathbb{R}^d$-valued random variables with index $\alpha > 0$, measure $\nu$ and normalizing sequence $(a_n)$ such that (iv) in Theorem 3.9 holds. Assume that the sequences $(A_n)_{n\in\mathbb{Z}}$ and $(C_n)_{n\in\mathbb{Z}}$ are independent of one another and furthermore that there is a $\beta > \alpha$ such that $A_0 A_{-1} \cdots A_{-n} \in L^\beta$ for all $n \in \mathbb{N}_0$ and*

$$\limsup_{n\to\infty} E\left(\|A_0 A_{-1} \cdots A_{-n}\|^\beta\right)^{1/(n+1)} < 1. \tag{4.19}$$

*Then the assertions of Theorem 4.1 hold. Moreover, $X_0$ (and therefore the "one"-dimensional marginal distribution of $X = (X_n)_{n\in\mathbb{Z}}$, i.e. the unique stationary solution of the stochastic difference equation $Y_n = A_n Y_{n-1} + C_n$) is multivariate regularly varying with index $\alpha$, normalizing sequence $(a_n)$ and measure*

$$\tilde{\nu}(\cdot) = \sum_{k=0}^\infty E\left(\nu \circ (A_0 A_{-1} \cdots A_{-k+1})^{-1}(\cdot)\right) = \nu(\cdot) + \sum_{k=1}^\infty E\left(\nu \circ (A_0 A_{-1} \cdots A_{-k+1})^{-1}(\cdot)\right). \tag{4.20}$$

*The whole sequence $X$ is also regularly varying with index $\alpha$.*

*Furthermore, $X_0$ is in $L^\beta$ for all $0 < \beta < \alpha$. If $C_0 \in L^\alpha$, then also $X_0 \in L^\alpha$.*

### 4.3.2 Light-Tailed Noise

A by now classical result of Kesten (1973) (and the accompanying paper Kesten (1974)) is that very light-tailed noise can result in a heavy-tailed stationary solution of the random difference equation $Y_n = A_n Y_{n-1} + C_n$ with $(A_n, C_n) \in M_d(\mathbb{R}) \times \mathbb{R}^d$ being an i.i.d. sequence, provided $A_n$ can have operator norm greater one. Extensions are to be found in LePage (1983), Saporta, Guivarc'h and LePage (2004) and Saporta (2004a). A very detailed discussion for the special case $d = 1$, including an alternative proof, is given in Goldie (1991). Furthermore examples that light-tailed noise can as well result in a light-tailed stationary solution, if $A_n$ is bound to be a contraction, are to be found in Goldie and Grübel (1996). Recently results in the spirit of those of Kesten were given in Klüppelberg and Pergamenchtchikov (2004) and the accompanying paper Klüppelberg and Pergamenchtchikov (2003) together with a streamlined proof. As they focused on the tail behaviour of autoregressive processes with random coefficients, we will not discuss their results here, but come back to them in a later section on the tails of random coefficient ARMA processes.

Let us now briefly repeat the main results below. Note that these results play an important role in many recent papers, e.g. Davis and Mikosch (1998), Mikosch and Stărică (2000), Basrak (2000), Basrak, Davis and Mikosch (2002b), Fasen, Klüppelberg and Lindner (2004) and Lindner and Maller (2004) to name a few. As the proofs are rather lengthy and technical we omit them and refer to the original literature instead. To compare the results to the ones from the previous section, where we analysed the case of a regularly varying noise, one only needs to employ Theorem 3.11. It is most noteworthy that the noise $(C_n)$ in all the coming theorems can be basically arbitrarily light-tailed. This indicates that the regular variation encountered in the stationary solution of the stochastic recurrence equation is not at all related to the noise, but emerges due to the possible occurrence of consecutive "large" values in the sequence $(A_n)$. Moreover, all following theorems presume that the joint sequence $(A_n, C_n)_{n\in\mathbb{N}}$ is i.i.d., but for fixed $n$ the random variables $A_n$ and $C_n$ may well depend upon each other.

The following summary of Kesten's results under the assumption that $(A_n, C_n)$ are positive is taken from Mikosch (2003).

**Theorem 4.15 (Kesten (1973, Theorems 3, 4))** *Let $((A_n, C_n))_{n\in\mathbb{Z}}$ be an i.i.d. sequence of random matrices $A_n \in M_d(\mathbb{R})$ with non-negative entries $A_{ij}$ and of $(\mathbb{R}^+)^d$-valued random variables $C_n \neq 0$ a.s. Assume that the following conditions are satisfied:*

*(i) There is some $\eta > 0$ such that $E(\|A_1\|_2^\eta) < 1$.*

*(ii) $A_1$ has no zero rows a.s.*

*(iii) The set*

$$\{\log \|a_n a_{n-1} \cdots a_1\|_2 : n \in \mathbb{N}, a_n a_{n-1} \cdots a_1 > 0 \text{ and } a_n, a_{n-1}, \ldots, a_1 \in supp(P_{A_1})\}$$

*generates a dense subgroup in $\mathbb{R}$ with respect to summation and the Euclidean topology. Here $P_{A_1}$ denotes the distribution of $A_1$ and $a_n a_{n-1} \cdots a_1 > 0$ means that all entries of the matrix are strictly positive.*

*(iv)* *There exists a* $\kappa > 0$ *such that*

$$E\left(\min_{i\in\{1,\ldots,d\}}\sum_{j=1}^{d} A_{ij}\right)^{\kappa} \geq d^{\kappa/2}$$

*and*

$$E\left(\|A_1\|_2^{\kappa}\log^{+}\|A_1\|_2\right) < \infty.$$

*Then the following statements hold:*

*(1)* *There exists a unique solution* $\tilde{\kappa} \in (0,\kappa]$ *to the equation*

$$\lim_{n\to\infty}\frac{1}{n}\log E(\|A_n\cdots A_1\|_2^{\tilde{\kappa}}) = 0.$$

*(2)* *If* $E(\|C_1\|_2^{\tilde{\kappa}}) < \infty$*, there exists a unique stationary solution* $(X_n)$ *to the stochastic recurrence equation* $Y_n = A_nY_{n-1} + C_n$ *which is given by (4.2).*

*(3)* *If* $E(\|C_1\|_2^{\tilde{\kappa}}) < \infty$*, then* $X_1$ *has the following regular variation (in the sense of Kesten) properties:*
*For all* $x \in \mathbb{R}^d\backslash\{0\}$*,* $\lim_{t\to\infty} t^{\tilde{\kappa}}P(\langle x, X_1\rangle > t) = w(x)$ *exists and is strictly positive for all* $(\mathbb{R}^+)^d$*-valued vectors* $x \neq 0$*.*

*In particular, all components of* $X_1$ *are regularly varying with index* $\tilde{\kappa}$*.*

Observe that condition $(i)$ implies strict negativity of the top Lyapunov coefficient.

Unfortunately, the general case, where the $A_n, C_n$ are no longer restricted to be non-negative, is considerably more involved. Below we first repeat the results for the $d$-dimensional case from Kesten and then the one-dimensional ones of Goldie (1991).

**Theorem 4.16 (Kesten (1973, Theorem 6))** *Let* $((A_n, C_n))_{n\in\mathbb{Z}}$ *be an i.i.d. sequence of matrices* $A_n \in M_d(\mathbb{R})$ *and* $\mathbb{R}^d$*-valued random variables* $C_n \neq 0$ *a.s. Assume that the following conditions are satisfied:*

*(i)* $E(\log^{+}\|A_1\|_2) < \infty$ *and the top Lyapunov exponent*

$$\gamma := \inf_{n\in\mathbb{N}_0}\left(\frac{1}{n}E\left(\log\|A_1A_2\cdots A_n\|_2\right)\right)$$

*is strictly negative.*

*(ii)* $A_1$ *is invertible a.s.*

*(iii)* *For every open set* $U \in \mathbb{S}^{d-1}$ *and* $x \in \mathbb{S}^{d-1}$ *there exists an* $n \in \mathbb{N}$ *such that*

$$P\left(\frac{x^{\mathsf{T}}A_1A_2\cdots A_n}{\|x^{\mathsf{T}}A_1A_2\cdots A_n\|_2} \in U\right) > 0.$$

(iv) *There exists an* $n \in \mathbb{N}$*, a cube* $K \subset \mathbb{R}^{d^2}$ *and an* $\eta > 0$ *such that the distribution of* $A_1 A_2 \cdots A_n$ *(after using an obvious isomorphism from* $M_d(\mathbb{R})$ *to* $\mathbb{R}^{d^2}$ *like vec) has a non-singular (w.r.t. the Lebesgue measure on* $\mathbb{R}^{d^2}$*) component with a density* $f$ *such that* $f(x) \geq \eta \, \forall \, x \in K$*.*

(v) *The set*

$$\{\log \|a_n a_{n-1} \cdots a_1\|_2 : n \in \mathbb{N}, a_n a_{n-1} \cdots a_1 \in M_\rho \text{ and } a_n, a_{n-1}, \ldots a_1 \in supp(P_{A_1})\}$$

*generates a dense subgroup in* $\mathbb{R}$ *with respect to summation and the Euclidean topology. Here* $P_{A_1}$ *denotes the distribution of* $A_1$ *and* $M_\rho \subset M_d(\mathbb{R})$ *are the* $d \times d$ *matrices over* $\mathbb{R}$ *that have the spectral radius as an algebraically simple eigenvalue and all other eigenvalues are strictly less in modulus.*

(vi) $P(C_1 = (I_d - A_1)r) < 1$ *for all* $r \in \mathbb{R}^d$*.*

(vii) *There exists a* $\kappa > 0$ *such that*

$$\begin{aligned} E\left(\left(\rho\left(\left(A_1 A_1^\intercal\right)^{-1}\right)\right)^{-\kappa/2}\right) &\geq 1, \\ E\left(\|A_1\|_2^\kappa \log^+ \|A_1\|_2\right) &< \infty, \\ E(\|C_1\|_2^\kappa) &< \infty. \end{aligned}$$

*Then the following statements hold:*

(1) *There exists a unique stationary solution* $(X_n)$ *to the stochastic recurrence equation* $Y_n = A_n Y_{n-1} + C_n$ *given by (4.2).*

(2) *Some* $\tilde{\kappa} \in (0, \kappa]$ *exists such that* $X_1$ *has the following regular variation (in the sense of Kesten) properties:*
*For all* $x \in \mathbb{R}^d \backslash \{0\}$*,* $\lim_{t\to\infty} t^{\tilde{\kappa}} P(\langle x, X_1 \rangle > t) = w(x)$ *exists and is strictly positive.*
*In particular, all components of* $X_1$ *are regularly varying with index* $\tilde{\kappa}$*.*

For refinements see LePage (1983) and Saporta (2004a) and note that the above definition of the top Lyapunov exponent is equivalent to the one we used previously.

**Theorem 4.17 (Goldie (1991, Theorem 4.1))** *Let* $((A_n, C_n))_{n\in\mathbb{Z}}$ *be an i.i.d. sequence of* $\mathbb{R}$*-valued random variables* $A_n$ *and* $C_n$*. Assume that there is some* $\kappa > 0$ *such that:*

(i) $E(|A_1|^\kappa) = 1$,

(ii) $E\left(|A_1|^\kappa \log^+ |A_1|\right) < \infty$,

(iii) $E(|C_1|^\kappa) < \infty$.

*If, moreover, the conditional law of* $\log |A_1|$ *given* $A_1 \neq 0$ *is non-arithmetic, i.e. not concentrated on any lattice* $r\mathbb{Z}$ *for some* $r \in \mathbb{R}$*, then the following statements hold:*

(1) *There exists a unique stationary solution* $(X_n)$ *to the stochastic recurrence equation* $Y_n = A_n Y_{n-1} + C_n$ *given by (4.2).*

*(2) $X_1$ has the following regular variation properties:*

$$P(X_1 > t) \sim K_+ t^{-\kappa}, \ t \to \infty \tag{4.21}$$

$$P(X_1 < -t) \sim K_- t^{-\kappa}, \ t \to \infty \tag{4.22}$$

*with some constants $K_+ \geq 0$ and $K_- \geq 0$. Furthermore $K_+ + K_- > 0$, iff $P(C_1 = (1 - A_1)r) < 1$ for each fixed $r \in \mathbb{R}$.*

*Let $(A, C) \stackrel{\mathscr{D}}{=} (A_1, C_1)$ be independent of $X \stackrel{\mathscr{D}}{=} X_1$. If $A \geq 0$ holds almost surely, then*

$$K_+ = \frac{E\left(\left((C + AX)^+\right)^\kappa - \left((AX)^+\right)^\kappa\right)}{\kappa E(|A|^\kappa \log |A|)} \tag{4.23}$$

$$K_- = \frac{E\left(\left((C + AX)^-\right)^\kappa - \left((AX)^-\right)^\kappa\right)}{\kappa E(|A|^\kappa \log |A|)} \tag{4.24}$$

*and else*

$$K_+ = K_- = \frac{E\left(|C + AX|^\kappa - |AX|^\kappa\right)}{2\kappa E(|A|^\kappa \log |A|)}.$$

As usual $\sim$ denotes asymptotic equivalence and $(x)^- = \max\{-x, 0\}$. Observe that $K_+ > 0$, resp. $K_- > 0$, imply that the upper, resp. lower, tail of the stationary solution is regularly varying with index $\kappa$. Of course, we do not have to resort to multivariate regular variation above, but use the one-dimensional Definition 3.6.

From the results of all three theorems above one can under appropriate conditions rather immediately conclude that the solution $(Y_n)_{n\in\mathbb{Z}}$ of the stochastic difference equation is regularly varying as a sequence (see Basrak, Davis and Mikosch (2002b, Cor. 2.7) for the details). In particular one needs to have that $Y_1$ is regularly varying in the sense of Kesten, i.e. the limits in the above theorems have to be non-zero, and that the conditions of Theorem 3.11 (ii), (iii) or (iv) are satisfied.

# Chapter 5

# Markov-Switching ARMA Models

The aim of this chapter is to introduce (multivariate) Markov switching autoregressive moving-average (MS-ARMA) time series models, provide conditions for the existence of a stationary solution of an MS-ARMA equation and study some of their properties.

## 5.1 Definition

In time series analysis a stationary stochastic process $X_t \in \mathbb{R}^d, t \in \mathbb{Z}$, is said to be a $d$-dimensional ARMA$(p,q)$ process (cf. Brockwell and Davis (1991, Definition 11.3.1 and Definition 3.1.2 for the univariate case)), if

$$X_t - \Phi_1 X_{t-1} - \ldots - \Phi_p X_{t-p} = Z_t + \Theta_1 Z_{t-1} + \ldots + \Theta_q Z_{t-q} \tag{5.1}$$

holds for all $t \in \mathbb{Z}$, where $\Phi_1, \ldots, \Phi_p, \Theta_1, \ldots, \Theta_q \in M_d(\mathbb{R})$ are the parameters and $(Z_t)_{t\in\mathbb{Z}} \sim WN(0, \Sigma)$ is $d$-dimensional white noise, i.e. a sequence of uncorrelated random variables in $\mathbb{R}^d$ with common mean 0 and covariance matrix $\Sigma$. The parameters $\Phi_1, \ldots, \Phi_p$ and $\Theta_1, \ldots, \Theta_q$ are called AR and MA coefficients, respectively. (5.1) is said to be an ARMA equation. The definition above implies that an ARMA process has mean 0. If a process $(X_t)_{t\in\mathbb{Z}}$ has mean $\mu \in \mathbb{R}^d$ and $(X_t - \mu)_{t\in\mathbb{Z}}$ is an ARMA$(p,q)$ process, one calls $X$ an ARMA$(p,q)$ process with mean $\mu$. There are also extensions of ARMA processes to the case where $(Z_t)$ is not (square) integrable (see e.g. Brockwell and Davis (1991, Section 13.3)).

ARMA$(p,q)$ processes are highly tractable and their behaviour and properties are well-known, see e.g. Brockwell and Davis (1991) and references therein. Yet, they are incapable of modelling non-linearities in time series. To deal with time series that exhibit only piecewise linear behaviour, several modifications of the ARMA model have been used. One such modification are MS-ARMA processes, where the ARMA coefficients are allowed to change over time according to a Markov chain. Another one is the autoregressive threshold (TAR) model (see e.g. Brachner (2004) and references therein), where the AR coefficients vary dependent upon the current value of the process.

**Definition 5.1 (MS-ARMA$(p,q)$ process)** *Let $p, q \in \mathbb{N}_0$, $p + q \geq 1$ and*

$$\Delta = (\mu_t, \Sigma_t, \Phi_{1t}, \ldots, \Phi_{pt}, \Theta_{1t}, \ldots, \Theta_{qt})_{t\in\mathbb{Z}} \tag{5.2}$$

*be a stationary and ergodic Markov chain with some (measurable) subset $E$ of the linear space $\mathbb{R}^d \times M_d(\mathbb{R})^{1+p+q}$ as state space. Moreover, let $\epsilon = (\epsilon_t)_{t\in\mathbb{Z}}$ be an i.i.d. sequence of $\mathbb{R}^d$-valued random variables independent of $\Delta$ and set $Z_t := \Sigma_t \epsilon_t$ (i.e. the matrix-vector product of $\Sigma_t$ and $\epsilon_t$). A stationary process $(X_t)_{t\in\mathbb{Z}}$ in $\mathbb{R}^d$ is called* MS-ARMA$(p, q, \Delta, \epsilon)$ *process, if*

$$X_t - \Phi_{1t}X_{t-1} - \cdots - \Phi_{pt}X_{t-p} = \mu_t + Z_t + \Theta_{1t}Z_{t-1} + \cdots + \Theta_{qt}Z_{t-q} \tag{5.3}$$

*holds for all $t \in \mathbb{Z}$.*

*Furthermore, a stationary process $(X_t)_{t\in\mathbb{Z}}$ is said to be an MS-ARMA$(p, q)$ process, if it is an* MS-ARMA$(p, q, \Delta, \epsilon)$ *process for some $\Delta$ and $\epsilon$ satisfying the above conditions.*

If $p$, respectively $q$, is zero, it is implicitly understood that the autoregressive, respectively moving-average, terms vanish. Actually, we shall presume $p \geq 1$ from now on in order to avoid degeneracies in the employed representations. This does not effect the generality of our results, as we can always simply set $\Phi_{1t} \equiv 0$, whenever $p$ was zero otherwise.

The different values that the Markov chain $\Delta$ can assume are called regimes and w.l.o.g. one can take the sequence $(\epsilon_t)$ to have zero mean, if $\epsilon_t \in L^1$. If $\epsilon_t$ is even in $L^2$, one has $\mathrm{cov}(Z_t|\Sigma_t) = \Sigma_t \mathrm{cov}(\epsilon_t)\Sigma_t^\top$.

The above definition is mainly taken from Francq and Zakoïan (2001). Yet, it is, in some respects, more general, since we do not require the Markov chain $\Delta$ to have only a finite state space and the noise to have a finite variance. But on the other hand we restrict ourselves to an i.i.d. noise sequence $(\epsilon_t)$ instead of a $d$-dimensional white noise sequence. Note, however, that the results in the next sections on stationarity conditions can immediately be extended to the case of $\epsilon$ being not an i.i.d. but a stationary and mixing sequence. The above definition is motivated by our interest in studying MS-ARMA processes in the presence of a noise sequence, that may have no finite second moment, and the fact that the results on the existence and uniqueness of a stationary solution to equation (5.3) for given $\Delta$ and $\epsilon$, given e.g. in Francq and Zakoïan (2001), carry immediately over to the more general situation characterized above as well. To the best of our knowledge, MS-ARMA$(p, q)$ processes with driving Markov Chains that may have non-finite state spaces have not been discussed explicitly in the available literature so far from a probabilistic point of view. However, regarding maximum likelihood estimation a compact state space has been allowed in Douc, Moulines and Rydén (2004). Note that the above definition implies that autoregressive models with i.i.d. random coefficients, as studied for example by Nicholls and Quinn (1982) or Klüppelberg and Pergamenchtchikov (2004), are special cases of MS-ARMA models.

## 5.2 Stationarity and Basic Properties of MS-ARMA Processes

In this section we will prove conditions under which there exists a unique stationary solution to an MS-ARMA equation (5.3) for a given parameter chain $\Delta$ and noise sequence $\epsilon$. This issue has already been studied by several authors under the assumption of a finite state space Markov chain $\Delta$. Francq and Zakoïan (2001) give the general stationarity condition, which we shall show to be valid for our set-up in the following. Moreover, they also give conditions for the solution to be second order stationary. Yao (2001) studies the existence of stationary and square-integrable MS-AR(p) (i.e. MS-ARMA(p,0)) processes and Yang (2000) solely studies covariance stationary processes and gives sufficient conditions that can be considerably weakened, as pointed out in Francq and Zakoïan (2002). The $L^2$-stationarity and -structure of MS-ARMA models has also been considered in Zhang and Stine (2001). Related work is furthermore Yao and Attali (2000), who study non-linear MS-AR processes, i.e. processes that satisfy $X_n = f_{\Delta_n}(X_{n-1}) + \epsilon_n$ for a Markov chain $\Delta$ and a noise sequence $\epsilon$. They especially look at the cases of sublinear or Lipschitz continuous functions $f_{\Delta_n}$. The ergodicity and stationarity of certain non-linear Markov switching autoregressions is furthermore considered in Francq and Roussignol (1998).

### 5.2.1 Sufficient Conditions for the Existence of Stationary MS-ARMA Processes

Using the general results on stochastic difference equations of the form $Y_n = A_n Y_{n-1} + C_n$, we are in a position to give sufficient conditions for the existence of a unique stationary and ergodic solution to a $d$-dimensional MS-ARMA$(p, q, \Delta, \epsilon)$ equation given by (5.3),

$$X_t - \Phi_{1t} X_{t-1} - \ldots - \Phi_{pt} X_{t-p} = \mu_t + Z_t + \Theta_{1t} Z_{t-1} + \ldots + \Theta_{qt} Z_{t-q},$$

where as in Definition 5.1 $\Delta = (\mu_t, \Sigma_t, \Phi_{1t}, \ldots, \Phi_{pt}, \Theta_{1t}, \ldots, \Theta_{qt})_{t \in \mathbb{Z}}$ is a stationary and ergodic Markov chain with some subset of $\mathbb{R}^d \times M_d(\mathbb{R})^{1+p+q}$ as state space, $(\epsilon_t)_{t \in \mathbb{Z}}$ is an i.i.d. sequence of $\mathbb{R}^d$-valued random variables independent of $\Delta$ and $Z_t := \Sigma_t \epsilon_t$.

The basic idea is to employ a higher dimensional representation, partly similar to the state space representation for ARMA-models (see e.g. Brockwell and Davis (1991, Examples 12.1.5, 12.1.6)), since then the result of Brandt (1986) given previously can be applied straightforwardly. Note that all zeros appearing below denote the zero in $M_{m,n}(\mathbb{R})$, resp. $\mathbb{R}^d$, with the appropriate dimensions $m, n$, resp. $d$, being obvious from the context. We define

$$\mathbf{X}_t = \begin{pmatrix} X_t \\ X_{t-1} \\ \vdots \\ X_{t-p+1} \\ Z_t \\ \vdots \\ Z_{t-q+1} \end{pmatrix} \in \mathbb{R}^{d(p+q)}, \tag{5.4}$$

$$\mathbf{m}_t = \begin{pmatrix} \mu_t \\ 0 \\ \vdots \\ 0 \end{pmatrix} \in \mathbb{R}^{d(p+q)}, \tag{5.5}$$

$$\boldsymbol{\Sigma}_t = \begin{pmatrix} \Sigma_t \\ \left.\begin{matrix} 0 \\ \vdots \\ 0 \end{matrix}\right\} p-1 \\ \Sigma_t \\ \left.\begin{matrix} 0 \\ \vdots \\ 0 \end{matrix}\right\} q-1 \end{pmatrix} \in M_{d(p+q),d}(\mathbb{R}), \tag{5.6}$$

$$\boldsymbol{\Phi}_t = \begin{pmatrix} \Phi_{1t} & \cdots & \Phi_{(p-1)t} & \Phi_{pt} \\ I_d & 0\cdots & \cdots & 0 \\ 0 & \ddots & \ddots & \vdots \\ 0 & \cdots 0 & I_d & 0 \end{pmatrix} \in M_{dp}(\mathbb{R}), \tag{5.7}$$

$$\boldsymbol{\Theta}_t = \begin{pmatrix} \Theta_{1t} & \cdots & \Theta_{(q-1)t} & \Theta_{qt} \\ 0 & \cdots & \cdots & 0 \\ 0 & \cdots & \cdots & \vdots \\ 0 & \cdots & \cdots & 0 \end{pmatrix} \in M_{dp,dq}(\mathbb{R}), \tag{5.8}$$

$$\mathbf{J} = \begin{pmatrix} 0 & \cdots & \cdots & 0 \\ I_d & 0 & \cdots & 0 \\ 0 & \ddots & 0\cdots & \vdots \\ 0 & \cdots 0 & I_d & 0 \end{pmatrix} \in M_{dq}(\mathbb{R}), \tag{5.9}$$

$$\begin{aligned} \mathbf{A}_t &= \begin{pmatrix} \Phi_{1t} & \cdots & \Phi_{(p-1)t} & \Phi_{pt} & \Theta_{1t} & \cdots & \Theta_{(q-1)t} & \Theta_{qt} \\ I_d & 0\cdots & \cdots & 0 & 0 & \cdots & \cdots & 0 \\ 0 & \ddots & \ddots & 0 & 0 & \cdots & \cdots & \vdots \\ 0 & \cdots 0 & I_d & 0 & 0 & \cdots & \cdots & \vdots \\ 0 & \cdots & \cdots & 0 & 0 & \cdots & \cdots & 0 \\ 0 & \cdots & \cdots & 0 & I_d & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \vdots & 0 & \ddots & 0\cdots & \vdots \\ 0 & \cdots & \cdots & 0 & 0 & \cdots 0 & I_d & 0 \end{pmatrix} \\ &= \begin{pmatrix} \boldsymbol{\Phi}_t & \boldsymbol{\Theta}_t \\ 0 & \mathbf{J} \end{pmatrix} \in M_{d(p+q),d(p+q)}(\mathbb{R}), \end{aligned} \tag{5.10}$$

$$\mathbf{C}_t = \mathbf{m}_t + \boldsymbol{\Sigma}_t \epsilon_t. \tag{5.11}$$

(In the case of a purely autoregressive MS-ARMA equation, i.e. $q = 0$, it is implicitly understood above that $\mathbf{J}_t$ and $\boldsymbol{\Theta}_t$ vanish, $\mathbf{X}_t = \left(X_t^\top, X_{t-1}^\top, \ldots, X_{t-p+1}^\top\right)^\top$, $\boldsymbol{\Sigma}_t = \left(\Sigma_t^\top, 0^\top, \ldots, 0^\top\right)^\top$ and $\mathbf{A}_t = \boldsymbol{\Phi}_t$. Moreover, recall that we presume $p \geq 1$ w.l.o.g.)

Then (5.3) has the representation

$$\mathbf{X}_t = \mathbf{A}_t \mathbf{X}_{t-1} + \mathbf{C}_t \tag{5.12}$$

in this higher dimensional set-up. We obviously have that any (stationary) solution of (5.3) leads via the above transformations to a (stationary) solution of (5.12) and, vice versa, that the first component of a (stationary) solution of (5.12) is also a (stationary) solution of (5.3). Moreover, an ergodic solution of (5.12) gives also an ergodic solution of (5.3), as a straightforward application of Lemma 2.25 to the map $\mathbf{X}_t = (X_t, \ldots, X_{t-p+1}, Z_t, \ldots, Z_{t-q+1})^\mathsf{T} \mapsto X_t$ shows.

Proposition 2.23 and Theorem 2.24 imply that the joint random sequence $(\Delta, \epsilon) = ((\Delta_t, \epsilon_t))_{t\in\mathbb{Z}}$ is stationary and ergodic and thus an obvious application of Lemma 2.25 shows that the transformed sequence $(\mathbf{A}, \mathbf{C}) = ((\mathbf{A}_t, \mathbf{C}_t))_{t\in\mathbb{Z}}$ is stationary and ergodic. Hence, we obtain the following result from Theorem 4.1 stating sufficient conditions for (5.3) to have a solution.

**Theorem 5.2** *The MS-ARMA$(p, q, \Delta, \epsilon)$ equation (5.3) has a unique stationary and ergodic solution, if $E(\log^+ \|\mathbf{A}_0\|)$, $E(\log^+ \|\mathbf{C}_0\|)$ are finite and the Lyapunov exponent satisfies*

$$\gamma = \inf_{t\in\mathbb{N}_0} \left( \frac{1}{t+1} E\left(\log \|\mathbf{A}_0\mathbf{A}_{-1}\cdots\mathbf{A}_{-t}\|\right)\right) < 0.$$

*The unique stationary solution $X = (X_t)$ is formed by the first $d$ coordinates of*

$$\mathbf{X}_t = \sum_{k=0}^{\infty} \mathbf{A}_t\mathbf{A}_{t-1}\cdots\mathbf{A}_{t-k+1}\mathbf{C}_{t-k}, \tag{5.13}$$

*which is the unique stationary and ergodic solution of (5.12). The series defining $\mathbf{X}$ converges absolutely a.s. (cf. Francq and Zakoïan (2001, Th. 1))*

*Let $\mathbf{V}_0$ be an arbitrary $\mathbb{R}^{d(p+q)}$-valued random variable defined on the same probability space as $(\Delta_t, \epsilon_t)_{t\in\mathbb{Z}}$ and define $(\mathbf{V}_t)_{t\in\mathbb{N}}$ recursively via (5.12) (or let $V_0, \ldots, V_{-p+1}, Z_0, \ldots, Z_{-q+1}$ be arbitrary $\mathbb{R}^d$ valued random variables and define $(V_t)_{t\in\mathbb{N}}$ via (5.3), $\mathbf{V}_t := (V_t, \ldots, V_{t-p+1}, Z_t, \ldots, Z_{t-q+1})^\mathsf{T}$). Then*

$$\|\mathbf{X}_t - \mathbf{V}_t\| \overset{\text{a.s.}}{\to} 0 \text{ as } t \to \infty \tag{5.14}$$

*and, in particular,*

$$\mathbf{V}_t \overset{\mathscr{D}}{\to} \mathbf{X}_0 \text{ as } t \to \infty, \tag{5.15}$$

*i.e. the distribution of $\mathbf{V}_t$ converges to the stationary distribution of $\mathbf{X}_t$.*

Note that in view of Lemma 4.2 it suffices to verify that $E(\log^+ \|\mathbf{A}_0\|)$, respectively $E(\log^+ \|\mathbf{C}_0\|)$, holds for some arbitrary norm on $M_{d(p+q)}(\mathbb{R})$, respectively $\mathbb{R}^{d(p+q)}$, and $\inf_{t\in\mathbb{N}} E\frac{1}{t+1}\left(\log \|\mathbf{A}_0\mathbf{A}_{-1}\cdots\mathbf{A}_{-t}\|\right) < 0$ for some (possibly different) algebra norm in order to be able to employ the above theorem. The next technical lemmata will lead to the insight that it is sufficient to study the behaviour of the AR-coefficients $(\mathbf{\Phi}_t)$ to be able to judge, whether $\inf_{t\in\mathbb{N}} \frac{1}{t+1} E\left(\log \|\mathbf{A}_0\mathbf{A}_{-1}\cdots\mathbf{A}_{-t}\|\right)$ is strictly negative or not. We follow, with some minor differences, the sketch provided in Francq and Zakoïan (2001), who give only the main ideas but no detailed proofs.

**Lemma 5.3 (cf. Francq and Zakoïan (2001, p. 343))** *For all $k \in \mathbb{N}$ and $t \in \mathbb{Z}$:*

$$\mathbf{A}_t\mathbf{A}_{t-1}\cdots\mathbf{A}_{t-k+1} = \begin{pmatrix} \mathbf{\Phi}_t\mathbf{\Phi}_{t-1}\cdots\mathbf{\Phi}_{t-k+1} & \mathbf{\Theta}_t\mathbf{J}^{k-1} + \sum_{l=0}^{k-2}\mathbf{\Phi}_t\mathbf{\Phi}_{t-1}\cdots\mathbf{\Phi}_{t-k+l+2}\mathbf{\Theta}_{t-k+l+1}\mathbf{J}^l \\ 0 & \mathbf{J}^k \end{pmatrix} \tag{5.16}$$

*If, moreover, $k \geq q+1$:*

$$\mathbf{A}_t\mathbf{A}_{t-1}\cdots\mathbf{A}_{t-k+1} = \begin{pmatrix} \mathbf{\Phi}_t\mathbf{\Phi}_{t-1}\cdots\mathbf{\Phi}_{t-k+1} & \sum_{l=0}^{q-1}\mathbf{\Phi}_t\mathbf{\Phi}_{t-1}\cdots\mathbf{\Phi}_{t-k+l+2}\mathbf{\Theta}_{t-k+l+1}\mathbf{J}^l \\ 0 & 0 \end{pmatrix} \tag{5.17}$$

**Proof:** For $k = 1$ (5.16) is the definition of $\mathbf{A}_t$ given in (5.10). Assume now that (5.16) holds for some $k \in \mathbb{N}$. Then we have for $k+1$

$$\begin{aligned}
\mathbf{A}_t\cdots\mathbf{A}_{t-k} &= \begin{pmatrix} \mathbf{\Phi}_t\mathbf{\Phi}_{t-1}\cdots\mathbf{\Phi}_{t-k+1} & \mathbf{\Theta}_t\mathbf{J}^{k-1} + \sum_{l=0}^{k-2}\mathbf{\Phi}_t\mathbf{\Phi}_{t-1}\cdots\mathbf{\Phi}_{t-k+l+2}\mathbf{\Theta}_{t-k+l+1}\mathbf{J}^l \\ 0 & \mathbf{J}^k \end{pmatrix} \\
&\times \begin{pmatrix} \mathbf{\Phi}_{t-k} & \mathbf{\Theta}_{t-k} \\ 0 & \mathbf{J} \end{pmatrix} \\
&= \begin{pmatrix} \mathbf{\Phi}_t\cdots\mathbf{\Phi}_{t-k} & \mathbf{\Phi}_t\cdots\mathbf{\Phi}_{t-k+1}\mathbf{\Theta}_{t-k} + \mathbf{\Theta}_t\mathbf{J}^k + \sum_{l=0}^{k-2}\mathbf{\Phi}_t\cdots\mathbf{\Phi}_{t-k+l+2}\mathbf{\Theta}_{t-k+l+1}\mathbf{J}^{l+1} \\ 0 & \mathbf{J}^{k+1} \end{pmatrix} \\
&= \begin{pmatrix} \mathbf{\Phi}_t\cdots\mathbf{\Phi}_{t-k} & \mathbf{\Theta}_t\mathbf{J}^k + \sum_{l=0}^{k-1}\mathbf{\Phi}_t\cdots\mathbf{\Phi}_{t-k+l+1}\mathbf{\Theta}_{t-k+l}\mathbf{J}^l \\ 0 & \mathbf{J}^{k+1} \end{pmatrix}.
\end{aligned}$$

Hence, (5.16) is shown by induction. (5.17) immediately follows by noting that $\mathbf{J}$ is nilpotent with index $q$, i.e. $\mathbf{J}^q = 0$. □

**Lemma 5.4 (cf. Francq and Zakoïan (2001, p. 343))** *For all natural $k \geq q+1$ and $t \in \mathbb{Z}$ the following identities hold:*

$$\mathbf{A}_t\cdots\mathbf{A}_{t-k+1} = \begin{pmatrix} \mathbf{\Phi}_t & 0 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} \mathbf{\Phi}_{t-1} & 0 \\ 0 & 0 \end{pmatrix}\cdots\begin{pmatrix} \mathbf{\Phi}_{t-k+q+2} & 0 \\ 0 & 0 \end{pmatrix}\mathbf{A}_{t-k+q+1}\cdots\mathbf{A}_{t-k+1}, \tag{5.18}$$

$$\begin{aligned}
&\begin{pmatrix} \mathbf{\Phi}_t & 0 \\ 0 & 0 \end{pmatrix}\cdots\begin{pmatrix} \mathbf{\Phi}_{t-k+1} & 0 \\ 0 & 0 \end{pmatrix} \\
&= \mathbf{A}_t\mathbf{A}_{t-1}\cdots\mathbf{A}_{t-k+q+2}\begin{pmatrix} \mathbf{\Phi}_{t-k+q+1} & 0 \\ 0 & 0 \end{pmatrix}\cdots\begin{pmatrix} \mathbf{\Phi}_{t-k+1} & 0 \\ 0 & 0 \end{pmatrix}
\end{aligned} \tag{5.19}$$

**Proof:** For $k = q + 1$ the first identity (5.18) obviously holds. Assume now that (5.18) is valid for some $k \geq q + 1$ and all $t$, then using (5.18) for $t - 1$ and $k$ one obtains for $k + 1$:

$$\begin{aligned}
& \begin{pmatrix} \mathbf{\Phi}_t & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} \mathbf{\Phi}_{t-1} & 0 \\ 0 & 0 \end{pmatrix} \cdots \begin{pmatrix} \mathbf{\Phi}_{t-(k+1)+q+2} & 0 \\ 0 & 0 \end{pmatrix} \mathbf{A}_{t-(k+1)+q+1} \cdots \mathbf{A}_{t-(k+1)+1} = \\
= \; & \begin{pmatrix} \mathbf{\Phi}_t & 0 \\ 0 & 0 \end{pmatrix} \mathbf{A}_{t-1} \cdots \mathbf{A}_{t-1-k+1} \\
\stackrel{(5.17)}{=} \; & \begin{pmatrix} \mathbf{\Phi}_t & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} \mathbf{\Phi}_{t-1} \cdots \mathbf{\Phi}_{t-1-k+1} & \sum_{l=0}^{q-1} \mathbf{\Phi}_{t-1} \cdots \mathbf{\Phi}_{t-1-k+l+2} \mathbf{\Theta}_{t-1-k+l+1} \mathbf{J}^l \\ 0 & 0 \end{pmatrix} \\
= \; & \begin{pmatrix} \mathbf{\Phi}_t \mathbf{\Phi}_{t-1} \cdots \mathbf{\Phi}_{t-(k+1)+1} & \sum_{l=0}^{q-1} \mathbf{\Phi}_t \mathbf{\Phi}_{t-1} \cdots \mathbf{\Phi}_{t-(k+1)+l+2} \mathbf{\Theta}_{t-(k+1)+l+1} \mathbf{J}^l \\ 0 & 0 \end{pmatrix} \\
\stackrel{(5.17)}{=} \; & \mathbf{A}_t \cdots \mathbf{A}_{t-(k+1)+1}.
\end{aligned}$$

This shows (5.18).
(5.19) is a simple consequence of (5.16):

$$\begin{aligned}
& \mathbf{A}_t \mathbf{A}_{t-1} \cdots \mathbf{A}_{t-k+q+2} \begin{pmatrix} \mathbf{\Phi}_{t-k+q+1} & 0 \\ 0 & 0 \end{pmatrix} \cdots \begin{pmatrix} \mathbf{\Phi}_{t-k+1} & 0 \\ 0 & 0 \end{pmatrix} \\
\stackrel{(5.16)}{=} \; & \begin{pmatrix} \mathbf{\Phi}_t \cdots \mathbf{\Phi}_{t-k+q+2} & \star \\ 0 & \star \end{pmatrix} \begin{pmatrix} \mathbf{\Phi}_{t-k+q+1} \cdots \mathbf{\Phi}_{t-k+1} & 0 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} \mathbf{\Phi}_t \cdots \mathbf{\Phi}_{t-k+1} & 0 \\ 0 & 0 \end{pmatrix} \\
= \; & \begin{pmatrix} \mathbf{\Phi}_t & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} \mathbf{\Phi}_{t-1} & 0 \\ 0 & 0 \end{pmatrix} \cdots \begin{pmatrix} \mathbf{\Phi}_{t-k+1} & 0 \\ 0 & 0 \end{pmatrix}.
\end{aligned}$$

□

**Proposition 5.5** *Assume that $E\left(\log^+ \|\mathbf{A}_0\|\right)$ is finite, then for arbitrary algebra norms on $M_{d(p+q)}(\mathbb{R})$ and $M_{dp}(\mathbb{R})$:*

$$\lim_{k\to\infty} \frac{1}{k} E\left(\log \|\mathbf{A}_0 \cdots \mathbf{A}_{-k+1}\|\right) = \lim_{k\to\infty} \frac{1}{k} E\left(\log \|\mathbf{\Phi}_0 \cdots \mathbf{\Phi}_{-k+1}\|\right).$$

**Proof:** W.l.o.g. we can assume that $\|\cdot\|$ is an operator norm (cf. Lemma 4.2) on $M_{d(p+q)}(\mathbb{R})$. For all $k \geq q + 1$ (5.18) gives

$$\begin{aligned}
& \frac{1}{k} E\left(\log \|\mathbf{A}_0 \cdots \mathbf{A}_{-k+1}\|\right) \\
= \; & \frac{1}{k} E\left(\log \left\| \begin{pmatrix} \mathbf{\Phi}_0 & 0 \\ 0 & 0 \end{pmatrix} \cdots \begin{pmatrix} \mathbf{\Phi}_{-k+q+2} & 0 \\ 0 & 0 \end{pmatrix} \mathbf{A}_{-k+q+1} \cdots \mathbf{A}_{-k+1} \right\| \right) \\
\leq \; & \frac{1}{k} E\left(\log \left\| \begin{pmatrix} \mathbf{\Phi}_0 & 0 \\ 0 & 0 \end{pmatrix} \cdots \begin{pmatrix} \mathbf{\Phi}_{-k+q+2} & 0 \\ 0 & 0 \end{pmatrix} \right\| \right) + \frac{1}{k} E\left(\log \|\mathbf{A}_{-k+q+1} \cdots \mathbf{A}_{-k+1}\|\right) \\
\leq \; & \frac{1}{k} E\left(\log \left\| \begin{pmatrix} \mathbf{\Phi}_0 & 0 \\ 0 & 0 \end{pmatrix} \cdots \begin{pmatrix} \mathbf{\Phi}_{-k+q+2} & 0 \\ 0 & 0 \end{pmatrix} \right\| \right) + (q+1) E\left(\frac{1}{k} \log^+ \|\mathbf{A}_0\|\right).
\end{aligned}$$

Since $\lim_{k\to\infty} (q+1) E\left(\frac{1}{k} \log^+ \|\mathbf{A}_0\|\right) = 0$, we thus obtain

$$\lim_{k\to\infty} E\left(\frac{1}{k} \log \|\mathbf{A}_0 \cdots \mathbf{A}_{-k+1}\|\right) \leq \lim_{k\to\infty} E\left(\frac{1}{k} \log \left\| \begin{pmatrix} \mathbf{\Phi}_0 & 0 \\ 0 & 0 \end{pmatrix} \cdots \begin{pmatrix} \mathbf{\Phi}_{-k+1} & 0 \\ 0 & 0 \end{pmatrix} \right\| \right).$$

Noting that the operator norm properties ensure

$$\left\| \begin{pmatrix} \mathbf{\Phi}_0 & 0 \\ 0 & 0 \end{pmatrix} \right\| \le \left\| \begin{pmatrix} \mathbf{\Phi}_0 & \mathbf{\Theta}_0 \\ 0 & \mathbf{J} \end{pmatrix} \right\| = \|\mathbf{A}_0\|,$$

an analogous argument using (5.19) shows

$$\lim_{k\to\infty} \frac{1}{k} E\left(\log \left\| \begin{pmatrix} \mathbf{\Phi}_0 & 0 \\ 0 & 0 \end{pmatrix} \cdots \begin{pmatrix} \mathbf{\Phi}_{-k+1} & 0 \\ 0 & 0 \end{pmatrix} \right\|\right) \le \lim_{k\to\infty} \frac{1}{k} E\left(\log \|\mathbf{A}_0 \cdots \mathbf{A}_{-k+1}\|\right).$$

Hence,

$$\lim_{k\to\infty} \frac{1}{k} E\left(\log \left\| \begin{pmatrix} \mathbf{\Phi}_0 & 0 \\ 0 & 0 \end{pmatrix} \cdots \begin{pmatrix} \mathbf{\Phi}_{-k+1} & 0 \\ 0 & 0 \end{pmatrix} \right\|\right) = \lim_{k\to\infty} \frac{1}{k} E\left(\log \|\mathbf{A}_0 \cdots \mathbf{A}_{-k+1}\|\right).$$

The operator norm $\|\cdot\|$ on $M_{d(p+q)}(\mathbb{R})$ induces an algebra norm on $M_{dp}(\mathbb{R})$ by setting $\|X\| = \left\| \begin{pmatrix} X & 0 \\ 0 & 0 \end{pmatrix} \right\|$ for $X \in M_{dp}(\mathbb{R})$. Since obviously $\|XY\| = \left\| \begin{pmatrix} X & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} Y & 0 \\ 0 & 0 \end{pmatrix} \right\|$, one obtains

$$\lim_{k\to\infty} \frac{1}{k} E\left(\log \|\mathbf{A}_0 \cdots \mathbf{A}_{-k+1}\|\right) = \lim_{k\to\infty} \frac{1}{k} E\left(\log \|\mathbf{\Phi}_0 \cdots \mathbf{\Phi}_{-k+1}\|\right).$$

Using this induced algebra norm and an application of Lemma 4.2 concludes the proof. □

**Corollary 5.6 (cf. Francq and Zakoïan (2001, p. 343))** *Assume that $E\left(\log^+ \|\mathbf{A}_0\|\right)$ is finite, then for arbitrary algebra norms on $M_{d(p+q)}(\mathbb{R})$ and $M_{dp}(\mathbb{R})$:*

$$\inf_{t\in\mathbb{N}_0} \left( \frac{1}{t+1} E\left(\log \|\mathbf{A}_0 \mathbf{A}_{-1} \cdots \mathbf{A}_{-t}\|\right) \right) = \inf_{t\in\mathbb{N}_0} \left( \frac{1}{t+1} E\left(\log \|\mathbf{\Phi}_0 \mathbf{\Phi}_{-1} \cdots \mathbf{\Phi}_{-t}\|\right) \right) \tag{5.20}$$

**Proof:** Combine Proposition 5.5 and Lemma 4.2. □

This shows that the Lyapunov coefficient of $(\mathbf{A}_t)$ can be replaced by the Lyapunov coefficient of $(\mathbf{\Phi}_t)$, the autoregressive part, in our sufficient conditions for the existence of a stationary solution of (5.3). For later reference we restate Theorem 5.2 with these modified conditions.

**Theorem 5.7** *The MS-ARMA$(p, q, \Delta, \epsilon)$ equation (5.3) has a unique stationary and ergodic solution, if $E(\log^+ \|\mathbf{A}_0\|)$, $E(\log^+ \|\mathbf{C}_0\|)$ are finite and the Lyapunov exponent satisfies*

$$\tilde{\gamma} = \inf_{t\in\mathbb{N}_0} \left( \frac{1}{t+1} E\left(\log \|\mathbf{\Phi}_0 \mathbf{\Phi}_{-1} \cdots \mathbf{\Phi}_{-t}\|\right) \right) < 0.$$

*The unique stationary solution $X = (X_t)$ is formed by the first $d$ coordinates of*

$$\mathbf{X}_t = \sum_{k=0}^{\infty} \mathbf{A}_t \mathbf{A}_{t-1} \cdots \mathbf{A}_{t-k+1} \mathbf{C}_{t-k},$$

*which is the unique stationary and ergodic solution of (5.12). The series defining $\mathbf{X}$ converges absolutely a.s. (cf. Francq and Zakoïan (2001, Th. 1))*

*Let $\mathbf{V}_0$ be an arbitrary $\mathbb{R}^{d(p+q)}$-valued random variable defined on the same probability space as $(\Delta_t, \epsilon_t)_{t\in\mathbb{Z}}$ and define $(\mathbf{V}_t)_{t\in\mathbb{N}}$ recursively via (5.12) (or let $V_0, \ldots, V_{-p+1}, Z_0, \ldots, Z_{-q+1}$ be arbitrary $\mathbb{R}^d$ valued random variables and define $(V_t)_{t\in\mathbb{N}}$ via (5.3), $\mathbf{V}_t := (V_t, \ldots, V_{t-p+1}, Z_t, \ldots, Z_{t-q+1})^\mathsf{T}$). Then*

$$\|\mathbf{X}_t - \mathbf{V}_t\| \stackrel{\text{a.s.}}{\to} 0 \text{ as } t \to \infty \tag{5.21}$$

*and, in particular,*

$$\mathbf{V}_t \stackrel{\mathscr{D}}{\to} \mathbf{X}_0 \text{ as } t \to \infty, \tag{5.22}$$

*i.e. the distribution of $\mathbf{V}_t$ converges to the stationary distribution of $\mathbf{X}_t$.*

### 5.2.2 Analysis of the Sufficient Conditions

The above theorem shows that, once the conditions $E(\log^+ \|\mathbf{A}_0\|)$, $E(\log^+ \|\mathbf{C}_0\|) < \infty$ are satisfied, only the behaviour of the AR part matters. This is comparable to the conditions for classical ARMA models (see Brockwell and Davis (1991, Th. 3.1.1+Remark 1, Th. 11.3.1)) to have a causal solution, which only impose restrictions on the AR coefficients. Note that the solution we obtain above for the MS-ARMA equation is also causal in the sense that it is constructed solely from the past and present of $(\Delta, \epsilon)$. That we need the condition $E \log^+ \|\mathbf{A}_0\| < \infty$ is due the fact that our ARMA coefficients are random. $E \log^+ \|\mathbf{A}_0\|$ depends on both the AR and MA coefficients. If the Markov chain $\Delta$ has only finitely many states or the state space is bounded, $E \log^+ \|\mathbf{A}_0\| < \infty$ is automatically fulfilled. Likewise, $E(\log^+ \|\mathbf{C}_0\|) < \infty$ is basically a condition on the noise sequence comparable to the condition for classical ARMA models (with infinite variance) that the noise is square integrable (cf. Brockwell and Davis (1991, p. 78)) or has Pareto-like tails (see Brockwell and Davis (1991, § 13.3)). Actually, assume that $\mu_0$ and $\Sigma_0$ are bounded and $(\epsilon_t)$ is a sequence of i.i.d. random variables in $L^\delta$ for some (finite) $\delta > 0$ (this is fulfilled by a square integrable noise or one with Pareto-like tails), then we have that $\mathbf{C}_0 = \mathbf{m}_0 + \mathbf{\Sigma}_0 \epsilon_0 \in L^\delta$ and thus $E \log^+ \|\mathbf{C}_0\| < \infty$, since $\log^+$ is ultimately bounded by any positive power (for any $\delta > 0$ using l'Hospital's rule: $\lim\limits_{x\to\infty} \log(x)/(x^\delta) = \lim\limits_{x\to\infty} 1/(\delta x^\delta) = 0$). Furthermore, it follows, using Hölder's inequality and Corollary 2.15, that $\mathbf{C}_0 \in L^\delta$, if $\mu_0 \in L^\delta$, $\Sigma_0 \in L^{\delta r}$ and $\epsilon_0 \in L^{\delta s}$ for some $r, s \geq 1$ such that $\frac{1}{r} + \frac{1}{s} = 1$, since $E(\|\Sigma_0\epsilon_0\|^\delta) \leq E((\|\Sigma_0\|\|\epsilon_0\|)^\delta) \leq E(\|\Sigma_0\|^{\delta r})^{1/r} E(\|\epsilon_0\|^{\delta s})^{1/s} < \infty$.

If the AR coefficients are not random but constants, $\tilde{\gamma} < 0$ translates into $\rho(\mathbf{\Phi}_0) = \lim_{n\to\infty} \|\mathbf{\Phi}_0^n\|^{1/n} < 1$ (for any algebra norm), which is equivalent to require that $\det(I_{dp} - z\mathbf{\Phi}) \neq 0$ for all $z \in \bar{B}_1(0)$. This is just the condition on a state space model to be causal (or stable) as defined in Brockwell and Davis (1991, p. 467). Moreover, for $d = 1$ $\det(I_{dp} - z\mathbf{\Phi}) \neq 0$ for all $z \in \bar{B}_1(0)$ is obviously equivalent to the condition that the AR-polynomial $1 - \Phi_1 z - \ldots - \Phi_p z^p$ vanishes nowhere on the closed unit disc; this is the sufficient condition for a classical ARMA-process to be causal (see Brockwell and Davis (1991, Th. 3.1.1 + Remark 1 and especially p. 468)).

The above considerations show that, if an MS-ARMA(p,q) model actually is an ARMA(p,q) model, the sufficient conditions for the existence of a stationary ("causal") solution become the sufficient conditions for the ARMA(p,q) model to have a causal solution.

Note that the latter are also necessary provided the AR and MA polynomial have no common zeros.

So far we have, however, no condition feasible in general, for when $\tilde{\gamma}$ (or $\gamma$) is actually strictly negative. For a constant AR part $\mathbf{\Phi}_t$ we have seen above that a spectral radius less than one gives $\tilde{\gamma} < 0$. From the representation of $\tilde{\gamma}$ as an infimum it obviously follows that $E(\log \|\mathbf{\Phi}_0\|) < 0$ suffices to ensure $\tilde{\gamma} < 0$ (recall from the remark after Theorem 4.1 that for $p = 1$ and $d = 1$ one even has $\tilde{\gamma} = E(\log |\mathbf{\Phi}_0|)$). Naturally, $\mathbf{\Phi}_0$ can be replaced by $\mathbf{A}_0$ and $\tilde{\gamma}$ by $\gamma$ in this considerations. At a first glance this condition seems to be only helpful for MS-ARMA(1,q) processes, since for $p > 1$ the matrix $\mathbf{\Phi}_0$ used in the higher dimensional representation is of the form

$$\begin{pmatrix} \star & \cdots & \star & \cdots & \star \\ I_d & 0 & \cdots & \cdots & 0 \\ 0 & \ddots & \ddots & \cdots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & I_d & 0 \end{pmatrix}$$

and will thus have operator norm one or larger for all norms on $\mathbb{R}^{dp}$, which assign the same length to all vectors in the canonical basis, and all usual norms satisfy this condition. When discussing feasible conditions for $\tilde{\gamma}$ to be negative in a later section, we shall, however, see that this observation is most valuable, since under certain conditions norms on $\mathbb{R}^d$ can be constructed such that a matrix of the above form has operator norm strictly less than one. Conditions requiring that matrices with a structure very similar to ours above may have (operator) norm less than one are heavily used in Basrak (2000, Section 3.2). Yet, Basrak (2000) does not discuss possible norms for which the conditions can actually be fulfilled. Similarly, $E(\log \|\mathbf{\Phi}_0 \cdots \mathbf{\Phi}_{-n}\|) < 0$ for some $n \in \mathbb{N}$ ensures $\tilde{\gamma} < 0$ using Lemma 4.2.

To the best of our knowledge there are no general necessary and sufficient conditions known for the existence of a stationary solution to a MS-ARMA(p,q) or a general stochastic difference equation of the type $Y_n = A_n Y_{n-1} + C_n$, so it is hard to say, whether or not our sufficient condition that the Lyapunov coefficient is strictly negative is close to necessity. For the stochastic difference equation $Y_n = A_n Y_{n-1} + C_n$ with $(A_n, C_n)_{n \in \mathbb{Z}}$ being an i.i.d. sequence in $M_d(\mathbb{R}) \times \mathbb{R}^d$ Bougerol and Picard (1992b) show that under some technical conditions the strict negativity of the Lyapunov coefficient actually is necessary for the existence of a stationary solution that can be represented as a function of past and present values of $(A_n, C_n)$ (i.e. a “causal” solution). This indicates that in our more general Markov chain set-up there is hope for our sufficient conditions to be close to necessity. Yet, generalizing the results of Bougerol and Picard (1992b) to a Markov Chain set-up seems to be rather involved and will not be pursued in the present thesis. Goldie and Maller (2000) give necessary and sufficient conditions for the existence of a stationary solution to the stochastic recurrence equation $Y_n = A_n Y_{n-1} + C_n$ with i.i.d. input $(A_n, C_n)$ in the one-dimensional case, i.e. that all random variables assume values in $\mathbb{R}$. Their conditions are considerably weaker than ours, since they do not demand that the solution should be “causal”. However, in our Markovian and the usual linear time series set-up it appears to be natural to study stationary solutions that are representable by past values of $(A_n, C_n)$. Thus we have chosen not to discuss other possible solutions. Related earlier work is also Vervaat (1979).

### 5.2.3 Existence of Moments

In the theorems to follow sufficient conditions for the existence of moments of an MS-ARMA$(p, q, \Delta, \epsilon)$ process are inferred from the general results given in Section 4.2. It is immediate that Lemmata 4.5 and 4.6 can be applied to both the sequences $(\mathbf{A}_t)$ and $(\mathbf{\Phi}_t)$. Observe that we now use $\tilde{p}$ when considering the finiteness of moments, as $p$ denotes already the autoregressive order of the MS-ARMA process.

The general Theorem 4.7 becomes for MS-ARMA processes:

**Theorem 5.8** *Assume the conditions of Theorem 5.2 or 5.7 are fulfilled. If moreover for some* $\tilde{p} \in (1, \infty]$

$$\sum_{k=0}^{\infty} \|\mathbf{A}_0 \cdots \mathbf{A}_{-k+1}\mathbf{C}_{-k}\|_{L^{\tilde{p}}} \tag{5.23}$$

*or for some* $\tilde{p} \in (0, 1)$

$$\sum_{k=0}^{\infty} E\left(\|\mathbf{A}_0 \cdots \mathbf{A}_{-k+1}\mathbf{C}_{-k}\|^{\tilde{p}}\right) \tag{5.24}$$

*converges, then the solution* $X_t$ *of the MS-ARMA equation (5.3) and its higher dimensional representation* $\mathbf{X}_t$ *are in* $L^{\tilde{p}}$*. Moreover, the series defining* $\mathbf{X}_t$ *(as given in Theorem 5.2 or 5.7) converges in* $L^{\tilde{p}}$*.*

It is clear that, if (5.23) holds for some $\tilde{p} \geq 1$, it holds for all $r \in [1, \tilde{p}]$ as well. However, if (5.24) holds for some $\tilde{p} \in (0, 1)$ this does not imply that it holds for all smaller values of $\tilde{p}$ as well. Yet, of course, if $X \in L^{\tilde{p}}$ for some $\tilde{p} > 0$ then $X \in L^r$ for all $r \in (0, \tilde{p}]$. Note, however, that the asymptotic conditions given in the next lemmata are much better behaved. If there is one $\tilde{p} \in (0, \infty]$ that fulfils the asymptotic condition, then the asymptotic conditions for all $s \in (0, \tilde{p}]$ are satisfied as well (use Jensen's inequality as in the proof of Lemma 3.20).

**Proof:** Combine Theorems 5.2/5.7 and 4.7 to obtain the results on $\mathbf{X}_t$. $X_t \in L^{\tilde{p}}$ is now a consequence of $\mathbf{X}_t \in L^{\tilde{p}}$ and Corollary 2.15. □

For later reference we also restate Propositions 4.9 and 4.10 for the special case of MS-ARMA processes.

**Proposition 5.9** *Let* $\tilde{p} \in (0, \infty)$*. If there exist* $r, s \geq 1$ *with* $1/r + 1/s = 1$*, such that* $\mathbf{A}_0 \cdots \mathbf{A}_{-k+1} \in L^{\tilde{p}r}\ \forall k \in \mathbb{N}$*,* $\limsup_{k\to\infty} E(\|\mathbf{A}_0 \cdots \mathbf{A}_{-k+1}\|^{\tilde{p}r})^{1/k} < 1$ *for* $0 < \tilde{p}r < \infty$*, resp.* $\lim_{k\to\infty} \|\mathbf{A}_0 \cdots \mathbf{A}_{-k+1}\|_{L^\infty}^{1/k} < 1$ *for* $\tilde{p}r = \infty$*, and* $\mathbf{C}_0 \in L^{\tilde{p}s}$*, then* $\gamma < 0$ *and (5.23) for* $\tilde{p} \geq 1$*, resp. (5.24) for* $0 < \tilde{p} < 1$*, hold.*

Again one especially obtains that, provided $\mathbf{A}_0 \in L^\infty$ (and thus $\mathbf{A}_0 \cdots \mathbf{A}_{-k+1} \in L^\infty$) and $\lim_{k\to\infty} \|\mathbf{A}_0 \cdots \mathbf{A}_{-k+1}\|_{L^\infty}^{1/k} < 1$, the MS-ARMA process $X_t$ and its higher dimensional representation $\mathbf{X}_t$ are in $L^{\tilde{p}}$, if $\mathbf{C}_0 \in L^{\tilde{p}}$. Note again that the condition $\mathbf{A}_0 \in L^\infty$ is automatically satisfied, if $\mathbf{A}_0$ has a finite or bounded state space.

**Proposition 5.10** *If* $\mathbf{A}_0 \in L^\infty\ \forall k \in \mathbb{N}$, $\lim_{k\to\infty} \|\mathbf{A}_0 \cdots \mathbf{A}_{-k+1}\|_{L^\infty}^{1/k} < 1$ *and* $\mathbf{C}_0 \in L^\infty$, *then* $\gamma < 0$ *and (5.23) with* $\tilde{p} = \infty$ *hold.*

Furthermore, note again that $\|\mathbf{A}_0\| < c$ a.s. for some $c < 1$ implies the validity of $\lim_{k\to\infty} \|\mathbf{A}_0 \cdots \mathbf{A}_{-k+1}\|_{L^\infty}^{1/k} < 1$ (cf. Lemma 4.11 and the brief discussion thereafter). If we have that $\mathbf{A} = (\mathbf{A}_t)$ is a sequence of independent random variables (as it is for example the case in the random coefficient autoregressive models studied in Nicholls and Quinn (1982), Klüppelberg and Pergamenchtchikov (2004) and other papers), $E(\|\mathbf{A}_0\|^{\tilde{p}}) < 1$ ensures $\lim_{k\to\infty} E(\|\mathbf{A}_0 \cdots \mathbf{A}_{-k+1}\|^{\tilde{p}})^{1/k} < 1$, since $\|\cdot\|$ is submultiplicative. These observations may seem to be of limited interest now, especially in view of the discussion in Section 5.2.2, but later on in the discussion of feasible sufficient conditions they turn out to be helpful.

The most straightforward moment conditions are obtainable under the assumption that $\mathbf{A} = (\mathbf{A}_t)$ and $\mathbf{C} = (\mathbf{C}_t)$ are independent. This happens, if $\Sigma_t$ and $\mu_t$ are constants or at least independent from the other components of the Markov chain $\Delta$. In this case one obtains the following simplification of Proposition 5.9.

**Proposition 5.11** *Let* $\mathbf{A}_0\mathbf{A}_{-1} \cdots \mathbf{A}_{-k+1}$ *be independent of* $\mathbf{C}_{-k}$ *for all* $k \in \mathbb{N}$ *and* $\tilde{p} \in (0, \infty)$. *If* $\mathbf{A}_0 \cdots \mathbf{A}_{-k+1} \in L^{\tilde{p}}\ \forall\, k$, $\mathbf{C}_0 \in L^{\tilde{p}}$ *and* $\limsup_{k\to\infty} E(\|\mathbf{A}_0 \cdots \mathbf{A}_{-k+1}\|^{\tilde{p}})^{1/k} < 1$, *then (5.23) for* $\tilde{p} \geq 1$, *resp. (5.24) for* $0 < \tilde{p} < 1$, *holds.*

The prerequisite independence is in particular satisfied, if $\mathbf{A}$ and $\mathbf{C}$ are independent or $(\mathbf{A}_k, \mathbf{C}_k)_{k\in\mathbb{Z}}$ is an i.i.d. sequence.
**Proof:** Proceed along the lines of the proof of Proposition 4.9, but instead of the Hölder inequality use the independence, which gives $E(\|\mathbf{A}_0 \cdots \mathbf{A}_{-k+1}\mathbf{C}_{-k}\|^{\tilde{p}}) \leq E(\|\mathbf{A}_0 \cdots \mathbf{A}_{-k+1}\|^{\tilde{p}})$ $E(\|\mathbf{C}_0\|^{\tilde{p}})$. □
Just as the Lyapunov coefficient $\gamma$ formed by the sequence $(\mathbf{A}_t)$ could be replaced by $\tilde{\gamma}$, the one formed by the AR-part $(\mathbf{\Phi}_t)$, when discussing the existence of stationary solutions, the asymptotic moment conditions on $(\mathbf{A}_t)$ can be replaced by analogous conditions on $(\mathbf{\Phi}_t)$. This relation is strongest in $L^\infty$.

**Proposition 5.12** *Let* $\mathbf{A}_0 \in L^\infty$. *Then* $\lim_{k\to\infty} \|\mathbf{A}_0 \cdots \mathbf{A}_{-k+1}\|_{L^\infty(\Omega,\mathcal{F},P,M_{d(p+q)}(\mathbb{R}),\|\cdot\|)}^{1/k} < 1$, *iff* $\lim_{k\to\infty} \|\mathbf{\Phi}_0 \cdots \mathbf{\Phi}_{-k+1}\|_{L^\infty(\Omega,\mathcal{F},P,M_{dp}(\mathbb{R}),\|\cdot\|)}^{1/k} < 1$.

Note that in view of Lemma 4.5 it does not matter which algebra norms $\|\cdot\|$ are actually used on $M_{d(p+q)}$ and $M_{dp}$. Moreover, in view of (5.18) and (5.19) there is a $k \in \mathbb{N}$ such that $\mathbf{A}_0 \cdots \mathbf{A}_{-m} = 0$ a.s. $\forall m \geq k$, iff there is a $k' \in \mathbb{N}$ such that $\mathbf{\Phi}_0 \cdots \mathbf{\Phi}_{-m} = 0$ a.s. $\forall m \geq k'$. Hence, we presume in the proof of this and the following propositions w.l.o.g. that there is no $k \in \mathbb{N}$ such that $\mathbf{A}_0 \cdots \mathbf{A}_{-k} = 0$ a.s.
**Proof:** By Theorem 2.14 and Corollary 2.15 $\mathbf{\Phi}_0$ is in $L^\infty$. Furthermore, for all $k \in \mathbb{N}$ we have $\mathbf{A}_0 \cdots \mathbf{A}_{-k+1} \in L^\infty$ and $\mathbf{\Phi}_0 \cdots \mathbf{\Phi}_{-k+1} \in L^\infty$. (5.18) gives:

$$\begin{aligned}\lim_{n\to\infty} \|\mathbf{A}_0 \cdots \mathbf{A}_{-n+1}\|_{L^\infty}^{1/n} \;\leq\; & \lim_{n\to\infty} \left\| \begin{pmatrix} \mathbf{\Phi}_0 & 0 \\ 0 & 0 \end{pmatrix} \cdots \begin{pmatrix} \mathbf{\Phi}_{-n+q+2} & 0 \\ 0 & 0 \end{pmatrix} \right\|_{L^\infty}^{1/n} \\ & \times \underbrace{\|\mathbf{A}_{-n+q+1} \cdots \mathbf{A}_{-n+1}\|_{L^\infty}^{1/n}}_{\to 1}\end{aligned}$$

$$
\begin{aligned}
&= \lim_{n\to\infty} \left\| \begin{pmatrix} \mathbf{\Phi}_0 & 0 \\ 0 & 0 \end{pmatrix} \cdots \begin{pmatrix} \mathbf{\Phi}_{-n+q+2} & 0 \\ 0 & 0 \end{pmatrix} \right\|_{L^\infty}^{1/n} \\
&= \lim_{n\to\infty} \left\| \begin{pmatrix} \mathbf{\Phi}_0 & 0 \\ 0 & 0 \end{pmatrix} \cdots \begin{pmatrix} \mathbf{\Phi}_{-(n-q-1)+1} & 0 \\ 0 & 0 \end{pmatrix} \right\|_{L^\infty}^{1/(n-q-1)} \\
&= \lim_{n\to\infty} \left\| \begin{pmatrix} \mathbf{\Phi}_0 & 0 \\ 0 & 0 \end{pmatrix} \cdots \begin{pmatrix} \mathbf{\Phi}_{-n+1} & 0 \\ 0 & 0 \end{pmatrix} \right\|_{L^\infty}^{1/n}.
\end{aligned}
$$

Since an analogous argument using (5.19) shows the above inequality with "$\geq$" instead of "$\leq$", we get using Lemma 4.5 and the same arguments as in the proof of Proposition 5.5

$$\lim_{k\to\infty} \|\mathbf{A}_0 \cdots \mathbf{A}_{-k+1}\|_{L^\infty}^{1/k} = \lim_{k\to\infty} \|\mathbf{\Phi}_0 \cdots \mathbf{\Phi}_{-k+1}\|_{L^\infty}^{1/k}.$$

□

Literally the same proof can be used for general positive $\tilde{p}$ instead of $\infty$ given $(\mathbf{A}_t)$ is an i.i.d. sequence. In this case one obtains:

**Proposition 5.13** *Let $(\mathbf{A}_t)$ be i.i.d, $\tilde{p} \in (0,\infty)$ and $\mathbf{A}_0 \in L^{\tilde{p}}$.*
*Then $\lim_{k\to\infty} E\left(\|\mathbf{A}_0 \cdots \mathbf{A}_{-k+1}\|^{\tilde{p}}\right)^{1/k} < 1$, iff $\lim_{k\to\infty} E\left(\|\mathbf{\Phi}_0 \cdots \mathbf{\Phi}_{-k+1}\|^{\tilde{p}}\right)^{1/k} < 1$.*

For general Markov-switching models one has to employ the Hölder inequality in one direction. This leads to the need of stronger assumptions and we cannot show equivalence of $\limsup_{k\to\infty} E(\|\mathbf{A}_0 \cdots \mathbf{A}_{-k+1}\|^{\tilde{p}})^{1/k} < 1$ and $\limsup_{k\to\infty} E(\|\mathbf{\Phi}_0 \cdots \mathbf{\Phi}_{-k+1}\|^{\tilde{p}})^{1/k} < 1$ in general.

**Proposition 5.14** *Let $\tilde{p} \in (0,\infty)$ and $\mathbf{A}_0 \cdots \mathbf{A}_{-k+1} \in L^{\tilde{p}}$ for all natural $k$.*

*(i) If $\limsup_{k\to\infty} E(\|\mathbf{A}_0 \cdots \mathbf{A}_{-k+1}\|^{\tilde{p}})^{1/k} < 1$, then $\limsup_{k\to\infty} E(\|\mathbf{\Phi}_0 \cdots \mathbf{\Phi}_{-k+1}\|^{\tilde{p}})^{1/k} < 1$.*

*(ii) If there are $r, s \in [1,\infty]$ with $\mathbf{\Phi}_0 \cdots \mathbf{\Phi}_{-k+1} \in L^{\tilde{p}r}$ $\forall k \in \mathbb{N}$, $\mathbf{A}_0 \cdots \mathbf{A}_{-q} \in L^{\tilde{p}s}$ and $\limsup_{k\to\infty} E(\|\mathbf{\Phi}_0 \cdots \mathbf{\Phi}_{-k+1}\|^{\tilde{p}r})^{1/k} < 1$ for $\tilde{p}r < \infty$, respectively $\lim_{k\to\infty} \|\mathbf{\Phi}_0 \cdots \mathbf{\Phi}_{-k+1}\|_{L^\infty}^{1/k} < 1$ for $\tilde{p}r = \infty$, then $\limsup_{k\to\infty} E(\|\mathbf{A}_0 \cdots \mathbf{A}_{-k+1}\|^{\tilde{p}})^{1/k} < 1$.*

**Proof:** Corollary 2.15 and Lemma 5.3 ensure $\mathbf{\Phi}_0 \cdots \mathbf{\Phi}_{-k+1} \in L^{\tilde{p}}$ for all natural $k$. From Lemma 5.3 one moreover obtains for any operator norm:

$$\left\| \begin{pmatrix} \mathbf{\Phi}_0 \cdots \mathbf{\Phi}_{-k+1} & 0 \\ 0 & 0 \end{pmatrix} \right\| \leq \|\mathbf{A}_0 \cdots \mathbf{A}_{-k+1}\|.$$

Using an induced algebra norm as in the proof of Proposition 5.5 and the asymptotic independence from the norm as given in Lemma 4.5 this gives (i).

To prove (ii) choose a strictly increasing sequence $(k_n)_{n\in\mathbb{N}}$ such that

$$\lim_{n\to\infty} E(\|\mathbf{A}_0 \cdots \mathbf{A}_{-k_n+1}\|^{\tilde{p}})^{1/k_n} = \limsup_{k\to\infty} E(\|\mathbf{A}_0 \cdots \mathbf{A}_{-k+1}\|^{\tilde{p}})^{1/k}$$

and

$$\left\| \begin{pmatrix} \mathbf{\Phi}_0 & 0 \\ 0 & 0 \end{pmatrix} \cdots \begin{pmatrix} \mathbf{\Phi}_{-k_n+q+2} & 0 \\ 0 & 0 \end{pmatrix} \right\|_{L^{\tilde{p}r}}^{1/k_n}$$

converges in $\overline{\mathbb{R}}$. Using (5.18) and the Hölder inequality we get

$$\begin{aligned}
\lim_{n\to\infty} E(\|\mathbf{A}_0\cdots\mathbf{A}_{-k_n+1}\|^{\tilde{p}})^{1/k_n} &\le \limsup_{n\to\infty} E\left(\left\|\begin{pmatrix}\mathbf{\Phi}_0 & 0\\ 0 & 0\end{pmatrix}\cdots\begin{pmatrix}\mathbf{\Phi}_{-k_n+q+2} & 0\\ 0 & 0\end{pmatrix}\right\|^{\tilde{p}}\right.\\
&\qquad \left.\times\, \|\mathbf{A}_{-k_n+q+1}\cdots\mathbf{A}_{-k_n+1}\|^{\tilde{p}}\right)^{1/k_n}\\
&\le \lim_{n\to\infty}\left\|\begin{pmatrix}\mathbf{\Phi}_0 & 0\\ 0 & 0\end{pmatrix}\cdots\begin{pmatrix}\mathbf{\Phi}_{-k_n+q+2} & 0\\ 0 & 0\end{pmatrix}\right\|_{L^{\tilde{p}r}}^{\tilde{p}/k_n}\\
&\qquad \times \underbrace{\|\mathbf{A}_0\cdots\mathbf{A}_{-q}\|_{L^{\tilde{p}s}}^{\tilde{p}/k_n}}_{\to 1}\\
&= \lim_{n\to\infty}\left\|\begin{pmatrix}\mathbf{\Phi}_0 & 0\\ 0 & 0\end{pmatrix}\cdots\begin{pmatrix}\mathbf{\Phi}_{-(k_n-q-1)+1} & 0\\ 0 & 0\end{pmatrix}\right\|_{L^{\tilde{p}r}}^{\tilde{p}/(k_n-q-1)}\\
&\le \limsup_{k\to\infty}\|\mathbf{\Phi}_0\cdots\mathbf{\Phi}_{-k+1}\|_{L^{\tilde{p}r}}^{\tilde{p}/k} < 1
\end{aligned}$$

using once more Lemma 4.5 and the construction from the proof of Proposition 5.5. □

The following corollary shows that, if $(\mathbf{A}_t)$ has a bounded (finite) state space, then it makes generally no difference, whether one studies the asymptotic behaviour of $(\mathbf{A}_t)$ or $(\mathbf{\Phi}_t)$.

**Corollary 5.15** *Let $\tilde{p} \in (0,\infty)$ and $\mathbf{A}_0 \in L^\infty$. Then $\limsup_{k\to\infty} E(\|\mathbf{A}_0\cdots\mathbf{A}_{-k+1}\|^{\tilde{p}})^{1/k} < 1$, iff $\limsup_{k\to\infty} E(\|\mathbf{\Phi}_0\cdots\mathbf{\Phi}_{-k+1}\|^{\tilde{p}})^{1/k} < 1$.*

**Proof:** Obviously $\mathbf{A}_0\cdots\mathbf{A}_{-k+1} \in L^{\tilde{p}}$ for all natural $k$. Now the result follows immediately from Proposition 5.14, since $r = 1$ and $s = \infty$ can be chosen in (ii). □

The above theorems show that in many cases, especially those usually occurring in applications, only the long run behaviour of the autoregressive part matters in the conditions ensuring finiteness of some moments of an MS-ARMA process.

## 5.3 Feasible Sufficient Conditions

In the previous section we have given conditions ensuring the existence of stationary MS-ARMA processes and the finiteness of moments. However, whereas most of the conditions involving only $\mathbf{A}_0$, $\mathbf{\Phi}_0$ or $\mathbf{C}_0$ are straightforward to check, it is in general far from trivial to check the essential conditions like $\lim_{n\to\infty}\frac{1}{n+1}E(\log\|\mathbf{A}_0\cdots\mathbf{A}_{-n}\|) < 0$ or $\limsup_{n\to\infty} E(\|\mathbf{A}_0\cdots\mathbf{A}_{-n+1}\|^{\tilde{p}})^{1/n} < 1$. Thus, the aim of the present section is to study some rather easy to check conditions, which imply these complicated ones. Yet, it should be obvious that all of these simplifications may well fail to indicate the existence of a stationary solution to an MS-ARMA equation, although there actually is one. The next section will then address the question of whether there is a general relation between the stationarity properties of the individual regimes and those of the overall Markov switching process.

### 5.3.1 Norm Conditions

In 5.2.2 we already noted that $E(\log \|\mathbf{A}_0\|) < 0$, resp. $E(\log \|\mathbf{\Phi}_0\|) < 0$, for some algebra norm would ensure the strict negativity of the Lyapunov coefficient $\gamma$, resp. $\tilde{\gamma}$. Similarly we saw in the section on moment conditions that $\|\mathbf{A}_0\| < c$ a.s. for some $c < 1$ or $E(\|\mathbf{A}_0\|^s) < 1$ in the case of i.i.d. $(\mathbf{A}_t)$ would ensure the validity of critical conditions for the finiteness of moments. Yet, we already noted that all norms over $\mathbb{R}^d$ usually considered have the property that they assign the same "length" to all the canonical basis vectors and thus lead to $\mathbf{A}_0$ and $\mathbf{\Phi}_0$ having operator norm of at least one, if one has a MS-ARMA(p,q) model with $p$ or $q$ being greater than one (resp. as far as $\mathbf{\Phi}_0$ alone is regarded, if $p > 1$). Yet, the following theorem shows that norms can be given such that a given matrix is within the unit circle provided all of its eigenvalues are less than one in modulus. Since the spectrum is involved, we study complex matrices and automatically consider $M_d(\mathbb{R})$ as a subset of $M_d(\mathbb{C})$.

**Theorem 5.16 (Ciarlet (1989, Th. 1.4-3))** *Let $A$ be any matrix in $M_d(\mathbb{C})$ with spectral radius $\rho(A) < 1$. Then for every $\epsilon > 0$ there is a norm $\|\cdot\|_\epsilon$ on $\mathbb{C}^d$ such that $\|A\|_\epsilon < \rho(A) + \epsilon$ holds, where $\|\cdot\|_\epsilon$ also denotes the induced operator norm.*

Note that as pointed out in 5.2.2 for a classical ARMA model the stationarity condition is exactly that $\mathbf{A}_0$ has spectral radius smaller than one. Since the above Theorem shows that for any matrix with spectrum within the unit circle there is an operator norm such that the matrix has norm less than one, the classical ARMA condition is equivalent to demand that $E(\log \|\mathbf{A}_0\|) < 0$ for some operator norm.

Furthermore one should bear in mind that the definitions of $\mathbf{A}_t$ and $\mathbf{\Phi}_t$ immediately imply $\sigma(\mathbf{A}_t) = \sigma(\mathbf{\Phi}_t) \cup \{0\}$ and that $\mathbf{\Phi}_t$ is invertible, iff $\Phi_{pt}$ is invertible, whereas $\mathbf{A}_t$ is never invertible (for $q > 0$).

**Proof (cf. Ciarlet (1989, pp. 29f)):** For $A \in M_d(\mathbb{C})$ there is a unitary $U \in M_d(\mathbb{C})$ such that

$$U^{-1}AU = \begin{pmatrix} \lambda_1 & t_{1,2} & \cdots & t_{1,d} \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & t_{d-1,d} \\ 0 & \cdots & 0 & \lambda_d \end{pmatrix}$$

with $\sigma(A) = \{\lambda_1, \ldots \lambda_d\}$ (Schur decomposition). For some $\epsilon > 0$ choose $\delta > 0$ so small that $\sum_{j=i+1}^{d} |\delta^{j-i} t_{i,j}| < \epsilon$ for $1 \le i \le d-1$ and set

$$D_\delta = \begin{pmatrix} \delta & 0 & \cdots & 0 \\ 0 & \delta^2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \delta^d \end{pmatrix}.$$

A straightforward calculation gives

$$D_\delta^{-1}U^{-1}AUD_\delta = \begin{pmatrix} \lambda_1 & t_{1,2}\delta^{2-1} & \cdots & t_{1,d}\delta^{d-1} \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & t_{d-1,d}\delta^{d-(d-1)} \\ 0 & \cdots & 0 & \lambda_d \end{pmatrix} = (c_{i,j})_{1 \le i,j \le d}$$

with $c_{i,i} = \lambda_i$, $c_{i,j} = t_{i,j}\delta^{j-i}$ for $j > i$ and $c_{i,j} = 0$ for $j < i$. This gives

$$\|D_\delta^{-1}U^{-1}AUD_\delta\|_\infty = \max_{1\le i\le d}\left(|\lambda_i| + \sum_{j=i+1}^{d} |\delta^{j-i}t_{i,j}|\right) < \rho(A) + \epsilon$$

by the choice of $\delta$. Define $\|\cdot\|_A : M_d(\mathbb{C}) \to \mathbb{R}$, $B \mapsto \|(UD_\delta)^{-1}B(UD_\delta)\|_\infty$, then $\|\cdot\|_A$ is obviously an algebra norm over $M_d(\mathbb{R})$ with $\|A\|_A < \rho(A) + \epsilon$.

So it remains to show that $\|\cdot\|_A$ is an operator norm. To this end define a norm $\|\cdot\|_A$ over $\mathbb{C}^d$ via $\|x\|_A := \|(UD_\delta)^{-1}x\|_\infty$. We have for all $B \in M_d(\mathbb{C})$ and $x \in \mathbb{C}^d$

$$\begin{aligned}\|Bx\|_A &= \|(UD_\delta)^{-1}Bx\|_\infty = \|(UD_\delta)^{-1}B(UD_\delta)(UD_\delta)^{-1}x\|_\infty \\ &\le \|(UD_\delta)^{-1}B(UD_\delta)\|_\infty\|(UD_\delta)^{-1}x\|_\infty = \|B\|_A\|x\|_A.\end{aligned}$$

Moreover, for $B \in M_d(\mathbb{C})$ there is a $u_B \neq 0 \in \mathbb{C}^d$ such that $\|(UD_\delta)^{-1}BUD_\delta u_B\|_\infty = \|(UD_\delta)^{-1}BUD_\delta\|_\infty\|u_B\|_\infty$. Set $v_B = UD_\delta u_B$, then $v_B \neq 0$ and

$$\|Bv_B\|_A = \|(UD_\delta)^{-1}BUD_\delta u_B\|_\infty = \|(UD_\delta)^{-1}BUD_\delta\|_\infty\|u_B\|_\infty = \|B\|_A\|v_B\|_A.$$

So $\|\cdot\|_A$ is the operator norm induced on $M_d(\mathbb{C})$ by the norm $\|\cdot\|_A$ on $\mathbb{C}^d$. □

The above results suggest the following possibility to check the existence of a stationary solution to an MS-ARMA$(p, q, \Delta, \epsilon)$ equation provided the state space of $(\mathbf{A}_t)$, resp. $(\mathbf{\Phi}_t)$, contains a matrix with spectrum within the unit ball. Take one such matrix from the state space of $(\mathbf{A}_t)$, resp. $(\mathbf{\Phi}_t)$, and an operator norm $\|\cdot\|$ such that this matrix has norm less than one (e.g. the one explicitly constructed in the above proof), then check, if $E(\log\|\mathbf{A}_0\|) < 0$, resp. $E(\log\|\mathbf{\Phi}_0\|) < 0$, using this norm. In general there will be many matrices in the state space of $(\mathbf{A}_t)$, resp. $(\mathbf{\Phi}_t)$, that satisfy $\rho(\cdot) < 1$, since in applications one will choose most (though probably not all) regimes to be stationary ARMA regimes. In principle one can repeat the procedure for all of those regimes, until the crucial stationarity condition on $E(\log\|\mathbf{A}_0\|)$, resp. $E(\log\|\mathbf{\Phi}_0\|)$, is fulfilled, since the outcome will in general be different for the different norms, as the norm constructed in the above proof is highly dependent on the individual matrix. If $(\mathbf{A}_t)$ is an i.i.d. sequence one can proceed similarly, to find an operator norm ensuring $E(\|\mathbf{A}_0\|^s) < 1$. If $\mathbf{A}_0$, resp. $\mathbf{\Phi}_0$, has almost sure only eigenvalues of modulus less than one (i.e. all individual regimes are stationary), one may also try to find an operator norm giving $\|\mathbf{A}_0\| < c$, resp. $\|\mathbf{\Phi}_0\| < c$, a.s. for some $c < 1$. It is clear that the search for such a norm is bound to fail, if there is a positive probability for the spectral radius to be greater than or equal to one.

However, there is also a simple negative result regarding the search for an appropriate norm making it straightforward to show that the Lyapunov exponent is strictly negative: Suppose the state space of $\mathbf{A}_0$ contains two matrices $A^{(1)}$ and $A^{(2)}$ such that $\rho(A^{(1)}) < 1$ and $\rho(A^{(1)}) < 1$, but $\rho(A^{(1)}A^{(2)}) \ge 1$ (for a concrete example of such matrices see Section 5.4.2), then there obviously cannot be any algebra norm $\|\cdot\|_*$ such that $\|A^{(1)}\|_* < 1$ and $\|A^{(2)}\|_* \le 1$. Otherwise $\|A^{(1)}A^{(2)}\|_* < 1$ would be a contradiction to $\rho(A^{(1)}A^{(2)}) \ge 1$ and the classical result of Beurling and Gelfand (cf. Werner (2002, Satz IX.1.3(e))) that the spectral radii of the elements of a Banach algebra are less than or equal to their norm.

One may still regard the above construction of an operator norm ensuring that some matrix is within the unit ball as rather complicated. Indeed, we think that the high degree of dependence of the obtained norm on the individual matrix is a major drawback. On the other hand no information on the special structure of $\mathbf{A}_0$, resp. $\mathbf{\Phi}_0$, is used above. Thus, we will now focus on the possibility of constructing operator norms that ensure that a (possibly uncountable) set of matrices, which has this very special structure, has norm less than one. Yet, this can, at least as it seems to us, only be done under rather restrictive conditions. In the end, however, this will give a criterion for the stationarity of an MS-ARMA process that is equivalent to the best known general stationarity criterion for TAR-models (cf. An and Huang (1996), Brachner (2004)). Actually, our results are slightly more general, since the TAR literature is on $\mathbb{R}$-valued processes only. In our general vector valued case we obtain some additional freedom from the possibility of choosing different norms. To the best of our knowledge the criterion we give below is new in the context of MS-ARMA processes, as are the constructions of the specific norms we give in the following two theorems.

Since it is notationally less involved and illustrates the basic idea, we study matrices with the structure of $\mathbf{\Phi}_t$ first.

**Theorem 5.17** *Let $d, p$ be natural numbers and $\mathcal{A} \subset M_{dp}(\mathbb{R})$ a set of matrices such that for each $A \in \mathcal{A}$ there are $A_1(A), \ldots, A_p(A) \in M_d(\mathbb{R})$ with*

$$A = \begin{pmatrix} A_1(A) & \cdots & A_{(p-1)}(A) & A_p(A) \\ I_d & 0 \cdots & \cdots & 0 \\ 0 & \ddots & \ddots & \vdots \\ 0 & \cdots 0 & I_d & 0 \end{pmatrix}.$$

*Assume moreover that there is a norm $\|\cdot\|_d$ on $\mathbb{R}^d$ and a $c < 1$ such that*

$$\sup_{A\in\mathcal{A}} \sum_{i=1}^{p} \|A_i(A)\|_d < c$$

*holds for the induced operator norm. Then there is a norm $\|\cdot\|_p$ on $\mathbb{R}^{dp} = (\mathbb{R}^d)^p$ and a $c' < 1$ such that*

$$\sup_{A\in\mathcal{A}} \|A\|_p < c'$$

*in the induced operator norm.*

*Especially, $\|X_0X_1\cdots X_k\|_p < (c')^{k+1}$ for all natural $k$ and sequences $(X_n)_{n\in\mathbb{N}_0}$ with elements in $\mathcal{A}$.*

**Proof:** Choose $c_1, \ldots, c_p \in \mathbb{R}$ such that $1 = c_1 > c_2 > \ldots > c_p > c$. Then

$$\sup_{A\in\mathcal{A}} \sum_{i=1}^{p} \frac{\|A_i(A)\|_d}{c_i} \leq \sup_{A\in\mathcal{A}} \sum_{i=1}^{p} \frac{\|A_i(A)\|_d}{c_p} < \frac{c}{c_p} < 1.$$

Now define a norm $\|\cdot\|_p$ on $(\mathbb{R}^d)^p$ via $\|(x_1^\mathsf{T}, \ldots, x_p^\mathsf{T})^\mathsf{T}\|_p = \max\{c_1\|x_1\|_d, \ldots, c_p\|x_p\|_d\}$ and identify $B((\mathbb{R}^d)^p)$ with $M_{dp}(\mathbb{R})$. For some $(x_1^\mathsf{T}, \ldots, x_p^\mathsf{T})^\mathsf{T} \in \mathbb{R}^{dp}$ and $A \in \mathcal{A}$ we have

$$\left\| A \begin{pmatrix} x_1 \\ \vdots \\ x_p \end{pmatrix} \right\|_p = \left\| \begin{pmatrix} \sum_{i=1}^p A_i(A)x_i \\ x_1 \\ \vdots \\ x_{p-1} \end{pmatrix} \right\|_p$$

$$
\begin{aligned}
&= \max\left\{c_1\left\|\sum_{i=1}^{p} A_i(A)x_i\right\|_d, c_2\|x_1\|_d, \ldots, c_p\|x_{p-1}\|_d\right\} \\
&= \max\left\{\left\|\sum_{i=1}^{p} A_i(A)x_i\right\|_d, \frac{c_2}{c_1}c_1\|x_1\|_d, \ldots, \frac{c_p}{c_{p-1}}c_{p-1}\|x_{p-1}\|_d\right\} \\
&\leq \max\left\{\left\|\sum_{i=1}^{p} A_i(A)x_i\right\|_d, \max_{2\leq k\leq p}\left\{\frac{c_k}{c_{k-1}}\right\}\|(x_1^\mathsf{T}, \ldots, x_p^\mathsf{T})^\mathsf{T}\|_p\right\}
\end{aligned}
$$

and, moreover,

$$
\begin{aligned}
\left\|\sum_{i=1}^{p} A_i(A)x_i\right\|_d &\leq \sum_{i=1}^{p}\|A_i(A)\|_d\|x_i\|_d \leq \max\{c_1\|x_1\|_d, \ldots, c_p\|x_p\|_d\}\sum_{i=1}^{p}\frac{\|A_i(A)\|_d}{c_i} \\
&\leq \left\|(x_1^\mathsf{T}, \ldots, x_p^\mathsf{T})^\mathsf{T}\right\|_p\sum_{i=1}^{p}\frac{\|A_i(A)\|_d}{c_p}.
\end{aligned}
$$

From this one deduces

$$
\begin{aligned}
\sup_{A\in\mathcal{A}}\|A\|_p &\leq \max\left\{\sup_{A\in\mathcal{A}}\sum_{i=1}^{p}\frac{\|A_i(A)\|_d}{c_p}, \max_{2\leq k\leq p}\left\{\frac{c_k}{c_{k-1}}\right\}\right\} \\
&\leq \max\left\{\frac{c}{c_p}, \max_{2\leq k\leq p}\left\{\frac{c_k}{c_{k-1}}\right\}\right\} =: c' < 1
\end{aligned}
$$

which concludes the proof. □

Under an additional assumption the above theorem can be generalized to matrices with the structure of $\mathbf{A}_t$.

**Theorem 5.18** *Let $d, p, q$ be natural numbers and $\mathcal{A} \subset M_{d(p+q)}(\mathbb{R})$ a set of matrices such that for each $A \in \mathcal{A}$ there are $A_1(A), \ldots, A_p(A), B_1(A), \ldots, B_q(A) \in M_d(\mathbb{R})$ with*

$$
A = \begin{pmatrix}
A_1(A) & \cdots & A_{p-1}(A) & A_p(A) & B_1(A) & \cdots & B_{q-1}(A) & B_q(A) \\
I_d & 0\cdots & \cdots & 0 & 0 & \cdots & \cdots & 0 \\
0 & \ddots & \ddots & 0 & 0 & \cdots & \cdots & \vdots \\
0 & \cdots 0 & I_d & 0 & 0 & \cdots & \cdots & \vdots \\
0 & \cdots & \cdots & 0 & 0 & \cdots & \cdots & 0 \\
0 & \cdots & \cdots & 0 & I_d & 0 & \cdots & 0 \\
\vdots & \ddots & \ddots & \vdots & 0 & \ddots & 0\cdots & \vdots \\
0 & \cdots & \cdots & 0 & 0 & \cdots 0 & I_d & 0
\end{pmatrix}.
$$

*Assume moreover that there is a norm $\|\cdot\|_d$ on $\mathbb{R}^d$ and a $c < 1$ such that*

$$
\sup_{A\in\mathcal{A}}\sum_{i=1}^{p}\|A_i(A)\|_d < c \text{ and } \sup_{A\in\mathcal{A}}\sum_{i=1}^{q}\|B_i(A)\|_d < \infty
$$

*hold for the induced operator norm. Then there is a norm* $\|\cdot\|_p$ *on* $\mathbb{R}^{d(p+q)} = (\mathbb{R}^d)^{(p+q)}$ *and a* $c' < 1$ *such that*

$$\sup_{A \in \mathcal{A}} \|A\|_p < c'$$

*in the induced operator norm.*

*Especially,* $\|X_0 X_1 \cdots X_k\|_p < (c')^{k+1}$ *for all natural* $k$ *and sequences* $(X_n)_{n\in\mathbb{N}_0}$ *with elements in* $\mathcal{A}$.

**Proof:** Choose $c_1, \ldots, c_p \in \mathbb{R}$ such that $1 = c_1 > c_2 > \ldots > c_p > c$. Then

$$\sup_{A\in\mathcal{A}} \sum_{i=1}^{p} \frac{\|A_i(A)\|_d}{c_i} \leq \sup_{A\in\mathcal{A}} \sum_{i=1}^{p} \frac{\|A_i(A)\|_d}{c_p} < \frac{c}{c_p} < 1.$$

So, choose moreover $M \in (c/c_p, 1)$ and $\tilde{c} \in \mathbb{R}^+$ such that

$$\sup_{A\in\mathcal{A}} \sum_{i=1}^{p} \frac{\|A_i(A)\|_d}{c_p} + \sup_{A\in\mathcal{A}} \sum_{i=1}^{q} \frac{\|B_i(A)\|_d}{\tilde{c}} < M < 1$$

and $c_{p+1}, \ldots, c_{p+q} \in \mathbb{R}$ with $c_{p+1} > \ldots > c_{p+q} > \tilde{c}$. Now define a norm $\|\cdot\|_p$ on $(\mathbb{R}^d)^{p+q}$ via

$$\|(x_1^\mathsf{T}, \ldots, x_p^\mathsf{T}, y_1^\mathsf{T}, \ldots, y_q^\mathsf{T})^\mathsf{T}\|_p = \max\{c_1\|x_1\|_d, \ldots, c_p\|x_p\|_d, c_{p+1}\|y_1\|_d, \ldots, c_{p+q}\|y_q\|_d\}$$

and identify $B((\mathbb{R}^d)^{p+q})$ with $M_{d(p+q)}(\mathbb{R})$. For some $(x_1^\mathsf{T}, \ldots, x_p^\mathsf{T}, y_1^\mathsf{T}, \ldots, y_q^\mathsf{T})^\mathsf{T} \in \mathbb{R}^{d(p+q)}$ and $A \in \mathcal{A}$ we have

$$\begin{aligned}
\left\| A \begin{pmatrix} x_1 \\ \vdots \\ x_p \\ y_1 \\ \vdots \\ y_q \end{pmatrix} \right\|_p &= \left\| \begin{pmatrix} \sum_{i=1}^p A_i(A)x_i + \sum_{i=1}^q B_i(A)y_i \\ x_1 \\ \vdots \\ x_{p-1} \\ 0 \\ y_1 \\ \vdots \\ y_{q-1} \end{pmatrix} \right\|_p \\
&= \max\left\{ c_1 \left\| \sum_{i=1}^p A_i(A)x_i + \sum_{i=1}^q B_i(A)y_i \right\|_d, c_2\|x_1\|_d, \ldots, c_p\|x_{p-1}\|_d, \right. \\
&\quad \left. 0, c_{p+2}\|y_1\|_d, \ldots, c_{p+q}\|y_{q-1}\|_d \right\} \\
&= \max\left\{ \left\| \sum_{i=1}^p A_i(A)x_i + \sum_{i=1}^q B_i(A)y_i \right\|_d, \frac{c_2}{c_1}c_1\|x_1\|_d, \ldots, \frac{c_p}{c_{p-1}}c_{p-1}\|x_{p-1}\|_d, \right. \\
&\quad \left. 0, \frac{c_{p+2}}{c_{p+1}}c_{p+1}\|y_1\|_d, \ldots, \frac{c_{p+q}}{c_{p+q-1}}c_{p+q-1}\|y_{q-1}\|_d \right\} \\
&\leq \max\left\{ \left\| \sum_{i=1}^p A_i(A)x_i + \sum_{i=1}^q B_i(A)y_i \right\|_d, \right. \\
&\quad \left. \max_{2\leq k\leq p+q, k\neq p+1} \left\{ \frac{c_k}{c_{k-1}} \right\} \|(x_1^\mathsf{T}, \ldots, x_p^\mathsf{T}, y_1^\mathsf{T}, \ldots, y_q^\mathsf{T})^\mathsf{T}\|_p \right\}
\end{aligned}$$

and, moreover,

$$\begin{aligned}
\left\| \sum_{i=1}^{p} A_i(A)x_i + \sum_{i=1}^{q} B_i(A)y_i \right\|_d &\leq \sum_{i=1}^{p} \|A_i(A)\|_d \|x_i\|_d + \sum_{i=1}^{p} \|B_i(A)\|_d \|y_i\|_d \\
&\leq \left( \sum_{i=1}^{p} \frac{\|A_i(A)\|_d}{c_i} + \sum_{i=1}^{q} \frac{\|B_i(A)\|_d}{c_{p+i}} \right) \\
&\quad \times \max\left\{c_1\|x_1\|_d, \ldots, c_p\|x_p\|_d, c_{p+1}\|y_1\|_d, \ldots, c_{p+q}\|y_q\|_d\right\} \\
&\leq \left( \sum_{i=1}^{p} \frac{\|A_i(A)\|_d}{c_i} + \sum_{i=1}^{q} \frac{\|B_i(A)\|_d}{c_{p+i}} \right) \\
&\quad \times \left\|(x_1^\mathsf{T}, \ldots, x_p^\mathsf{T}, y_1^\mathsf{T}, \ldots, y_q^\mathsf{T})^\mathsf{T}\right\|_p .
\end{aligned}$$

From this one deduces

$$\begin{aligned}
\sup_{A\in\mathcal{A}} \|A\|_p &\leq \max\left\{ \sup_{A\in\mathcal{A}} \sum_{i=1}^{p} \frac{\|A_i(A)\|_d}{c_p} + \sup_{A\in\mathcal{A}} \sum_{i=1}^{q} \frac{\|B_i(A)\|_d}{\tilde{c}}, \max_{2\leq k\leq p+q, k\neq p+1} \left\{ \frac{c_k}{c_{k-1}} \right\} \right\} \\
&\leq \max\left\{ M, \max_{2\leq k\leq p+q, k\neq p+1} \left\{ \frac{c_k}{c_{k-1}} \right\} \right\} =: c' < 1
\end{aligned}$$

which concludes the proof. □

It is now straightforward to deduce conditions for the strict negativity of the Lyapunov coefficient and the existence of moments of Markov-switching models.

**Corollary 5.19** *Assume there is a* $\bar{c} < 1$ *and a norm* $\|\cdot\|$ *on* $\mathbb{R}^d$ *such that* $\sum_{i=1}^{p} \|\Phi_{i0}\| \leq \bar{c}$ *a.s. Then* $\tilde{\gamma} < 0$, $\mathbf{\Phi}_0 \in L^\infty$, $\lim_{k\to\infty} \|\mathbf{\Phi}_0 \cdots \mathbf{\Phi}_{-k+1}\|_{L^\infty}^{1/k} < 1$ *and* $\limsup_{k\to\infty} E(\|\mathbf{\Phi}_0 \cdots \mathbf{\Phi}_{-k+1}\|^s)^{1/k} < 1$ *for all* $s \in \mathbb{R}^+$.

**Proof:** Apply Theorem 5.17 on the subset $\mathcal{A} = \{\mathbf{\Phi}_0 : \sum_{i=1}^{p} \|\Phi_{i0}\| \leq \bar{c}\}$ of the state space of $\mathbf{\Phi}_0$ to obtain an operator norm $\|\cdot\|$ which ensures $\|\mathbf{\Phi}_0\| < c'$ a.s. for some $c' < 1$. This immediately implies the above claims. □

An analogous result for $(\mathbf{A}_t)$ follows from Theorem 5.18:

**Corollary 5.20** *Assume that there is a* $\bar{c} < 1$, $M \in \mathbb{R}^+$ *and a norm* $\|\cdot\|$ *on* $\mathbb{R}^d$ *such that* $\sum_{i=1}^{p} \|\Phi_{i0}\| \leq \bar{c}$ *and* $\sum_{i=1}^{q} \|\Theta_{i0}\| \leq M$ *a.s. Then it holds that* $\gamma < 0$, $\mathbf{A}_0 \in L^\infty$, $\lim_{k\to\infty} \|\mathbf{A}_0 \cdots \mathbf{A}_{-k+1}\|_{L^\infty}^{1/k} < 1$ *and* $\limsup_{k\to\infty} E(\|\mathbf{A}_0 \cdots \mathbf{A}_{-k+1}\|^s)^{1/k} < 1$ *for all* $s \in \mathbb{R}^+$.

Recalling Theorem 5.8, Proposition 5.9 and the comment thereafter leads to the following theorem giving feasible sufficient conditions for the existence of a stationary solution of an MS-ARMA equation and moments thereof. The condition on $\sum_{i=1}^{p} \|\Phi_{i0}\|$ corresponds to the general stationarity condition for TAR models given in e.g. An and Huang (1996) and Brachner (2004), as promised above. The additional condition on $\sum_{i=1}^{p} \|\Theta_{i0}\|$ is only necessary because we in general consider models that may have a moving average component, whereas TAR models are purely autoregressive ones.

**Theorem 5.21** *Assume that there is a $\bar{c} < 1$, $M \in \mathbb{R}^+$ and a norm $\|\cdot\|$ on $\mathbb{R}^d$ such that $\sum_{i=1}^p \|\Phi_{i0}\| \leq \bar{c}$ and $\sum_{i=1}^q \|\Theta_{i0}\| \leq M$ a.s. Let, moreover, $E(\log^+ \|\mathbf{C}_0\|)$ be finite. Then all conditions of Theorem 5.2 or 5.7 are satisfied and thus there is a unique stationary and ergodic solution $X = (X_t)$ to the MS-ARMA$(p, q, \Delta, \epsilon)$ equation (5.3).*

*If $\mathbf{C}_0 \in L^{\tilde{p}}$ for some $\tilde{p} \in (0, \infty]$, then the solution $X_t$ of the MS-ARMA equation (5.3) and its higher dimensional representation $\mathbf{X}_t$ are in $L^{\tilde{p}}$. Moreover, the series defining $\mathbf{X}_t$ (as given in Theorem 5.2 or 5.7) converges in $L^{\tilde{p}}$.*

Although the above results on norms seem to be tailor-made to obtain Theorem 5.21, they can also be of use, if $\sum_{i=1}^p \|\Phi_{i0}\| \leq \bar{c}$ with $\bar{c} < 1$ does not hold almost sure. However, assume the latter condition is satisfied for a subset $\mathcal{A}$ of the state space of $\mathbf{A}_t$, resp. $\mathbf{\Phi}_t$, that has a positive probability. Then one can use the above results to construct a norm such that $\|\mathbf{A}_t\| < c$, resp. $\|\mathbf{\Phi}_t\| < c$, for some $c < 1$ on this subset and calculate e.g. $E(\log \|\mathbf{A}_0\|)$, resp. $E(\log \|\mathbf{\Phi}_0\|)$, for this norm, since there now clearly is some hope that the latter are less than zero, especially if the probability of $\mathcal{A}$ is close to one.

### 5.3.2 A Spectral Radius Condition

Now we turn to studying conditions ensuring stationarity and finiteness of moments of MS-ARMA processes that presume a finite state space of the driving Markov chain $\Delta$. Note that in the literature (apart from the ML-estimator discussion in Douc, Moulines and Rydén (2004)) only this finite state space case has been discussed and used for modelling purposes until now. The spectral radius condition we give below and some similar conditions have to the best of our knowledge first appeared in Karlsen (1990a) and have been heavily used in several papers (e.g. Zhang and Stine (2001), Francq and Zakoïan (2001), Francq and Zakoïan (2002)) to obtain wide sense stationarity, i.e. the expected value, variance and covariances are independent of time (see e.g. Brockwell and Davis (1991, Definition 1.3.2)), of an MS-ARMA$(p, q, \Delta, \epsilon)$ process. The only work employing such a condition to ensure (strict) stationarity seems to be Yao (2001), who studies MS-AR(p) processes driven by a finite state space Markov chain. For non-linear Markov switching autoregressions somewhat similar conditions are used in Francq and Roussignol (1998) and Yao and Attali (2000). Restricted to linear autoregression their approach means replacing the tensor products appearing below by norms.

For the remainder of this section we presume that the stationary and ergodic Markov chain $\Delta$ has $l$ possible states $\Delta^{(i)}$ given by $\Delta^{(i)} = (\mu^{(i)}, \Sigma^{(i)}, \Phi_1^{(i)}, \ldots, \Phi_p^{(i)}, \Theta_1^{(i)}, \ldots, \Theta_q^{(i)})$ for $i = 1, 2, \ldots, l$. Moreover, the stationary distribution of $\Delta$ is called $\pi$ and in the usual notation for a finite state space Markov chain we write $\pi = (\pi^{(1)}, \pi^{(2)}, \ldots, \pi^{(l)})$ and $P = (p_{ij})_{1 \leq i,j \leq l}$ with $p_{ij} = P(\Delta_t = \Delta^{(j)} | \Delta_{t-1} = \Delta^{(i)})$ for the transition matrix. Furthermore, we define

$$\mathbf{\Phi}^{(i)} = \begin{pmatrix} \Phi_1^{(i)} & \cdots & \Phi_{p-1}^{(i)} & \Phi_p^{(i)} \\ I_d & 0 \cdots & \cdots & 0 \\ 0 & \ddots & \ddots & \vdots \\ 0 & \cdots 0 & I_d & 0 \end{pmatrix} \in M_{dp}(\mathbb{R}), \tag{5.25}$$

$$\mathbf{A}^{(i)} = \begin{pmatrix} \Phi_1^{(i)} & \cdots & \Phi_{p-1}^{(i)} & \Phi_p^{(i)} & \Theta_1^{(i)} & \cdots & \Theta_{q-1}^{(i)} & \Theta_q^{(i)} \\ I_d & 0\cdots & \cdots & 0 & 0 & \cdots & \cdots & 0 \\ 0 & \ddots & \ddots & 0 & 0 & \cdots & \cdots & \vdots \\ 0 & \cdots 0 & I_d & 0 & 0 & \cdots & \cdots & \vdots \\ 0 & \cdots & \cdots & 0 & 0 & \cdots & \cdots & 0 \\ 0 & \cdots & \cdots & 0 & I_d & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \vdots & 0 & \ddots & 0\cdots & \vdots \\ 0 & \cdots & \cdots & 0 & 0 & \cdots 0 & I_d & 0 \end{pmatrix}, \tag{5.26}$$

$$\mathbf{\Sigma}^{(i)} = \left(\Sigma^{(i)\mathsf{T}}, 0, \ldots, 0, \Sigma^{(i)\mathsf{T}}, 0, \ldots, 0\right)^{\mathsf{T}} \in M_{d(p+q),d}(\mathbb{R}) \tag{5.27}$$

and

$$\mathbf{m}^{(i)} = \left(\mu^{(i)\mathsf{T}}, 0, \ldots, 0\right)^{\mathsf{T}} \in \mathbb{R}^{d(p+q)}. \tag{5.28}$$

We also need some more notions from standard matrix algebra. In the following $\otimes$ denotes the tensor (Kronecker) product of matrices, $tr(\cdot)$ the trace of a square matrix and $vec(\cdot)$ the vectorized form of a matrix (recall $vec\left((a_{ij})_{1\le i\le m, 1\le j\le n}\right) = (a_{11}, a_{21}, \ldots, a_{m1}, a_{12}, \ldots, a_{mn})^{\mathsf{T}}$, cf. e.g. Nicholls and Quinn (1982, p. 11)). Observe that we have $tr \in B(M_d(\mathbb{R}), \mathbb{R})$, respectively $B(M_d(\mathbb{C}), \mathbb{C})$, and $vec \in B(M_{m,n}(\mathbb{R}), \mathbb{R}^{mn})$, respectively $B(M_{m,n}(\mathbb{C}), \mathbb{C}^{mn})$. $vec$ is actually a topological isomorphism (regardless of the actual norms used). Moreover, one obtains a scalar product over $M_d(\mathbb{R})$ $(M_d(\mathbb{C}))$ by setting $\langle A, B\rangle = tr(AB^{\mathsf{T}})$ for $A, B \in M_d(\mathbb{R})$ $(M_d(\mathbb{C}))$. The norm induced by this scalar product (sometimes called Froebenius norm) is an algebra, but not an operator norm (see Heuser (1992, pp. 126, 128)) and will be denoted by $\|\cdot\|_t$ in the following (note $\|(a_{ij})\|_t^2 = \sum_{i,j} |a_{ij}|^2$). Since we finally look at the spectral radius, we again simply regard the real matrices as a subset of the complex ones.

Via a technical lemma we shall now show that the Lyapunov coefficient $\gamma$ is strictly negative, if

$$\begin{aligned} Q_{\mathbf{A}} &:= \begin{pmatrix} \mathbf{A}^{(1)} \otimes \mathbf{A}^{(1)} & 0 & \cdots & 0 \\ 0 & \mathbf{A}^{(2)} \otimes \mathbf{A}^{(2)} & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \mathbf{A}^{(l)} \otimes \mathbf{A}^{(l)} \end{pmatrix} \left(P^{\mathsf{T}} \otimes I_{(d(p+q))^2}\right) \\ &= \begin{pmatrix} p_{11}\mathbf{A}^{(1)} \otimes \mathbf{A}^{(1)} & p_{21}\mathbf{A}^{(1)} \otimes \mathbf{A}^{(1)} & \cdots & p_{l1}\mathbf{A}^{(1)} \otimes \mathbf{A}^{(1)} \\ p_{12}\mathbf{A}^{(2)} \otimes \mathbf{A}^{(2)} & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & p_{l(l-1)}\mathbf{A}^{(l-1)} \otimes \mathbf{A}^{(l-1)} \\ p_{1l}\mathbf{A}^{(l)} \otimes \mathbf{A}^{(l)} & \cdots & p_{(l-1)l}\mathbf{A}^{(l)} \otimes \mathbf{A}^{(l)} & p_{ll}\mathbf{A}^{(l)} \otimes \mathbf{A}^{(l)} \end{pmatrix} \end{aligned} \tag{5.29}$$

or

$$Q_{\mathbf{\Phi}} := \begin{pmatrix} \mathbf{\Phi}^{(1)} \otimes \mathbf{\Phi}^{(1)} & 0 & \cdots & 0 \\ 0 & \mathbf{\Phi}^{(2)} \otimes \mathbf{\Phi}^{(2)} & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \mathbf{\Phi}^{(l)} \otimes \mathbf{\Phi}^{(l)} \end{pmatrix} \left(P^{\mathsf{T}} \otimes I_{(dp)^2}\right) \tag{5.30}$$

$$= \begin{pmatrix} p_{11}\mathbf{\Phi}^{(1)} \otimes \mathbf{\Phi}^{(1)} & p_{21}\mathbf{\Phi}^{(1)} \otimes \mathbf{\Phi}^{(1)} & \cdots & p_{l1}\mathbf{\Phi}^{(1)} \otimes \mathbf{\Phi}^{(1)} \\ p_{12}\mathbf{\Phi}^{(2)} \otimes \mathbf{\Phi}^{(2)} & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & p_{l(l-1)}\mathbf{\Phi}^{(l-1)} \otimes \mathbf{\Phi}^{(l-1)} \\ p_{1l}\mathbf{\Phi}^{(l)} \otimes \mathbf{\Phi}^{(l)} & \cdots & p_{(l-1)l}\mathbf{\Phi}^{(l)} \otimes \mathbf{\Phi}^{(l)} & p_{ll}\mathbf{\Phi}^{(l)} \otimes \mathbf{\Phi}^{(l)} \end{pmatrix}$$

has spectral radius less than one. Furthermore, we will show that this also implies that the MS-ARMA process is square integrable, if $\epsilon_t$ is so.

**Lemma 5.22** *Set* $\mathbb{I}_{\mathbf{A}} = \left(I_{(d(p+q))^2}, I_{(d(p+q))^2}, \ldots, I_{(d(p+q))^2}\right) \in M_{(d(p+q))^2, l(d(p+q))^2}$ *and* $\mathbb{I}_{\mathbf{\Phi}} = \left(I_{(dp)^2}, I_{(dp)^2}, \ldots, I_{(dp)^2}\right) \in M_{(dp)^2, l(dp)^2}$. *Then the following identities hold for all* $t \in \mathbb{Z}$ *and* $k \in \mathbb{N}_0$*:*

$$\begin{pmatrix} E\left(vec\left(\mathbf{A}_t\mathbf{A}_{t-1}\cdots\mathbf{A}_{t-k}\mathbf{A}_{t-k}^\top\cdots\mathbf{A}_{t-1}^\top\mathbf{A}_t^\top\right) I_{\{\Delta^{(1)}\}}(\Delta_t)\right) \\ E\left(vec\left(\mathbf{A}_t\mathbf{A}_{t-1}\cdots\mathbf{A}_{t-k}\mathbf{A}_{t-k}^\top\cdots\mathbf{A}_{t-1}^\top\mathbf{A}_t^\top\right) I_{\{\Delta^{(2)}\}}(\Delta_t)\right) \\ \vdots \\ E\left(vec\left(\mathbf{A}_t\mathbf{A}_{t-1}\cdots\mathbf{A}_{t-k}\mathbf{A}_{t-k}^\top\cdots\mathbf{A}_{t-1}^\top\mathbf{A}_t^\top\right) I_{\{\Delta^{(l)}\}}(\Delta_t)\right) \end{pmatrix} = Q_{\mathbf{A}}^k \begin{pmatrix} \pi^{(1)} vec\left(\mathbf{A}^{(1)}{\mathbf{A}^{(1)}}^\top\right) \\ \pi^{(2)} vec\left(\mathbf{A}^{(2)}{\mathbf{A}^{(2)}}^\top\right) \\ \vdots \\ \pi^{(l)} vec\left(\mathbf{A}^{(l)}{\mathbf{A}^{(l)}}^\top\right) \end{pmatrix}, \tag{5.31}$$

$$\begin{pmatrix} E\left(vec\left(\mathbf{\Phi}_t\mathbf{\Phi}_{t-1}\cdots\mathbf{\Phi}_{t-k}\mathbf{\Phi}_{t-k}^\top\cdots\mathbf{\Phi}_{t-1}^\top\mathbf{\Phi}_t^\top\right) I_{\{\Delta^{(1)}\}}(\Delta_t)\right) \\ E\left(vec\left(\mathbf{\Phi}_t\mathbf{\Phi}_{t-1}\cdots\mathbf{\Phi}_{t-k}\mathbf{\Phi}_{t-k}^\top\cdots\mathbf{\Phi}_{t-1}^\top\mathbf{\Phi}_t^\top\right) I_{\{\Delta^{(2)}\}}(\Delta_t)\right) \\ \vdots \\ E\left(vec\left(\mathbf{\Phi}_t\mathbf{\Phi}_{t-1}\cdots\mathbf{\Phi}_{t-k}\mathbf{\Phi}_{t-k}^\top\cdots\mathbf{\Phi}_{t-1}^\top\mathbf{\Phi}_t^\top\right) I_{\{\Delta^{(l)}\}}(\Delta_t)\right) \end{pmatrix} = Q_{\mathbf{\Phi}}^k \begin{pmatrix} \pi^{(1)} vec\left(\mathbf{\Phi}^{(1)}{\mathbf{\Phi}^{(1)}}^\top\right) \\ \pi^{(2)} vec\left(\mathbf{\Phi}^{(2)}{\mathbf{\Phi}^{(2)}}^\top\right) \\ \vdots \\ \pi^{(l)} vec\left(\mathbf{\Phi}^{(l)}{\mathbf{\Phi}^{(l)}}^\top\right) \end{pmatrix}, \tag{5.32}$$

$$E\left(vec\left(\mathbf{A}_t\mathbf{A}_{t-1}\cdots\mathbf{A}_{t-k}\mathbf{A}_{t-k}^\top\cdots\mathbf{A}_{t-1}^\top\mathbf{A}_t^\top\right)\right) = \mathbb{I}_{\mathbf{A}} Q_{\mathbf{A}}^k \begin{pmatrix} \pi^{(1)} vec\left(\mathbf{A}^{(1)}{\mathbf{A}^{(1)}}^\top\right) \\ \pi^{(2)} vec\left(\mathbf{A}^{(2)}{\mathbf{A}^{(2)}}^\top\right) \\ \vdots \\ \pi^{(l)} vec\left(\mathbf{A}^{(l)}{\mathbf{A}^{(l)}}^\top\right) \end{pmatrix} \tag{5.33}$$

*and*

$$E\left(vec\left(\mathbf{\Phi}_t\mathbf{\Phi}_{t-1}\cdots\mathbf{\Phi}_{t-k}\mathbf{\Phi}_{t-k}^\top\cdots\mathbf{\Phi}_{t-1}^\top\mathbf{\Phi}_t^\top\right)\right) = \mathbb{I}_{\mathbf{\Phi}} Q_{\mathbf{\Phi}}^k \begin{pmatrix} \pi^{(1)} vec\left(\mathbf{\Phi}^{(1)}{\mathbf{\Phi}^{(1)}}^\top\right) \\ \pi^{(2)} vec\left(\mathbf{\Phi}^{(2)}{\mathbf{\Phi}^{(2)}}^\top\right) \\ \vdots \\ \pi^{(l)} vec\left(\mathbf{\Phi}^{(l)}{\mathbf{\Phi}^{(l)}}^\top\right) \end{pmatrix}. \tag{5.34}$$

**Proof:** We only prove the claims regarding $\mathbf{A}$, since the results on $\mathbf{\Phi}$ are obtained along the same lines. (5.31) and the definition of $\mathbb{I}_{\mathbf{A}}$ immediately give (5.33):

$$
\begin{aligned}
&\mathbb{I}_{\mathbf{A}} Q_{\mathbf{A}}^k \begin{pmatrix} \pi^{(1)} vec\left(\mathbf{A}^{(1)} {\mathbf{A}^{(1)}}^{\mathsf{T}}\right) \\ \pi^{(2)} vec\left(\mathbf{A}^{(2)} {\mathbf{A}^{(2)}}^{\mathsf{T}}\right) \\ \vdots \\ \pi^{(l)} vec\left(\mathbf{A}^{(l)} {\mathbf{A}^{(l)}}^{\mathsf{T}}\right) \end{pmatrix} \\
= \;& \sum_{i=1}^{l} I_{(d(p+q))^2} E\left(vec\left(\mathbf{A}_t \mathbf{A}_{t-1} \cdots \mathbf{A}_{t-k} \mathbf{A}_{t-k}^{\mathsf{T}} \cdots \mathbf{A}_{t-1}^{\mathsf{T}} \mathbf{A}_t^{\mathsf{T}}\right) I_{\{\Delta^{(i)}\}}(\Delta_t)\right) \\
= \;& E\left(vec\left(\mathbf{A}_t \mathbf{A}_{t-1} \cdots \mathbf{A}_{t-k} \mathbf{A}_{t-k}^{\mathsf{T}} \cdots \mathbf{A}_{t-1}^{\mathsf{T}} \mathbf{A}_t^{\mathsf{T}}\right) \sum_{i=1}^{l} I_{\{\Delta^{(i)}\}}(\Delta_t)\right) \\
= \;& E\left(vec\left(\mathbf{A}_t \mathbf{A}_{t-1} \cdots \mathbf{A}_{t-k} \mathbf{A}_{t-k}^{\mathsf{T}} \cdots \mathbf{A}_{t-1}^{\mathsf{T}} \mathbf{A}_t^{\mathsf{T}}\right)\right).
\end{aligned}
$$

Obviously (5.31) holds for $k = 0$ and else is equivalent to

$$
\begin{aligned}
&Q_{\mathbf{A}}^k \begin{pmatrix} \pi^{(1)} vec\left(\mathbf{A}^{(1)} {\mathbf{A}^{(1)}}^{\mathsf{T}}\right) \\ \pi^{(2)} vec\left(\mathbf{A}^{(2)} {\mathbf{A}^{(2)}}^{\mathsf{T}}\right) \\ \vdots \\ \pi^{(l)} vec\left(\mathbf{A}^{(l)} {\mathbf{A}^{(l)}}^{\mathsf{T}}\right) \end{pmatrix} \\
= &\begin{pmatrix} \sum\limits_{i_1, i_2, \ldots, i_k = 1}^{l} \pi^{(i_k)} p_{i_k i_{k-1}} p_{i_{k-1} i_{k-2}} \cdots p_{i_1 1} vec\left(\mathbf{A}^{(1)} \mathbf{A}^{(i_1)} \ldots \mathbf{A}^{(i_k)} {\mathbf{A}^{(i_k)}}^{\mathsf{T}} \ldots {\mathbf{A}^{(i_1)}}^{\mathsf{T}} {\mathbf{A}^{(1)}}^{\mathsf{T}}\right) \\ \sum\limits_{i_1, i_2, \ldots, i_k = 1}^{l} \pi^{(i_k)} p_{i_k i_{k-1}} p_{i_{k-1} i_{k-2}} \cdots p_{i_1 2} vec\left(\mathbf{A}^{(2)} \mathbf{A}^{(i_1)} \ldots \mathbf{A}^{(i_k)} {\mathbf{A}^{(i_k)}}^{\mathsf{T}} \ldots {\mathbf{A}^{(i_1)}}^{\mathsf{T}} {\mathbf{A}^{(2)}}^{\mathsf{T}}\right) \\ \vdots \\ \sum\limits_{i_1, i_2, \ldots, i_k = 1}^{l} \pi^{(i_k)} p_{i_k i_{k-1}} p_{i_{k-1} i_{k-2}} \cdots p_{i_1 l} vec\left(\mathbf{A}^{(l)} \mathbf{A}^{(i_1)} \ldots \mathbf{A}^{(i_k)} {\mathbf{A}^{(i_k)}}^{\mathsf{T}} \ldots {\mathbf{A}^{(i_1)}}^{\mathsf{T}} {\mathbf{A}^{(l)}}^{\mathsf{T}}\right) \end{pmatrix},
\end{aligned}
$$

which we establish by an induction argument. For $k$ equal to one we have

$$
\begin{pmatrix} p_{11} \mathbf{A}^{(1)} \otimes \mathbf{A}^{(1)} & \cdots & p_{l1} \mathbf{A}^{(1)} \otimes \mathbf{A}^{(1)} \\ \vdots & \ddots & \vdots \\ p_{1l} \mathbf{A}^{(l)} \otimes \mathbf{A}^{(l)} & \cdots & p_{ll} \mathbf{A}^{(l)} \otimes \mathbf{A}^{(l)} \end{pmatrix} \begin{pmatrix} \pi^{(1)} vec\left(\mathbf{A}^{(1)} {\mathbf{A}^{(1)}}^{\mathsf{T}}\right) \\ \pi^{(2)} vec\left(\mathbf{A}^{(2)} {\mathbf{A}^{(2)}}^{\mathsf{T}}\right) \\ \vdots \\ \pi^{(l)} vec\left(\mathbf{A}^{(l)} {\mathbf{A}^{(l)}}^{\mathsf{T}}\right) \end{pmatrix}
$$

$$
= \begin{pmatrix} \sum_{i_1=1}^{l} \pi^{(i_1)} p_{i_1 1} \mathbf{A}^{(1)} \otimes \mathbf{A}^{(1)} vec\left(\mathbf{A}^{(i_1)} \mathbf{A}^{(i_1)\top}\right) \\ \sum_{i_1=1}^{l} \pi^{(i_1)} p_{i_1 2} \mathbf{A}^{(2)} \otimes \mathbf{A}^{(2)} vec\left(\mathbf{A}^{(i_1)} \mathbf{A}^{(i_1)\top}\right) \\ \vdots \\ \sum_{i_1=1}^{l} \pi^{(i_1)} p_{i_1 l} \mathbf{A}^{(l)} \otimes \mathbf{A}^{(l)} vec\left(\mathbf{A}^{(i_1)} \mathbf{A}^{(i_1)\top}\right) \end{pmatrix}.
$$

The general identity $vec(ABC) = (C^\top \otimes A)vec(B)$ (cf. Nicholls and Quinn (1982, Th. A.1.1)) gives $\mathbf{A}^{(j)} \otimes \mathbf{A}^{(j)} vec\left(\mathbf{A}^{(i_1)} \mathbf{A}^{(i_1)\top}\right) = vec\left(\mathbf{A}^{(j)} \mathbf{A}^{(i_1)} \mathbf{A}^{(i_1)\top} \mathbf{A}^{(j)\top}\right)$ and thus the above claimed identity is established for $k = 1$.

Assume now that the identity holds for some $k \in \mathbb{N}$. Then

$$
Q_{\mathbf{A}}^{k+1} \begin{pmatrix} \pi^{(1)} vec\left(\mathbf{A}^{(1)} \mathbf{A}^{(1)\top}\right) \\ \pi^{(2)} vec\left(\mathbf{A}^{(2)} \mathbf{A}^{(2)\top}\right) \\ \vdots \\ \pi^{(l)} vec\left(\mathbf{A}^{(l)} \mathbf{A}^{(l)\top}\right) \end{pmatrix} = Q_{\mathbf{A}} Q_{\mathbf{A}}^{k} \begin{pmatrix} \pi^{(1)} vec\left(\mathbf{A}^{(1)} \mathbf{A}^{(1)\top}\right) \\ \pi^{(2)} vec\left(\mathbf{A}^{(2)} \mathbf{A}^{(2)\top}\right) \\ \vdots \\ \pi^{(l)} vec\left(\mathbf{A}^{(l)} \mathbf{A}^{(l)\top}\right) \end{pmatrix}
$$

$$
= \begin{pmatrix} p_{11} \mathbf{A}^{(1)} \otimes \mathbf{A}^{(1)} & \cdots & p_{l1} \mathbf{A}^{(1)} \otimes \mathbf{A}^{(1)} \\ \vdots & \ddots & \vdots \\ p_{1l} \mathbf{A}^{(l)} \otimes \mathbf{A}^{(l)} & \cdots & p_{ll} \mathbf{A}^{(l)} \otimes \mathbf{A}^{(l)} \end{pmatrix}
$$

$$
\times \begin{pmatrix} \sum_{i_2, i_3, \ldots, i_{k+1}=1}^{l} \pi^{(i_{k+1})} p_{i_{k+1} i_k} p_{i_k i_{k-1}} \cdots p_{i_2 1} vec\left(\mathbf{A}^{(1)} \mathbf{A}^{(i_2)} \ldots \mathbf{A}^{(i_{k+1})} \mathbf{A}^{(i_{k+1})\top} \ldots \mathbf{A}^{(i_2)\top} \mathbf{A}^{(1)\top}\right) \\ \sum_{i_2, i_3, \ldots, i_{k+1}=1}^{l} \pi^{(i_{k+1})} p_{i_{k+1} i_k} p_{i_k i_{k-1}} \cdots p_{i_2 2} vec\left(\mathbf{A}^{(2)} \mathbf{A}^{(i_2)} \ldots \mathbf{A}^{(i_{k+1})} \mathbf{A}^{(i_{k+1})\top} \ldots \mathbf{A}^{(i_2)\top} \mathbf{A}^{(2)\top}\right) \\ \vdots \\ \sum_{i_2, i_3, \ldots, i_{k+1}=1}^{l} \pi^{(i_{k+1})} p_{i_{k+1} i_k} p_{i_k i_{k-1}} \cdots p_{i_2 l} vec\left(\mathbf{A}^{(l)} \mathbf{A}^{(i_2)} \ldots \mathbf{A}^{(i_{k+1})} \mathbf{A}^{(i_{k+1})\top} \ldots \mathbf{A}^{(i_2)\top} \mathbf{A}^{(l)\top}\right) \end{pmatrix}
$$

$$
= \begin{pmatrix} \sum_{i_1, i_2, \ldots, i_{k+1}=1}^{l} \pi^{(i_{k+1})} p_{i_{k+1} i_k} \cdots p_{i_2 i_1} p_{i_1 1} \mathbf{A}^{(1)} \otimes \mathbf{A}^{(1)} vec\left(\mathbf{A}^{(i_1)} \ldots \mathbf{A}^{(i_{k+1})} \mathbf{A}^{(i_{k+1})\top} \ldots \mathbf{A}^{(i_1)\top}\right) \\ \sum_{i_1, i_2, \ldots, i_{k+1}=1}^{l} \pi^{(i_{k+1})} p_{i_{k+1} i_k} \cdots p_{i_2 i_1} p_{i_1 2} \mathbf{A}^{(2)} \otimes \mathbf{A}^{(2)} vec\left(\mathbf{A}^{(i_1)} \ldots \mathbf{A}^{(i_{k+1})} \mathbf{A}^{(i_{k+1})\top} \ldots \mathbf{A}^{(i_1)\top}\right) \\ \vdots \\ \sum_{i_1, i_2, \ldots, i_{k+1}=1}^{l} \pi^{(i_{k+1})} p_{i_{k+1} i_k} \cdots p_{i_2 i_1} p_{i_1 l} \mathbf{A}^{(l)} \otimes \mathbf{A}^{(l)} vec\left(\mathbf{A}^{(i_1)} \ldots \mathbf{A}^{(i_{k+1})} \mathbf{A}^{(i_{k+1})\top} \ldots \mathbf{A}^{(i_1)\top}\right) \end{pmatrix}
$$

and using again $vec(ABC) = (C^\top \otimes A)vec(B)$ to obtain $\mathbf{A}^{(j)} \otimes \mathbf{A}^{(j)} vec\left(\mathbf{A}^{(i_1)} \cdots \mathbf{A}^{(i_{k+1})} \mathbf{A}^{(i_{k+1})\top} \cdots \mathbf{A}^{(i_1)\top}\right) = vec\left(\mathbf{A}^{(j)} \mathbf{A}^{(i_1)} \cdots \mathbf{A}^{(i_{k+1})} \mathbf{A}^{(i_{k+1})\top} \cdots \mathbf{A}^{(i_1)\top} \mathbf{A}^{(j)\top}\right)$ finally concludes the proof. $\square$

**Theorem 5.23** *Assume that* $\rho(Q_{\mathbf{A}}) < 1$ *or* $\rho(Q_{\mathbf{\Phi}}) < 1$. *Then it holds that* $\gamma = \tilde{\gamma} < 0$ *and* $\limsup_{k\to\infty} E(\|\mathbf{\Phi}_0 \cdots \mathbf{\Phi}_{-k+1}\|^2)^{1/k} < 1$, $\limsup_{k\to\infty} E(\|\mathbf{A}_0 \cdots \mathbf{A}_{-k+1}\|^2)^{1/k} < 1$.

The converse implication does not hold, as will be exemplified in Section 5.4.1.

**Proof:** By Corollary 5.6 we have $\gamma = \tilde{\gamma}$ and as a consequence of Corollary 5.15 we have $\limsup_{k\to\infty} E(\|\mathbf{\Phi}_0 \cdots \mathbf{\Phi}_{-k+1}\|^2)^{1/k} < 1$, iff $\limsup_{k\to\infty} E(\|\mathbf{A}_0 \cdots \mathbf{A}_{-k+1}\|^2)^{1/k} < 1$. Thus in view of Lemma 4.5 it suffices to show that $\rho(Q_{\mathbf{A}}) < 1$ implies

$$\limsup_{k\to\infty} E(\|\mathbf{A}_0 \cdots \mathbf{A}_{-k+1}\|^2)^{1/k} < 1$$

and $\rho(Q_{\mathbf{\Phi}}) < 1$ gives

$$\limsup_{k\to\infty} E(\|\mathbf{\Phi}_0 \cdots \mathbf{\Phi}_{-k+1}\|^2)^{1/k} < 1$$

for some algebra norm. We only show the first implication, since for the other one one proceeds totally analogously.

From the last lemma and the fact that *tr* and *vec* are linear mappings we obtain

$$\begin{aligned}
E(\|\mathbf{A}_0 \cdots \mathbf{A}_{-k+1}\|_t^2) &= E\left(tr(\mathbf{A}_0 \cdots \mathbf{A}_{-k+1}\mathbf{A}_{-k+1}^{\mathsf{T}} \cdots \mathbf{A}_0^{\mathsf{T}})\right) \\
&= tr\left(E(\mathbf{A}_0 \cdots \mathbf{A}_{-k+1}\mathbf{A}_{-k+1}^{\mathsf{T}} \cdots \mathbf{A}_0^{\mathsf{T}})\right) \\
&= tr\left(vec^{-1}\left(E\left(vec(\mathbf{A}_0 \cdots \mathbf{A}_{-k+1}\mathbf{A}_{-k+1}^{\mathsf{T}} \cdots \mathbf{A}_0^{\mathsf{T}})\right)\right)\right) \\
&= tr\left(vec^{-1}\left(\mathbb{I}_{\mathbf{A}} Q_{\mathbf{A}}^{k-1}\begin{pmatrix} \pi^{(1)} vec\left(\mathbf{A}^{(1)}\mathbf{A}^{(1)^{\mathsf{T}}}\right) \\ \pi^{(2)} vec\left(\mathbf{A}^{(2)}\mathbf{A}^{(2)^{\mathsf{T}}}\right) \\ \vdots \\ \pi^{(l)} vec\left(\mathbf{A}^{(l)}\mathbf{A}^{(l)^{\mathsf{T}}}\right)\end{pmatrix}\right)\right) \\
&\leq \|tr \circ vec^{-1} \circ \mathbb{I}_{\mathbf{A}}\| \|Q_{\mathbf{A}}^{k-1}\| \left\|\begin{pmatrix} \pi^{(1)} vec\left(\mathbf{A}^{(1)}\mathbf{A}^{(1)^{\mathsf{T}}}\right) \\ \pi^{(2)} vec\left(\mathbf{A}^{(2)}\mathbf{A}^{(2)^{\mathsf{T}}}\right) \\ \vdots \\ \pi^{(l)} vec\left(\mathbf{A}^{(l)}\mathbf{A}^{(l)^{\mathsf{T}}}\right)\end{pmatrix}\right\|
\end{aligned}$$

for $k \in \mathbb{N}$. So there is a $C \in \mathbb{R}^+$ such that $E(\|\mathbf{A}_0 \cdots \mathbf{A}_{-k+1}\|_t^2)^{1/k} \leq C^{1/k}\|Q_{\mathbf{A}}^{k-1}\|^{1/k}$. Since $\lim_{k\to\infty} \|Q_{\mathbf{A}}^{k-1}\|^{1/k} = \rho(Q_{\mathbf{A}})$, this gives $\limsup_{k\to\infty} E(\|\mathbf{A}_0 \cdots \mathbf{A}_{-k+1}\|^2)^{1/k} < 1$. □

Actually it is irrelevant whether one studies $Q_{\mathbf{A}}$ or $Q_{\mathbf{\Phi}}$ as the following result from Francq and Zakoïan (2001) shows.

**Lemma 5.24** *The spectral radii of* $Q_{\mathbf{A}}$ *and* $Q_{\mathbf{\Phi}}$ *are the same.*

**Proof:** see Francq and Zakoïan (2001, Appendix A) □

We now turn to establishing that $\rho(Q_{\mathbf{A}}) < 1$ also implies that the MS-ARMA process is square integrable, if the noise sequence $(\epsilon_t)$ is so. We follow the discussion in Francq and Zakoïan (2001) who showed that this condition implies wide side sense stationarity (called second order stationarity by them). However, they did not observe that this condition also gives strict stationarity, although our above calculations are very similar to theirs. For MS-AR(p) processes Yao (2001) obtained strict stationarity and the existence of the second moment under the condition that the matrix

$$\left(P^{\mathsf{T}} \otimes I_{(dp)^2}\right) \begin{pmatrix} \mathbf{\Phi}^{(1)} \otimes \mathbf{\Phi}^{(1)} & 0 & \cdots & 0 \\ 0 & \mathbf{\Phi}^{(2)} \otimes \mathbf{\Phi}^{(2)} & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \mathbf{\Phi}^{(l)} \otimes \mathbf{\Phi}^{(l)} \end{pmatrix}$$

has spectral radius less than one. But since $\sigma(AB)\backslash\{0\} = \sigma(BA)\backslash\{0\}$ for any two matrices $A, B \in M_n(\mathbb{R})$ this is equivalent to $\rho(Q_{\mathbf{\Phi}}) < 1$.

**Lemma 5.25 (cf. Francq and Zakoïan (2001, pp. 344ff))** *Assume that $\epsilon_t \in L^2$ and $E(\epsilon_t) = 0$. Set $\mathbb{I}_{\mathbf{A}} = \left(I_{(d(p+q))^2}, I_{(d(p+q))^2}, \ldots, I_{(d(p+q))^2}\right) \in M_{(d(p+q))^2, l(d(p+q))^2}$. Then the following identities hold for all $t \in \mathbb{Z}$ and $k \in \mathbb{N}_0$:*

$$\begin{aligned} &\begin{pmatrix} E\left(vec\left(\mathbf{A}_t\mathbf{A}_{t-1}\cdots\mathbf{A}_{t-k+1}\mathbf{C}_{t-k}\mathbf{C}_{t-k}^{\mathsf{T}}\mathbf{A}_{t-k+1}^{\mathsf{T}}\cdots\mathbf{A}_{t-1}^{\mathsf{T}}\mathbf{A}_t^{\mathsf{T}}\right) I_{\{\Delta^{(1)}\}}(\Delta_t)\right) \\ E\left(vec\left(\mathbf{A}_t\mathbf{A}_{t-1}\cdots\mathbf{A}_{t-k+1}\mathbf{C}_{t-k}\mathbf{C}_{t-k}^{\mathsf{T}}\mathbf{A}_{t-k+1}^{\mathsf{T}}\cdots\mathbf{A}_{t-1}^{\mathsf{T}}\mathbf{A}_t^{\mathsf{T}}\right) I_{\{\Delta^{(2)}\}}(\Delta_t)\right) \\ \vdots \\ E\left(vec\left(\mathbf{A}_t\mathbf{A}_{t-1}\cdots\mathbf{A}_{t-k+1}\mathbf{C}_{t-k}\mathbf{C}_{t-k}^{\mathsf{T}}\mathbf{A}_{t-k+1}^{\mathsf{T}}\cdots\mathbf{A}_{t-1}^{\mathsf{T}}\mathbf{A}_t^{\mathsf{T}}\right) I_{\{\Delta^{(l)}\}}(\Delta_t)\right) \end{pmatrix} \\ = \; & Q_{\mathbf{A}}^k \left( \begin{pmatrix} \pi^{(1)}\mathbf{\Sigma}^{(1)} \otimes \mathbf{\Sigma}^{(1)} \\ \pi^{(2)}\mathbf{\Sigma}^{(2)} \otimes \mathbf{\Sigma}^{(2)} \\ \vdots \\ \pi^{(l)}\mathbf{\Sigma}^{(l)} \otimes \mathbf{\Sigma}^{(l)} \end{pmatrix} vec(E(\epsilon_0\epsilon_0^{\mathsf{T}})) + \begin{pmatrix} \pi^{(1)}\mathbf{m}^{(1)} \otimes \mathbf{m}^{(1)} \\ \pi^{(2)}\mathbf{m}^{(2)} \otimes \mathbf{m}^{(2)} \\ \vdots \\ \pi^{(l)}\mathbf{m}^{(l)} \otimes \mathbf{m}^{(l)} \end{pmatrix} \right), \end{aligned}$$

*and*

$$\begin{aligned} & E\left(vec\left(\mathbf{A}_t\mathbf{A}_{t-1}\cdots\mathbf{A}_{t-k+1}\mathbf{C}_{t-k}\mathbf{C}_{t-k}^{\mathsf{T}}\mathbf{A}_{t-k+1}^{\mathsf{T}}\cdots\mathbf{A}_{t-1}^{\mathsf{T}}\mathbf{A}_t^{\mathsf{T}}\right)\right) \\ = \; & \mathbb{I}_{\mathbf{A}} Q_{\mathbf{A}}^k \left( \begin{pmatrix} \pi^{(1)}\mathbf{\Sigma}^{(1)} \otimes \mathbf{\Sigma}^{(1)} \\ \pi^{(2)}\mathbf{\Sigma}^{(2)} \otimes \mathbf{\Sigma}^{(2)} \\ \vdots \\ \pi^{(l)}\mathbf{\Sigma}^{(l)} \otimes \mathbf{\Sigma}^{(l)} \end{pmatrix} vec(E(\epsilon_0\epsilon_0^{\mathsf{T}})) + \begin{pmatrix} \pi^{(1)}\mathbf{m}^{(1)} \otimes \mathbf{m}^{(1)} \\ \pi^{(2)}\mathbf{m}^{(2)} \otimes \mathbf{m}^{(2)} \\ \vdots \\ \pi^{(l)}\mathbf{m}^{(l)} \otimes \mathbf{m}^{(l)} \end{pmatrix} \right). \end{aligned}$$

**Proof (cf. Francq and Zakoïan (2001, pp. 344ff)):** Again the second identity follows immediately from the first. For the first we employ an induction argument. For $k = 0$ and all $t \in \mathbb{Z}$:

$$\begin{pmatrix} E\left(vec\left(\mathbf{C}_t\mathbf{C}_t^{\mathsf{T}}\right) I_{\{\Delta^{(1)}\}}(\Delta_t)\right) \\ E\left(vec\left(\mathbf{C}_t\mathbf{C}_t^{\mathsf{T}}\right) I_{\{\Delta^{(2)}\}}(\Delta_t)\right) \\ \vdots \\ E\left(vec\left(\mathbf{C}_t\mathbf{C}_t^{\mathsf{T}}\right) I_{\{\Delta^{(l)}\}}(\Delta_t)\right) \end{pmatrix}$$

$$
\begin{aligned}
&= \begin{pmatrix} E\left(vec\left(\left(\mathbf{m}^{(1)}+\mathbf{\Sigma}^{(1)}\epsilon_t\right)\left(\mathbf{m}^{(1)}+\mathbf{\Sigma}^{(1)}\epsilon_t\right)^{\mathsf{T}}\right)I_{\{\Delta^{(1)}\}}(\Delta_t)\right) \\ E\left(vec\left(\left(\mathbf{m}^{(2)}+\mathbf{\Sigma}^{(2)}\epsilon_t\right)\left(\mathbf{m}^{(2)}+\mathbf{\Sigma}^{(2)}\epsilon_t\right)^{\mathsf{T}}\right)I_{\{\Delta^{(2)}\}}(\Delta_t)\right) \\ \vdots \\ E\left(vec\left(\left(\mathbf{m}^{(l)}+\mathbf{\Sigma}^{(l)}\epsilon_t\right)\left(\mathbf{m}^{(l)}+\mathbf{\Sigma}^{(l)}\epsilon_t\right)^{\mathsf{T}}\right)I_{\{\Delta^{(l)}\}}(\Delta_t)\right) \end{pmatrix} \\
&= \begin{pmatrix} \pi^{(1)}vec\left(\mathbf{m}^{(1)}\mathbf{m}^{(1)^{\mathsf{T}}}+\mathbf{m}^{(1)}\left(\mathbf{\Sigma}^{(1)}E(\epsilon_t)\right)^{\mathsf{T}}+\mathbf{\Sigma}^{(1)}E(\epsilon_t)\mathbf{m}^{(1)^{\mathsf{T}}}+\mathbf{\Sigma}^{(1)}E(\epsilon_t\epsilon_t^{\mathsf{T}})\mathbf{\Sigma}^{(1)^{\mathsf{T}}}\right) \\ \pi^{(2)}vec\left(\mathbf{m}^{(2)}\mathbf{m}^{(2)^{\mathsf{T}}}+\mathbf{m}^{(2)}\left(\mathbf{\Sigma}^{(2)}E(\epsilon_t)\right)^{\mathsf{T}}+\mathbf{\Sigma}^{(2)}E(\epsilon_t)\mathbf{m}^{(2)^{\mathsf{T}}}+\mathbf{\Sigma}^{(2)}E(\epsilon_t\epsilon_t^{\mathsf{T}})\mathbf{\Sigma}^{(2)^{\mathsf{T}}}\right) \\ \vdots \\ \pi^{(l)}vec\left(\mathbf{m}^{(l)}\mathbf{m}^{(l)^{\mathsf{T}}}+\mathbf{m}^{(l)}\left(\mathbf{\Sigma}^{(l)}E(\epsilon_t)\right)^{\mathsf{T}}+\mathbf{\Sigma}^{(l)}E(\epsilon_t)\mathbf{m}^{(l)^{\mathsf{T}}}+\mathbf{\Sigma}^{(l)}E(\epsilon_t\epsilon_t^{\mathsf{T}})\mathbf{\Sigma}^{(l)^{\mathsf{T}}}\right) \end{pmatrix} \\
&= \begin{pmatrix} \pi^{(1)}vec\left(\mathbf{m}^{(1)}\mathbf{m}^{(1)^{\mathsf{T}}}+\mathbf{\Sigma}^{(1)}E(\epsilon_t\epsilon_t^{\mathsf{T}})\mathbf{\Sigma}^{(1)^{\mathsf{T}}}\right) \\ \pi^{(2)}vec\left(\mathbf{m}^{(2)}\mathbf{m}^{(2)^{\mathsf{T}}}+\mathbf{\Sigma}^{(2)}E(\epsilon_t\epsilon_t^{\mathsf{T}})\mathbf{\Sigma}^{(2)^{\mathsf{T}}}\right) \\ \vdots \\ \pi^{(l)}vec\left(\mathbf{m}^{(l)}\mathbf{m}^{(l)^{\mathsf{T}}}+\mathbf{\Sigma}^{(l)}E(\epsilon_t\epsilon_t^{\mathsf{T}})\mathbf{\Sigma}^{(l)^{\mathsf{T}}}\right) \end{pmatrix} \\
&= \begin{pmatrix} \pi^{(1)}\mathbf{\Sigma}^{(1)}\otimes\mathbf{\Sigma}^{(1)} \\ \pi^{(2)}\mathbf{\Sigma}^{(2)}\otimes\mathbf{\Sigma}^{(2)} \\ \vdots \\ \pi^{(l)}\mathbf{\Sigma}^{(l)}\otimes\mathbf{\Sigma}^{(l)} \end{pmatrix} vec(E(\epsilon_0\epsilon_0^{\mathsf{T}})) + \begin{pmatrix} \pi^{(1)}\mathbf{m}^{(1)}\otimes\mathbf{m}^{(1)} \\ \pi^{(2)}\mathbf{m}^{(2)}\otimes\mathbf{m}^{(2)} \\ \vdots \\ \pi^{(l)}\mathbf{m}^{(l)}\otimes\mathbf{m}^{(l)} \end{pmatrix},
\end{aligned}
$$

where we again used $vec(ABC) = (C^{\mathsf{T}} \otimes A)vec(B)$ (cf. Nicholls and Quinn (1982, Th. A.1.1)) and, moreover, the obvious $vec\left(m^{(j)}m^{(j)^{\mathsf{T}}}\right) = m^{(j)} \otimes m^{(j)}$. So, the identity is established for $k = 0$. Assume now that it holds for some $k \in \mathbb{N}_0$. Then

$$
\begin{aligned}
& Q_{\mathbf{A}}^{k+1}\left(\begin{pmatrix} \pi^{(1)}\mathbf{\Sigma}^{(1)}\otimes\mathbf{\Sigma}^{(1)} \\ \pi^{(2)}\mathbf{\Sigma}^{(2)}\otimes\mathbf{\Sigma}^{(2)} \\ \vdots \\ \pi^{(l)}\mathbf{\Sigma}^{(l)}\otimes\mathbf{\Sigma}^{(l)} \end{pmatrix} vec(E(\epsilon_0\epsilon_0^{\mathsf{T}})) + \begin{pmatrix} \pi^{(1)}\mathbf{m}^{(1)}\otimes\mathbf{m}^{(1)} \\ \pi^{(2)}\mathbf{m}^{(2)}\otimes\mathbf{m}^{(2)} \\ \vdots \\ \pi^{(l)}\mathbf{m}^{(l)}\otimes\mathbf{m}^{(l)} \end{pmatrix}\right) \\
&= \begin{pmatrix} p_{11}\mathbf{A}^{(1)}\otimes\mathbf{A}^{(1)} & \cdots & p_{l1}\mathbf{A}^{(1)}\otimes\mathbf{A}^{(1)} \\ \vdots & \ddots & \vdots \\ p_{1l}\mathbf{A}^{(l)}\otimes\mathbf{A}^{(l)} & \cdots & p_{ll}\mathbf{A}^{(l)}\otimes\mathbf{A}^{(l)} \end{pmatrix} \\
&\times \begin{pmatrix} E\left(vec\left(\mathbf{A}_{t-1}\mathbf{A}_{t-2}\cdots\mathbf{A}_{t-k}\mathbf{C}_{t-k-1}\mathbf{C}_{t-k-1}^{\mathsf{T}}\mathbf{A}_{t-k}^{\mathsf{T}}\cdots\mathbf{A}_{t-2}^{\mathsf{T}}\mathbf{A}_{t-1}^{\mathsf{T}}\right)I_{\{\Delta^{(1)}\}}(\Delta_{t-1})\right) \\ E\left(vec\left(\mathbf{A}_{t-1}\mathbf{A}_{t-2}\cdots\mathbf{A}_{t-k}\mathbf{C}_{t-k-1}\mathbf{C}_{t-k-1}^{\mathsf{T}}\mathbf{A}_{t-k}^{\mathsf{T}}\cdots\mathbf{A}_{t-2}^{\mathsf{T}}\mathbf{A}_{t-1}^{\mathsf{T}}\right)I_{\{\Delta^{(2)}\}}(\Delta_{t-1})\right) \\ \vdots \\ E\left(vec\left(\mathbf{A}_{t-1}\mathbf{A}_{t-2}\cdots\mathbf{A}_{t-k}\mathbf{C}_{t-k-1}\mathbf{C}_{t-k-1}^{\mathsf{T}}\mathbf{A}_{t-k}^{\mathsf{T}}\cdots\mathbf{A}_{t-2}^{\mathsf{T}}\mathbf{A}_{t-1}^{\mathsf{T}}\right)I_{\{\Delta^{(l)}\}}(\Delta_{t-1})\right) \end{pmatrix}
\end{aligned}
$$

$$= \begin{pmatrix} \sum_{i=1}^{l} p_{i1}\mathbf{A}^{(1)} \otimes \mathbf{A}^{(1)} E\left(vec\left(\mathbf{A}_{t-1}\cdots\mathbf{A}_{t-k}\mathbf{C}_{t-k-1}\mathbf{C}_{t-k-1}^{\mathsf{T}}\mathbf{A}_{t-k}^{\mathsf{T}}\cdots\mathbf{A}_{t-1}^{\mathsf{T}}\right) I_{\{\Delta^{(i)}\}}(\Delta_{t-1})\right) \\ \sum_{i=1}^{l} p_{i2}\mathbf{A}^{(2)} \otimes \mathbf{A}^{(2)} E\left(vec\left(\mathbf{A}_{t-1}\cdots\mathbf{A}_{t-k}\mathbf{C}_{t-k-1}\mathbf{C}_{t-k-1}^{\mathsf{T}}\mathbf{A}_{t-k}^{\mathsf{T}}\cdots\mathbf{A}_{t-1}^{\mathsf{T}}\right) I_{\{\Delta^{(i)}\}}(\Delta_{t-1})\right) \\ \vdots \\ \sum_{i=1}^{l} p_{il}\mathbf{A}^{(l)} \otimes \mathbf{A}^{(l)} E\left(vec\left(\mathbf{A}_{t-1}\cdots\mathbf{A}_{t-k}\mathbf{C}_{t-k-1}\mathbf{C}_{t-k-1}^{\mathsf{T}}\mathbf{A}_{t-k}^{\mathsf{T}}\cdots\mathbf{A}_{t-1}^{\mathsf{T}}\right) I_{\{\Delta^{(i)}\}}(\Delta_{t-1})\right) \end{pmatrix}$$

$$= \begin{pmatrix} E\left(vec\left(\mathbf{A}_t\mathbf{A}_{t-1}\cdots\mathbf{A}_{t-k}\mathbf{C}_{t-k-1}\mathbf{C}_{t-k-1}^{\mathsf{T}}\mathbf{A}_{t-k}^{\mathsf{T}}\cdots\mathbf{A}_{t-1}^{\mathsf{T}}\mathbf{A}_t^{\mathsf{T}}\right) I_{\{\Delta^{(1)}\}}(\Delta_t)\right) \\ E\left(vec\left(\mathbf{A}_t\mathbf{A}_{t-1}\cdots\mathbf{A}_{t-k}\mathbf{C}_{t-k-1}\mathbf{C}_{t-k-1}^{\mathsf{T}}\mathbf{A}_{t-k}^{\mathsf{T}}\cdots\mathbf{A}_{t-1}^{\mathsf{T}}\mathbf{A}_t^{\mathsf{T}}\right) I_{\{\Delta^{(2)}\}}(\Delta_t)\right) \\ \vdots \\ E\left(vec\left(\mathbf{A}_t\mathbf{A}_{t-1}\cdots\mathbf{A}_{t-k}\mathbf{C}_{t-k-1}\mathbf{C}_{t-k-1}^{\mathsf{T}}\mathbf{A}_{t-k}^{\mathsf{T}}\cdots\mathbf{A}_{t-1}^{\mathsf{T}}\mathbf{A}_t^{\mathsf{T}}\right) I_{\{\Delta^{(l)}\}}(\Delta_t)\right) \end{pmatrix}.$$

In the last step one uses (exemplified for the first coordinate):

$$\begin{aligned} & \sum_{i=1}^{l} p_{i1}\mathbf{A}^{(1)} \otimes \mathbf{A}^{(1)} E\left(vec\left(\mathbf{A}_{t-1}\cdots\mathbf{A}_{t-k}\mathbf{C}_{t-k-1}\mathbf{C}_{t-k-1}^{\mathsf{T}}\mathbf{A}_{t-k}^{\mathsf{T}}\cdots\mathbf{A}_{t-1}^{\mathsf{T}}\right) I_{\{\Delta^{(i)}\}}(\Delta_{t-1})\right) \\ = & \sum_{i=1}^{l} p_{i1} E\left(vec\left(\mathbf{A}^{(1)}\mathbf{A}_{t-1}\cdots\mathbf{A}_{t-k}\mathbf{C}_{t-k-1}\mathbf{C}_{t-k-1}^{\mathsf{T}}\mathbf{A}_{t-k}^{\mathsf{T}}\cdots\mathbf{A}_{t-1}^{\mathsf{T}}\mathbf{A}^{(1)^{\mathsf{T}}}\right) I_{\{\Delta^{(i)}\}}(\Delta_{t-1})\right) \\ = & \sum_{i,j=1}^{l} p_{j1} \underbrace{E\left(vec\left(\mathbf{A}^{(1)}\mathbf{A}_{t-1}\cdots\cdots\mathbf{A}_{t-1}^{\mathsf{T}}\mathbf{A}^{(1)^{\mathsf{T}}}\right) I_{\{\Delta^{(i)}\}}(\Delta_{t-1})|\Delta_{t-1}=\Delta^{(j)}\right)}_{=0 \text{ for } i\neq j} \pi^{(j)} \\ \overset{(*)}{=} & \sum_{i,j=1}^{l} E\left(vec\left(\mathbf{A}^{(1)}\mathbf{A}_{t-1}\cdots\cdots\mathbf{A}_{t-1}^{\mathsf{T}}\mathbf{A}^{(1)^{\mathsf{T}}}\right) I_{\{\Delta^{(i)}\}}(\Delta_{t-1}) I_{\{\Delta^{(1)}\}}(\Delta_t)|\Delta_{t-1}=\Delta^{(j)}\right) \pi^{(j)} \\ = & \sum_{i=1}^{l} E\left(vec\left(\mathbf{A}^{(1)}\mathbf{A}_{t-1}\cdots\mathbf{C}_{t-k-1}\mathbf{C}_{t-k-1}^{\mathsf{T}}\cdots\mathbf{A}_{t-1}^{\mathsf{T}}\mathbf{A}^{(1)^{\mathsf{T}}}\right) I_{\{\Delta^{(i)}\}}(\Delta_{t-1}) I_{\{\Delta^{(1)}\}}(\Delta_t)\right) \\ = & E\left(vec\left(\mathbf{A}^{(1)}\mathbf{A}_{t-1}\cdots\mathbf{A}_{t-k}\mathbf{C}_{t-k-1}\mathbf{C}_{t-k-1}^{\mathsf{T}}\mathbf{A}_{t-k}^{\mathsf{T}}\cdots\mathbf{A}_{t-1}^{\mathsf{T}}\mathbf{A}^{(1)^{\mathsf{T}}}\right) I_{\{\Delta^{(1)}\}}(\Delta_t)\right) \end{aligned}$$

In $(*)$ we used $p_{j1} = E(I_{\{\Delta^{(1)}\}}(\Delta_t)|\Delta_{t-1} = \Delta^{(j)})$ and the independence of $\Delta_t$ from $\Delta_{t-k-1}, \ldots, \Delta_{t-1}$ given $\Delta_{t-1}$.

Thus the claimed identity is proved for all $k \in \mathbb{N}_0$ by induction. □

**Theorem 5.26** *Assume that $\Delta$ has a finite state space, $\epsilon_0 \in L^2$, $E(\epsilon_0) = 0$ and $\rho(Q_{\mathbf{A}}) < 1$. Then all conditions of Theorem 5.2 or 5.7 are satisfied and thus there is a unique stationary and ergodic solution $X = (X_t)$ to the MS-ARMA($p, q, \Delta, \epsilon$) equation (5.3). Furthermore, the solution $X_t$ of the MS-ARMA equation (5.3) and its higher dimensional representation $\mathbf{X}_t$ are in $L^2$ and the series defining $\mathbf{X}_t$ (as given in Theorem 5.2 or 5.7) converges in $L^2$.*

**Proof (cf. Francq and Zakoïan (2001) for the $L^2$ part):** The finite state space and $\epsilon_0 \in L^2$ imply $E(\log^+ \|\mathbf{A}_0\|) < \infty$ and $E(\log^+ \|\mathbf{C}_0\|) < \infty$ as already noted in Section

5.2.2. $\gamma < 0$ and $\tilde{\gamma} < 0$ follows by Theorem 5.23. So, it only remains to show (5.23) with $p = 2$. It is clear that $\mathbf{A}_0 \cdots \mathbf{A}_{-k+1}\mathbf{C}_{-k} \in L^2$ for all natural $k$. To be able to apply Theorem 5.8 it thus suffices to show $\limsup_{k\to\infty} E(\|\mathbf{A}_0 \cdots \mathbf{A}_{-k+1}\mathbf{C}_{-k}\|^2)^{1/k} < 1$ (cf. Lemma 4.8) and the later is obviously independent of the particular norm employed. Noting that $\|x\|_2^2 = tr(xx^\mathsf{T})$ for any $x \in \mathbb{R}^n$ we obtain:

$$E(\|\mathbf{A}_0 \cdots \mathbf{A}_{-k+1}\mathbf{C}_{-k}\|_2^2) = E\left(tr\left(\mathbf{A}_0\mathbf{A}_{-1}\cdots\mathbf{A}_{-k+1}\mathbf{C}_{-k}\mathbf{C}_{-k}^\mathsf{T}\mathbf{A}_{-k+1}^\mathsf{T}\cdots\mathbf{A}_{-1}^\mathsf{T}\mathbf{A}_0^\mathsf{T}\right)\right)$$

$$= tr \circ vec^{-1}\left(E\left(vec\left(\mathbf{A}_0\mathbf{A}_{-1}\cdots\mathbf{A}_{-k+1}\mathbf{C}_{-k}\mathbf{C}_{-k}^\mathsf{T}\mathbf{A}_{-k+1}^\mathsf{T}\cdots\mathbf{A}_{-1}^\mathsf{T}\mathbf{A}_0^\mathsf{T}\right)\right)\right)$$

$$= tr \circ vec^{-1}\left(\mathbb{I}_\mathbf{A} Q_\mathbf{A}^k\left(\begin{pmatrix} \pi^{(1)}\mathbf{\Sigma}^{(1)} \otimes \mathbf{\Sigma}^{(1)} \\ \pi^{(2)}\mathbf{\Sigma}^{(2)} \otimes \mathbf{\Sigma}^{(2)} \\ \vdots \\ \pi^{(l)}\mathbf{\Sigma}^{(l)} \otimes \mathbf{\Sigma}^{(l)} \end{pmatrix} vec(E(\epsilon_0\epsilon_0^\mathsf{T})) + \begin{pmatrix} \pi^{(1)}\mathbf{m}^{(1)} \otimes \mathbf{m}^{(1)} \\ \pi^{(2)}\mathbf{m}^{(2)} \otimes \mathbf{m}^{(2)} \\ \vdots \\ \pi^{(l)}\mathbf{m}^{(l)} \otimes \mathbf{m}^{(l)} \end{pmatrix}\right)\right)$$

$$\leq \|tr \circ vec^{-1} \circ \mathbb{I}_\mathbf{A}\|\|Q_\mathbf{A}^k\| \left\|\begin{pmatrix} \pi^{(1)}\mathbf{\Sigma}^{(1)} \otimes \mathbf{\Sigma}^{(1)} \\ \pi^{(2)}\mathbf{\Sigma}^{(2)} \otimes \mathbf{\Sigma}^{(2)} \\ \vdots \\ \pi^{(l)}\mathbf{\Sigma}^{(l)} \otimes \mathbf{\Sigma}^{(l)} \end{pmatrix} vec(E(\epsilon_0\epsilon_0^\mathsf{T})) + \begin{pmatrix} \pi^{(1)}\mathbf{m}^{(1)} \otimes \mathbf{m}^{(1)} \\ \pi^{(2)}\mathbf{m}^{(2)} \otimes \mathbf{m}^{(2)} \\ \vdots \\ \pi^{(l)}\mathbf{m}^{(l)} \otimes \mathbf{m}^{(l)} \end{pmatrix}\right\|.$$

So there is a $C \in \mathbb{R}^+$ such that $E(\|\mathbf{A}_0 \cdots \mathbf{A}_{-k+1}\mathbf{C}_{-k}\|_2^2)^{1/k} \leq C^{1/k}\|Q_\mathbf{A}^k\|^{1/k}$ for all natural $k$. Since $\lim_{k\to\infty}\|Q_\mathbf{A}^k\|^{1/k} = \rho(Q_\mathbf{A})$, this gives $\limsup_{k\to\infty} E(\|\mathbf{A}_0 \cdots \mathbf{A}_{-k+1}\mathbf{C}_{-k}\|^2)^{1/k} < 1$. □

A thorough discussion of this spectral radius condition for second order stationarity is to be found in Francq and Zakoïan (2001), which also contains several examples. In particular, it is shown there that $\rho(Q_\mathbf{A}) < 1$ is also a necessary condition for the existence of a second order stationary solution to an MS-ARMA(1,1) equation under some technical assumptions. Some explicit moment calculations can also be found in Timmermann (2000). A thorough discussion on this topic is, however, beyond the scope of the present thesis focusing on theoretical properties of MS-ARMA models. Yet, in Section 5.4.1 we will give an example showing that $\rho(Q_\mathbf{A})$ can be greater one, although the Lyapunov coefficient is strictly negative.

### 5.3.3 Simulation

Another way to check, whether the stationarity or moment existence conditions are satisfied for a particular model, is to use simulations. Below we only briefly state the main ideas as a more thorough analysis is again beyond the scope of this thesis.

For the top Lyapunov coefficient it seems to be advisable to use that

$$\lim_{n\to\infty}\frac{1}{n+1}\log\|\mathbf{A}_n\cdots\mathbf{A}_0\| = \gamma \ \ a.s. \tag{5.35}$$

as shown in Furstenberg and Kesten (1960, Th. 2), since for this it is sufficient to simulate only one realization of the chain $(\mathbf{A}_n)_{n\in\mathbb{N}_0}$. The simulation of $\frac{1}{n+1}\log\|\mathbf{A}_n\cdots\mathbf{A}_0\|$ should be stopped when the sequence appears to have converged or alternatively one can use the results of Goldsheid (1991), who studies the asymptotics and gives a central limit theorem, to decide when to stop the simulation and even to construct confidence intervals. Unfortunately, it is less straightforward, how to verify the moment conditions like $\limsup_{k\to\infty} E(\|\mathbf{A}_0\cdots\mathbf{A}_{-k+1}\mathbf{C}_{-k}\|^{\tilde{p}})^{1/k} < 1$, $\limsup_{k\to\infty} E(\|\mathbf{A}_0\cdots\mathbf{A}_{-k+1}\|^{\tilde{p}})^{1/k} < 1$ or $\limsup_{k\to\infty} \|\mathbf{A}_0\cdots\mathbf{A}_{-k+1}\|_{L^\infty}^{1/k} < 1$ via simulations. In general no simple simulation scheme seems to be possible and to the best of our knowledge this question has not been addressed in the literature until now. Hence, any further analysis should be very welcome. However, if $\mathbf{A}$ is an i.i.d. sequence Lemma 4.6 (ii) is applicable and as the "lim sup" now actually is a real limit that is even equal to the infimum over the whole sequence, it seems to be advisable to simulate $E(\|\mathbf{A}_0\cdots\mathbf{A}_{-n+1}\|^s)^{1/n}$ for fixed (large enough) $n\in\mathbb{N}$ and check whether this gives a value strictly less than one. All the usual Monte Carlo simulation tools (see e.g. Asmussen (1999) for an overview) are available in order to improve the quality of the simulations, to draw inferences and, especially, to give (asymptotic) confidence bands. As the $L^\infty$ case involves studying an (essential) supremum, a similar approach using Lemma 4.6 (i) seems not to be possible to analyse $\limsup_{k\to\infty} \|\mathbf{A}_0\cdots\mathbf{A}_{-k+1}\|_{L^\infty}^{1/k} < 1$.

## 5.4 Global and Local Stationarity

The aim of this section is to analyse the relation between local stationarity and the global stationarity of an MS-ARMA(p,q) process. By *local stationarity* we mean that each regime corresponds to a causal ARMA process and by *global stationarity* that the overall MS-ARMA process is stationary and expressible as a measurable function of past and present values of $(\Delta_t)$ and $(\epsilon_t)$, i.e. $X_t = f(\Delta_t, \Delta_{t-1}, \ldots, \epsilon_t, \epsilon_{t-1}, \ldots)$. Extending standard ARMA terminology we also call such an MS-ARMA process *causal*. It will turn out that local stationarity is neither sufficient nor necessary for global stationarity in the sense that the MS-ARMA process is defined by the causal series representation given in Theorem 5.2. For the sake of notational ease we will restrict ourselves to $d = 1$, i.e. one-dimensional processes. It is, however, obvious how the results carry over to $d > 1$.

### 5.4.1 MS-ARMA(1,q): Local Stationarity Sufficient but not Necessary

Let us first consider the MS-ARMA(1,0) case and assume local stationarity, i.e. each regime is of the form $Y_t = \Phi_{1t}Y_{t-1}+\mu_t+\Sigma_t\epsilon_t$ with $|\Phi_{1t}| < 1$. Then $E(\log|\mathbf{\Phi}_0|) < 0$ and thus the top Lyapunov coefficient is strictly negative and we obtain global stationarity (given $E(\log^+|\mu_0 + \Sigma_0\epsilon_0|)$ is finite). To illustrate that local stationarity is not necessary let $(\Delta_t)$ have only two possible states and stationary distribution $(\pi^{(1)}, \pi^{(2)})$. Then $E(\log|\mathbf{\Phi}_0|) < 0$ translates into $\pi^{(1)}\log|\mathbf{\Phi}^{(1)}| + \pi^{(2)}\log|\mathbf{\Phi}^{(2)}| < 0$ and this is equivalent to $|\mathbf{\Phi}^{(1)}|^{\pi^{(1)}} < |\mathbf{\Phi}^{(2)}|^{-\pi^{(2)}}$. From the last equation it is immediate to see that $|\mathbf{\Phi}^{(1)}|$ can be arbitrarily large provided $|\mathbf{\Phi}^{(2)}|$ is close enough to zero (and vice versa). So local stationarity is not at all necessary for global stationarity. But things are slightly different, if second moments are considered. In Francq and Zakoïan (2001, p. 351) it is shown that the spectral radius condition of the previous section is necessary for an MS-ARMA(1,0) process (with finite state space of $\Delta$) to be second order stationary. For a special transition matrix structure the domain of $(\mathbf{\Phi}^{(1)}, \mathbf{\Phi}^{(2)})$ giving second order stationarity (provided $(\mathbf{C}_t)$ is independent of $(\mathbf{A}_t)$ and has finite second moment) is studied in Yao (2001) and depicted in Figure 1 of this article. It is clear that the second-order stationarity condition is more restrictive than the above one ensuring only the top Lyapunov exponent to be strictly negative. Especially, the results of Yao (2001) give that for a fixed transition matrix of $\Delta$ the possible set of $(\mathbf{\Phi}^{(1)}, \mathbf{\Phi}^{(2)})$ ensuring finite second moments is bounded, whereas above we obtained that one of the $\mathbf{\Phi}^{(i)}$ can be arbitrarily large.

Let us illustrate the relations between the individual stationarity conditions by some examples. For the sake of simplicity we presume $\Sigma_t = 1$, $\mu_t = 0$, that there are two possible states $\mathbf{\Phi}^{(1)}$ and $\mathbf{\Phi}^{(2)}$ and that the transition matrix is given by

$$P = \begin{pmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{pmatrix} = \begin{pmatrix} \bar{p} & 1-\bar{p} \\ 1-\bar{p} & \bar{p} \end{pmatrix}$$

for some $\bar{p} \in [0, 1)$. Then the stationary distribution is given by $(\pi^{(1)}, \pi^{(2)}) = (1/2, 1/2)$. Recall that for the real-valued MS-AR(1) $\gamma = E(\log |\Phi_{1t}|)$, so we can actually calculate the Lyapunov coefficient rather easily.

**Example 1:** Take $\bar{p} = 3/4$, $\mathbf{\Phi}^{(1)} = 1/2$ and $\mathbf{\Phi}^{(2)} = 11/10$. We obtain $E(\log |\Phi_{1t}|) = (1/2)\log((1/2)\cdot(11/10)) = (1/2)\log(11/20) < 0$. For the condition from Theorem 5.23 one calculates $\rho(Q_{\mathbf{A}}) = (219 + \sqrt{23761})/400 \approx 0.9328650868$ (using Maple) and for the one from the upcoming Lemma 5.28 $\rho(Q_{|A|}) = 3/5 + \sqrt{34}/20 \approx 0.8915475948$. So, all three conditions show strict negativity of the Lyapunov coefficient.

**Example 2:** Let us examine the effect of increasing the probability of remaining in the current regime. Take $\bar{p} = 49/50$, $\mathbf{\Phi}^{(1)} = 1/2$ and $\mathbf{\Phi}^{(2)} = 11/10$. We obtain $E(\log |\Phi_{1t}|) = (1/2)\log((1/2)\cdot(11/10)) = (1/2)\log(11/20) < 0$ again. However, for the condition from Theorem 5.23 one calculates $\rho(Q_{\mathbf{A}}) = (3577 + \sqrt{5534929})/5000 \approx 1.185928596$ and for the one from the upcoming Lemma 5.28 $\rho(Q_{|\mathbf{A}|}) = (98 + \sqrt{1354})/125 \approx 1.078373912$. So, all two spectral radius conditions fail to show the strict negativity of the Lyapunov coefficient, although $E(\log |\Phi_{1t}|) < 0$ holds.

**Example 3:** Let us now examine the effect of increasing the explosiveness of the second regime. Take $\bar{p} = 3/4$, $\mathbf{\Phi}^{(1)} = 1/2$ and $\mathbf{\Phi}^{(2)} = 3/2$. We obtain $E(\log |\Phi_{1t}|) = (1/2)\log((1/2)\cdot(3/2)) = (1/2)\log(3/4) < 0$. However, for the condition from Theorem 5.23 one calculates $\rho(Q_{\mathbf{A}}) = (15 + 3\sqrt{17})/16 \approx 1.710582305$ and for the one from the upcoming Lemma 5.28 $\rho(Q_{|\mathbf{A}|}) = (3 + \sqrt{3})/4 \approx 1.183012702$. So, all two spectral radius conditions fail to show the strict negativity of the Lyapunov coefficient again, although $E(\log |\Phi_{1t}|) < 0$ holds.

Observe that, actually, one has $E(\log |\Phi_{1t}|) < 0$ for any value of $\mathbf{\Phi}^{(2)}$ strictly less than two in modulus, if $\mathbf{\Phi}^{(1)} = 1/2$. The above calculations show that both spectral radius conditions considered may well fail to show the strict negativity of the Lyapunov coefficient. As $\rho(Q_{\mathbf{A}}) < 1$ under some technical conditions is shown to be necessary for the existence of a second moment of an MS-ARMA(1,1) process in Francq and Zakoïan (2001, Example 3), the above Examples 2 and 3 lead, apart from degenerate cases, to stationary, but not second-order stationary MS-ARMA processes when the noise $\epsilon_t$ is in $L^2$. Some simulations of the above considered processes are to be found in Sections 5.6.1 and 5.6.2.

To sum up the above discussion, we note that for an MS-ARMA(1,q) process local stationarity is sufficient but not necessary to ensure a strictly negative $\gamma$ and thus global stationarity.

### 5.4.2 General MS-ARMA(p,q): Local Stationarity neither Sufficient nor Necessary

Let us now turn to the case $p > 1$. From Theorem 5.21 one obtains that under rather heavy additional conditions local stationarity gives the global one. Later we will give an example showing that local stationarity is generally not sufficient. However, let us first briefly discuss that local stationarity is not necessary for the global one either.

#### Non-necessity

For the sake of simplicity let us again assume that $\Delta$ has only two possible states and give a concrete example. Let

$$\mathbf{\Phi}^{(1)} = \begin{pmatrix} \frac{1}{4} & \frac{1}{8} \\ 1 & 0 \end{pmatrix}.$$

Since $1/4 + 1/8 < 1/2 < 1$ it is clear that it is a causal ARMA(2,q) regime. Using the proof of Theorem 5.17 we construct a norm such that $\mathbf{\Phi}^{(1)}$ is within the unit circle. To this end choose $c = 1/2, c_1 = 1, c_2 = 3/4$ in the construction there. The obtained norm is thus given by $\|(x_1, x_2)^\mathsf{T}\| = \max(|x_1|, (3/4)|x_2|)$ and $\|\mathbf{\Phi}^{(1)}\| \leq 3/4$. Let now $\mathbf{\Phi}^{(2)}$ be of the following form

$$\mathbf{\Phi}_a^{(2)} = \begin{pmatrix} a & 0 \\ 1 & 0 \end{pmatrix}$$

where $a$ is greater than or equal to one. So the second regime corresponds to a non-causal ARMA(1,q) process (note $\rho(\mathbf{\Phi}_a^{(2)}) = a$). Elementary calculations immediately give that $\|\mathbf{\Phi}_a^{(2)}\| = a$ and the sufficient condition $E(\log \|\mathbf{\Phi}_0\|) < 0$ ensuring $\gamma < 0$ is equivalent to $a^{\pi^{(2)}} < (4/3)^{\pi^{(1)}}$. The latter gives that there indeed are possible values for $a$ strictly greater than one which result in global stationarity of the MS-ARMA process and the applicability of Theorem 5.2 in particular. It may be unsatisfactory that the second regime was restricted to an ARMA(1,q) one. Yet, from continuity arguments it is immediate that one can also combine the first regime with regimes of the form

$$\mathbf{\Phi}_{a,b}^{(2)} = \begin{pmatrix} a & b \\ 1 & 0 \end{pmatrix},$$

where $a$ is greater than one and $b$ sufficiently close to zero. For small $|b|$ this gives a non-causal ARMA(2,q) as second regime, but still a strictly negative Lyapunov exponent for the MS-ARMA(2,q) process.

#### Non-sufficiency

We now turn to an example showing that despite local stationarity, i.e. all regimes correspond to causal ARMA processes, the Lyapunov coefficient can be non-negative and the series representation in Theorem 5.2 may not give a stationary solution to the MS-ARMA equation. Under square-integrability conditions on the noise any causal ARMA process is also second order stationary. This may lead to the idea that, provided all regimes are causal ARMA processes, the overall MS-ARMA process should also be causal and, especially, be square-integrable for an $L^2$ noise. That local stationarity does not necessarily result in $L^2$ stationarity for MS-ARMA is demonstrated in Francq and Zakoïan (2001). Let us briefly repeat their counterexample. Assume that we have a stationary and ergodic Markov chain $\Delta$ with only two possible states and transition matrix

$$P = \begin{pmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{pmatrix}$$

and a zero-mean $L^2$ noise sequence $\epsilon$. Note that for all stochastic matrices with $p_{ii} \neq 1,\ i = 1, 2$, there exists a two state ergodic and stationary Markov chain having it as transition matrix (recall that aperiodicity is not required for ergodicity in our sense). Let further the regimes $\Delta^{(1)}$ and $\Delta^{(2)}$ be given by the following two regimes

$$X_t = \Phi_1^{(1)} X_{t-1} + \Phi_2^{(1)} X_{t-2} + \epsilon_t$$

and

$$X_t = \Phi_1^{(2)} X_{t-1} + \epsilon_t,$$

each of which shall correspond to a causal ARMA process. Assume now that there exists an MS-ARMA(2,0) process $(X_t)$ that is stationary, in $L^2$ and solves the MS-ARMA equation (5.3) with the above given $\Delta$ and $\epsilon$. Moreover, assume that $X$ is a causal solution, i.e. $X_t$ can be represented as a measurable function of $\Delta_t, \Delta_{t-1}, \ldots$ and $\epsilon_t, \epsilon_{t-1}, \ldots$. Then the conditional expectation $E\left(X_t^2 | \Delta_t = \Delta^{(1)}, \Delta_{t-1} = \Delta^{(2)}\right)$ exists and is the same at all times $t$. One calculates:

$$\begin{aligned}
& E\left(X_t^2 | \Delta_t = \Delta^{(1)}, \Delta_{t-1} = \Delta^{(2)}\right) \\
= \; & E\left(\left(\left(\Phi_1^{(1)}\Phi_1^{(2)} + \Phi_2^{(1)}\right) X_{t-2} + \epsilon_t + \Phi_1^{(1)}\epsilon_{t-1}\right)^2 | \Delta_t = \Delta^{(1)}, \Delta_{t-1} = \Delta^{(2)}\right) \\
= \; & E\left(\left(\left(\Phi_1^{(1)}\Phi_1^{(2)} + \Phi_2^{(1)}\right) X_{t-2}\right)^2 + \left(\epsilon_t + \Phi_1^{(1)}\epsilon_{t-1}\right)^2 \right. \\
& \left. + \left(\Phi_1^{(1)}\Phi_1^{(2)} + \Phi_2^{(1)}\right) X_{t-2}\left(\epsilon_t + \Phi_1^{(1)}\epsilon_{t-1}\right) | \Delta_t = \Delta^{(1)}, \Delta_{t-1} = \Delta^{(2)}\right).
\end{aligned}$$

Using that $X_{t-2}$ and $\{\epsilon_t, \epsilon_{t-1}\}$, as well as $\Delta$ and $\epsilon$ are independent and that $\epsilon_t$ has zero mean, we obtain:

$$\begin{aligned}
& E\left(X_t^2 | \Delta_t = \Delta^{(1)}, \Delta_{t-1} = \Delta^{(2)}\right) \\
= \; & E\left(\left(\left(\Phi_1^{(1)}\Phi_1^{(2)} + \Phi_2^{(1)}\right) X_{t-2}\right)^2 + \left(\epsilon_t + \Phi_1^{(1)}\epsilon_{t-1}\right)^2 | \Delta_t = \Delta^{(1)}, \Delta_{t-1} = \Delta^{(2)}\right) \\
\geq \; & \left(\Phi_1^{(1)}\Phi_1^{(2)} + \Phi_2^{(1)}\right)^2 E\left(X_{t-2}^2 | \Delta_t = \Delta^{(1)}, \Delta_{t-1} = \Delta^{(2)}\right) \\
= \; & \left(\Phi_1^{(1)}\Phi_1^{(2)} + \Phi_2^{(1)}\right)^2 \sum_{i,j=1}^{2} P\left(\Delta_{t-2} = \Delta^{(i)}, \Delta_{t-3} = \Delta^{(j)} | \Delta_t = \Delta^{(1)}, \Delta_{t-1} = \Delta^{(2)}\right) \\
& \times E\left(X_{t-2}^2 | \Delta_t = \Delta^{(1)}, \Delta_{t-1} = \Delta^{(2)}, \Delta_{t-2} = \Delta^{(i)}, \Delta_{t-3} = \Delta^{(j)}\right) \cdot
\end{aligned}$$

$$\begin{aligned}
&\geq && \left(\Phi_1^{(1)}\Phi_1^{(2)} + \Phi_2^{(1)}\right)^2 E\left(X_{t-2}^2 | \Delta_t = \Delta^{(1)}, \Delta_{t-1} = \Delta^{(2)}, \Delta_{t-2} = \Delta^{(1)}, \Delta_{t-3} = \Delta^{(2)}\right) \\
& && \times P\left(\Delta_{t-2} = \Delta^{(1)}, \Delta_{t-3} = \Delta^{(2)} | \Delta_t = \Delta^{(1)}, \Delta_{t-1} = \Delta^{(2)}\right) \\
&\overset{\Delta \text{ Markov}}{=} && \left(\Phi_1^{(1)}\Phi_1^{(2)} + \Phi_2^{(1)}\right)^2 E\left(X_{t-2}^2 | \Delta_{t-2} = \Delta^{(1)}, \Delta_{t-3} = \Delta^{(2)}\right) p_{12}p_{21} \\
&= && \left(\Phi_1^{(1)}\Phi_1^{(2)} + \Phi_2^{(1)}\right)^2 p_{12}p_{21} E\left(X_t^2 | \Delta_t = \Delta^{(1)}, \Delta_{t-1} = \Delta^{(2)}\right).
\end{aligned}$$

Thus it must hold that $\left(\Phi_1^{(1)}\Phi_1^{(2)} + \Phi_2^{(1)}\right)^2 p_{12}p_{21} \leq 1$. Yet, take e.g. $\Phi_1^{(1)} = 9/5$, $\Phi_2^{(1)} = -9/10$ and $\Phi_1^{(2)} = -1/5$. This gives $\left(\Phi_1^{(1)}\Phi_1^{(2)} + \Phi_2^{(1)}\right)^2 = (63/50)^2$, which implies that $p_{12}p_{21} \leq (63/50)^{-2} \approx 0.630$ needs to hold. For $p_{12}, p_{21} \in [0.8, 1]$ the latter is, however, obviously violated. Thus there cannot be any stationary, causal MS-ARMA process with finite second moments for the above chosen parameter values. But note that one obtains $\rho(\mathbf{\Phi}^{(1)}) = |(9/10) \pm (3/10)i| = 3/\sqrt{10} < 1$ and $\rho(\mathbf{\Phi}^{(2)}) = 1/5$ and thus both regimes correspond to causal ARMA ones. In Francq and Zakoïan (2001) a simulation is undertaken to examine this explosive behaviour. From the simulated path they conclude that the critical changes leading to the explosion occur, when the regime switches, which is rather often the case for the above parameters. The precise reason will become obvious when we now study this counterexample further. For a very concrete and thus highly tractable example we show below that the series representation for a solution to an MS-ARMA equation as given in Theorem 5.2 does not converge a.s. and thus the Lyapunov exponent is non-negative, although the individual regimes are causal.

For the above studied parameters we have

$$\mathbf{\Phi}^{(1)} = \begin{pmatrix} 9/5 & -9/10 \\ 1 & 0 \end{pmatrix}$$

and

$$\mathbf{\Phi}^{(2)} = \begin{pmatrix} -1/5 & 0 \\ 1 & 0 \end{pmatrix}.$$

The crucial observation now is that

$$R := \mathbf{\Phi}^{(1)}\mathbf{\Phi}^{(2)} = \begin{pmatrix} -63/50 & 0 \\ -1/5 & 0 \end{pmatrix}$$

and

$$S := \mathbf{\Phi}^{(2)}\mathbf{\Phi}^{(1)} = \begin{pmatrix} -9/25 & 9/50 \\ 9/5 & -9/10 \end{pmatrix}$$

both have spectral radius $63/50 > 1$. Since $\mathbf{\Phi}^{(1)}$ and $\mathbf{\Phi}^{(2)}$ have spectral radius less than one, ${\mathbf{\Phi}^{(i)}}^k$ is a contraction (i.e. has operator norm less than 1) for all large enough $k \in \mathbb{N}$ and all norms on $\mathbb{R}^2$. That is why, there is a causal solution to the respective ARMA equations representable by an absolutely converging series as in Theorem 5.2. But if we switch between the two regimes regularly no contraction but an explosion may be obtained in the long run, since $R$ and $S$ have $-63/50$ as an eigenvalue. This seems to be the precise reason, why there was no "causal" second-order stationary solution possible above.

Fixing $p_{12}$ and $p_{21}$ to the value one, we obtain an ergodic and periodic Markov chain $\Delta$, which has stationary distribution $(\pi^{(1)}, \pi^{(2)}) = (0.5, 0.5)$. Let us further assume that the noise $\epsilon$ is not random at all, but $\epsilon_t = 1$ for all times. For this model we now study the series

$$\mathbf{X}_t = \sum_{k=0}^{\infty} \mathbf{A}_t \mathbf{A}_{t-1} \cdots \mathbf{A}_{t-k+1} \mathbf{C}_{t-k}, \tag{5.36}$$

that is the stationary and ergodic solution of an MS-ARMA process as obtained in Theorem 5.2, in some more detail. Note that

$$\mathbf{A}^{(i)} = \mathbf{\Phi}^{(i)} \text{ and } \mathbf{C}_t = (1,0)^{\mathsf{T}}.$$

W.l.o.g. we can restrict attention to $\mathbf{X}_0$. One readily calculates for $n \in \mathbb{N}$

$$R^n \mathbf{C}_0 = \begin{pmatrix} \left(\frac{-63}{50}\right)^n \\ \frac{-1}{5}\left(\frac{-63}{50}\right)^{n-1} \end{pmatrix}$$

and

$$\mathbf{A}^{(2)} R^n \mathbf{C}_0 = \begin{pmatrix} \frac{-1}{5}\left(\frac{-63}{50}\right)^n \\ \left(\frac{-63}{50}\right)^n \end{pmatrix}.$$

So, both $R^n\mathbf{C}_0$ and $\mathbf{A}^{(2)} R^n \mathbf{C}_0$ diverge and even escape to infinity in norm for $n \to \infty$.

Since our Markov chain $\Delta$ switches necessarily all the time between its two states, there are only two possible, mutual exclusive, cases having probability one half each. In the first case we have $\mathbf{A}_{-2k} = \mathbf{A}^{(1)}$ and $\mathbf{A}_{-2k+1} = \mathbf{A}^{(2)}$ for all $k \in \mathbb{N}_0$. Hence, $\mathbf{A}_0 \cdots \mathbf{A}_{-2k+1}\mathbf{C}_{-2k} = R^k\mathbf{C}_0$ for $k \in \mathbb{N}$. In the other case $\mathbf{A}_{-2k} = \mathbf{A}^{(2)}$ and $\mathbf{A}_{-2k+1} = \mathbf{A}^{(1)}$ for all $k \in \mathbb{N}_0$. Thus, $\mathbf{A}_0\mathbf{A}_{-1} \cdots \mathbf{A}_{-2k}\mathbf{C}_{-2k-1} = \mathbf{A}^{(2)}R^k\mathbf{C}_0$ for $k \in \mathbb{N}$. This shows that in any case the summands of the series in equation (5.36) do not converge to zero. Hence, the series (5.36) is almost sure divergent.

This result shows that the series (5.36) does not provide a stationary solution to the MS-ARMA equation. So it is possible to combine causal ARMA regimes in such a way to an MS-ARMA equation that there is no stationary and causal solution in the sense of Theorem 5.2.

Above we used a trivial deterministic noise, since this allowed for relatively simple explicit calculations. However, in the above example one also concludes immediately from Theorem 5.2 that the Lyapunov coefficient cannot be strictly negative (otherwise Theorem 5.2 would imply absolute convergence of (5.36)). Thus, for the above used Markov chain $\Delta$ and any i.i.d. noise sequence $\epsilon$ Theorem 5.2 cannot be applied to obtain a solution to the MS-ARMA$(2, 0, \Delta, \epsilon)$ equation.

## 5.5 Geometric Ergodicity and Strong Mixing

Since strong mixing has important consequences for extreme value analysis as recalled in the preliminary Section 2.6, we now turn to analysing, when an MS-ARMA process is strongly mixing. To this end, we study the geometric ergodicity of an appropriate Markov chain. Our results extend the ones for random coefficient autoregressions dating back to Feigin and Tweedie (1985), which will be briefly discussed later on, and partly those of Yao and Attali (2000) to driving Markov chains with a state space that is not necessarily finite. However, while Yao and Attali (2000) studied possibly non-linear Markov-switching autoregressions, we restrict ourselves to the linear case and only employ geometric ergodicity, since being mainly interested in extremal behaviour $V$-uniform ergodicity, as considered by Yao and Attali (2000), seems to add only rather limited extra value in our eyes.

It is obvious that $\mathbf{X}_t = \mathbf{A}_t\mathbf{X}_{t-1} + \mathbf{C}_t = \mathbf{A}_t\mathbf{X}_{t-1} + \mathbf{m}_t + \mathbf{\Sigma}_t\epsilon_t$ is in general not a Markov chain, since the transitions do depend upon the state the driving chain $\Delta$ is in. One may, however, note that for an i.i.d. driving chain $\Delta$ the sequence $\mathbf{X} = (\mathbf{X}_t)$ is indeed a Markov chain. To obtain general results it is necessary to study the process $(\mathbf{X}_t, \Delta_t)$ as in Yao and Attali (2000). It is immediate to see that $(\mathbf{X}_t, \Delta_t)$ is a (homogeneous) Markov chain. Actually, the same is true for general stochastic difference equations with Markovian input, but we shall restrict ourselves to MS-ARMA processes. The results we give in the following can, however, also be extended to this more general case with the necessary changes being rather straightforward. For the sake of simplicity we shall denote the state space of $\Delta$ by $E$ (recall that this is a subset of $\mathbb{R}^d \times M_d(\mathbb{R})^{1+p+q}$) and the Borel-$\sigma$-algebra restricted to $E$ by $\mathcal{E}$. Thus, the state space of $(\mathbf{X}_t, \Delta_t)$ is $\mathbb{R}^{d(p+q)} \times E$ and equipped with the $\sigma$-algebra $\mathcal{B}(\mathbb{R}^{d(p+q)}) \times \mathcal{E}$.

Moreover, we formulate the results of this section in such a way that they can be employed to the case where the chain $(\mathbf{X}_t, \Delta_t)$ is started at time zero with initial (possibly random) values $(\mathbf{X}_0, \Delta_0)$. However, this applies solely to this section.

Below we summarize first the results of Yao and Attali (2000) for the case of a finite state space $E$, as the results we obtain for the general case later need considerably more technical conditions. In particular, we will in the general set-up first study the weak Feller property and irreducibility in some detail before turning our attention to geometric ergodicity and strong mixing. Finally, we repeat the results for random coefficient autoregressions for our particular set-up.

### 5.5.1 Geometric Ergodicity and Strong Mixing for a Finite State Space Chain $\Delta$

In this paragraph we again employ the notation introduced in 5.3.2 for $\Delta$ having a finite state space and throughout presume $\Delta$ to be a positive recurrent and stationary Markov chain. Furthermore, we omit any proofs and refer the interested reader to Yao and Attali (2000), since all theorems given are just adaptations of their results to the linear case.

**Theorem 5.27 (Yao and Attali (2000, Th. 1))** *Assume that the state space $E$ of $\Delta$ is finite, that $\Sigma_t = I_d$ and that $\epsilon_1$ has a strictly positive density w.r.t. the Lebesgue measure on $\mathbb{R}^d$ and is in $L^\eta$ for some $\eta > 0$. If there is a norm $\|\cdot\|$ on $\mathbb{R}^{d(p+q)}$ such that $E(\log\|\mathbf{A}_1\|) < 0$, then $(\mathbf{X}_t, \Delta_t)$ is geometrically ergodic.*

*In particular, if $(\Delta_t)_{t\in\mathbb{Z}}$ is stationary and ergodic, then all conditions of Theorem 5.2 are fulfilled and the chain $(\mathbf{X}_t, \Delta_t)_{t\in\mathbb{Z}}$, the higher dimensional representation $\mathbf{X}$ of the MS-ARMA process as well as the MS-ARMA process $X$ itself are strongly mixing.*

**Proof:** $E(\log\|\mathbf{A}_1\|) < 0$ gives $\gamma < 0$. That the other conditions of Theorem 5.2 are satisfied, is obvious from the finiteness of $E$ and $\epsilon_1 \in L^\eta$. For the remainder of the proof see

Yao and Attali (2000). □

The following result is in some respect very similar to ours from Section 5.3.2 and can be used as another feasible way to check the strict negativity of the Lyapunov exponent when considering only finitely many states for the driving chain $\Delta$. Define

$$\begin{aligned} Q_{\|\mathbf{A}\|} &:= \begin{pmatrix} \|\mathbf{A}^{(1)}\| & 0 & \cdots & 0 \\ 0 & \|\mathbf{A}^{(2)}\| & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \|\mathbf{A}^{(l)}\| \end{pmatrix} P^{\mathsf{T}} \\ &= \begin{pmatrix} p_{11}\|\mathbf{A}^{(1)}\| & p_{21}\|\mathbf{A}^{(1)}\| & \cdots & p_{l1}\|\mathbf{A}^{(1)}\| \\ p_{12}\|\mathbf{A}^{(2)}\| & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & p_{l(l-1)}\|\mathbf{A}^{(l-1)}\| \\ p_{1l}\|\mathbf{A}^{(l)}\| & \cdots & p_{(l-1)l}\|\mathbf{A}^{(l)}\| & p_{ll}\|\mathbf{A}^{(l)}\| \end{pmatrix}. \end{aligned}$$

**Lemma 5.28 (cf. Yao and Attali (2000, Lemma 2))** *It holds that*

$$E(\log\|\mathbf{A}_1\|) \leq \log\rho(Q_{\|\mathbf{A}\|}).$$

*Thus* $\rho(Q_{\|\mathbf{A}\|})$ *ensures* $E(\log\|\mathbf{A}_1\|) < 0$ *and, moreover,* $\gamma < 0$.

As noted already in Section 5.4.1, the converse implication does not hold, i.e. $E(\log\|\mathbf{A}_1\|)$ can be strictly negative and $\rho(Q_{\|\mathbf{A}\|})$ positive.

## 5.5.2 The General State Space Case

In this section we return back to the general case for the state space $E$. The first step is to study, when the chain $(\mathbf{X}_t, \Delta_t)$ is weakly Fellerian. The following two propositions give a necessary as well as a sufficient criterion, but unfortunately we have not been able to provide a necessary and sufficient condition.

**Proposition 5.29 (Necessary condition)** *Assume that* $(\Delta_1|\Delta_0, \mathbf{X}_0) \stackrel{\mathscr{D}}{=} (\Delta_1|\Delta_0)$ *(i.e. given* $\Delta_0$ *the random variables* $\Delta_1$ *and* $\mathbf{X}_0$ *are independent). If* $(\mathbf{X}_t, \Delta_t)_{t\in\mathbb{Z}}$ *is weakly Fellerian, then* $\Delta = (\Delta_t)_{t\in\mathbb{Z}}$ *is a weak Feller chain.*

The conditional independence assumption is, in particular, satisfied, if either the chain $(\mathbf{X}_t, \Delta_t)$ has been started with independent $\Delta_0$ and $\mathbf{X}_0$ or $\mathbf{X}_0$ can be represented as a measurable function of the presence and infinite past of the driving chain $\Delta$ and the noise $\epsilon$, i.e. $\mathbf{X}_0 = f(\Delta_0, \Delta_{-1}, \Delta_{-2}, \ldots, \epsilon_0, \epsilon_{-1}, \epsilon_{-2}, \ldots)$. In all considerations of the previous sections in this chapter we have only looked at MS-ARMA processes with the latter property.

**Proof:** Let $g : E \to \mathbb{R}$ be bounded and continuous. Define $\tilde{g} : \mathbb{R}^{d(p+q)} \times E \to \mathbb{R}$ via $\tilde{g}(x, \delta) = g(\delta)$. Then $\tilde{g}$ is bounded and continuous and

$$E(g(\Delta_1)|\Delta_0 = \delta) = E(\tilde{g}(\mathbf{X}_1, \Delta_1)|\mathbf{X}_0 = x, \Delta_0 = \delta)$$

is continuous, since $(\mathbf{X}_t, \Delta_t)$ is weakly Fellerian. Thus $\Delta$ is a weak Feller chain. □

**Proposition 5.30 (Sufficient Condition)** *Assume that there is some measurable function $F$ such that $\Delta_t = F(\Delta_{t-1}, u_t)$ holds, where $(u_t)_{t\in\mathbb{Z}}$ is an i.i.d. sequence assuming values in a measurable space $(G, \mathcal{G})$ and $F(\cdot, u)$ is continuous for any fixed $u \in G$. Then $(\mathbf{X}_t, \Delta_t)$ is a weak Feller chain.*

To demand the existence of a function $F$ such that $\Delta_t = F(\Delta_{t-1}, u_t)$ may seem to be rather restrictive. Yet, one should note that the condition is still very flexible and many Markov chains are of this type (cf. Meyn and Tweedie (1993, Sec. 2.2 and Ch. 7) and Duflo (1997, p. 183)). Compared to the non-linear state space models studied in Meyn and Tweedie (1993, Sec. 2.2 and Ch. 7) our assumptions are even weaker, since we do not request any continuous differentiability. If $\Delta_t = F(\Delta_{t-1}, u_t)$ holds, then $\Delta$ is weakly Fellerian as pointed out in Meyn and Tweedie (1993, Prop. 6.1.2). This should also become obvious when inspecting the proof of the above proposition below. Furthermore, if $\Delta$ has a countable state space consisting only of isolated points, then the above condition is satisfied. Recall also that $\Delta$ and $\epsilon$ are always assumed to be independent.

**Proof:** Since projections are continuous, there are functions $F_{\mathbf{A}}, F_{\mathbf{m}}, F_{\mathbf{\Sigma}}$ such that $\mathbf{A}_t = F_{\mathbf{A}}(\Delta_{t-1}, u_t)$, $\mathbf{m}_t = F_{\mathbf{m}}(\Delta_{t-1}, u_t)$, $\mathbf{\Sigma}_t = F_{\mathbf{\Sigma}}(\Delta_{t-1}, u_t)$ and $F_{\mathbf{A}}, F_{\mathbf{m}}$, $F_{\mathbf{\Sigma}}$, are continuous in $\Delta_{t-1}$. Thus, we obtain that

$$(\mathbf{X}_t, \Delta_t) = (F_{\mathbf{A}}(\Delta_{t-1}, u_t)\mathbf{X}_{t-1} + F_{\mathbf{m}}(\Delta_{t-1}, u_t) + F_{\mathbf{\Sigma}}(\Delta_{t-1}, u_t)\epsilon_t, F(\Delta_{t-1}, u_t))$$

is a continuous function of $(\mathbf{X}_{t-1}, \Delta_{t-1})$.

Let $g : \mathbb{R}^{d(p+q)} \times E \to \mathbb{R}$ be bounded and continuous and denote $P_{\epsilon_0,u_0}$ the distribution of $(\epsilon_0, u_0)$, then

$$E\left(g(\mathbf{X}_1, \Delta_1)|\mathbf{X}_0 = x, \Delta_0 = \delta\right) = \int_{\mathbb{R}^d\times G} g\left(F_{\mathbf{A}}(\delta, u)x + F_{\mathbf{m}}(\delta, u) + F_{\mathbf{\Sigma}}(\delta, u)\eta\right) dP_{\epsilon_0,u_0}(\eta, u)$$

is a continuous function of $(x, \delta)$, as the continuity lemma from standard integration theory (see, for instance, Bauer (1992, Lemma 16.1)) shows. □

Now we turn to giving a sufficient condition for the existence of a measure $\mu$ such that $(\mathbf{X}_t, \Delta_t)$ is $\mu$-irreducible. In the following $\lambda^r$ denotes the Lebesgue measure on $\mathbb{R}^r$.

**Proposition 5.31** *Let $P^n_\Delta$ denote the transition kernel of the Markov chain $\Delta$ and be $\mu_\Delta$ a nondegenerate measure on $(E, \mathcal{E})$ such that for any $A \in \mathcal{E}$ with $\mu_\Delta(A) > 0$ and $x \in E$*

$$\sum_{n=p+q}^{\infty} P^n_\Delta(x, A) > 0 \tag{5.37}$$

*holds. Assume that $\epsilon_0$ has a strictly positive density w.r.t. $\lambda^d$ and, moreover, that $\Sigma_t$ is invertible for all possible states of $\Delta_t$. Then $(\mathbf{X}_t, \Delta_t)$ is $\lambda^{d(p+q)} \otimes \mu_\Delta$-irreducible.*

**Proof:** Condition (5.37) immediately implies that $\Delta$ is $\mu_\Delta$-irreducible. Inspecting the iteration $\mathbf{X}_t = \mathbf{A}_t\mathbf{X}_{t-1} + \mathbf{C}_t$, it is obvious that under the above assumptions $\mathbf{X}_{p+q+k}$ can reach any set of positive Lebesgue measure for all $k \in \mathbb{N}_0$ with strictly positive probability regardless of the value $(\mathbf{X}_0, \Delta_0)$ and the evolution of the chain $(\Delta_t)$. Combining this with the fact that for every set $A$ with positive measure $\mu_\Delta$ there is an $n \geq p + q$ such that $P^n_\Delta(x, A) > 0$, yields the result. □

The crucial condition is (5.37), but, actually, in many cases of interest it should suffice to demand that $\Delta$ be ($\mu$)-irreducible, for instance:

**Lemma 5.32** *Assume that $\Delta$ has a countable state space $E$ and is irreducible. Then (5.37) is satisfied.*

**Proof:** The case card$(E) = 1$ is trivial. Let thus $\Delta^{(1)}$ and $\Delta^{(2)}$ be two different possible states of $\Delta$. Then there are $m, n \in \mathbb{N}$ such that $P^m(\Delta^{(1)}, \Delta^{(2)}) > 0$ and $P^n(\Delta^{(2)}, \Delta^{(1)}) > 0$ and thus $P^{km+(k-1)n}(\Delta^{(1)}, \Delta^{(2)}) > 0$ and $P^{k(m+n)}(\Delta^{(1)}, \Delta^{(1)}) > 0$ for all $k \in \mathbb{N}$. $\square$

Now we give a theorem on the geometric ergodicity of $(\mathbf{X}_t, \Delta_t)$ under rather technical conditions that appear to be the weakest necessary for our proof to work. Ways of actually checking the Feller chain and irreducibility conditions are given in the previous propositions.

**Theorem 5.33** *Assume that $(\mathbf{X}_t, \Delta_t)$ is a weak Feller chain, $(\Delta_1|\mathbf{X}_0, \Delta_0) \stackrel{\mathscr{D}}{=} (\Delta_1|\Delta_0)$ $(*)$, the state space $E$ of $\Delta$ is compact and there exists a nondegenerate measure $\mu_\Delta$ on $(E, \mathcal{E})$ such that $(\mathbf{X}_t, \Delta_t)$ is $\lambda^{d(p+q)} \otimes \mu_\Delta$-irreducible. If, moreover, there is an $\eta \in (0, 1]$ and $c < 1$ such that*

$$E(\|\mathbf{A}_1\|^\eta|\Delta_0 = \delta) \leq c \;\forall\; \delta \in E \tag{5.38}$$

*for some norm $\|\cdot\|$ on $\mathbb{R}^{d(p+q)}$ and $\epsilon_1 \in L^\eta$, then $(\mathbf{X}_t, \Delta_t)$ is geometrically ergodic.*

*In particular, if $(\Delta_t)_{t\in\mathbb{Z}}$ is stationary and ergodic, then $(*)$ is automatically satisfied provided all other conditions are. Moreover, all conditions of Theorem 5.2 are fulfilled and $(\mathbf{X}_t, \Delta_t)_{t\in\mathbb{Z}}$, the higher dimensional representation $\mathbf{X}$ of the MS-ARMA process as well as the MS-ARMA process $X$ itself are strongly mixing (with geometric rate).*

The condition for the very last assertion is due to the fact that we temporarily allowed for starting the process at time zero with initial value $(\mathbf{X}_0, \Delta_0)$ in this very section. (5.38) appears to be very restrictive. Yet, note that the conditions of Yao and Attali (2000, Th. 1) are (in our linear set-up) equivalent to assume the existence of an $0 < \eta \leq 1$ such that $E(\|\mathbf{A}_1\|^\eta) < 1$, as can be immediately seen from the results given in Basrak (2000, p. 78). The reason, why we have to resort to using a condition involving conditional expectations, is that arguments involving suprema over the state space, which were employed to prove the finite state space results of Yao and Attali (2000), do, as far as we can see, not necessarily work in the non-finite state space case. Unfortunately, we have not been able to find any nicer general conditions.

**Proof:** Consider the continuous function $g : \mathbb{R}^{d(p+q)} \to \mathbb{R}^+$, $x \mapsto \|x\|^\eta + 1$ and note that we have $\|a + b\|^\eta \leq \|a\|^\eta + \|b\|^\eta$ for all $a, b \in \mathbb{R}^{d(p+q)}$ as $0 < \eta \leq 1$ (cf. Loève (1977, p. 157)). Thus, for any $x \in \mathbb{R}^{d(p+q)}$ and $\delta \in E$

$$\begin{aligned} E(\|\mathbf{X}_1\|^\eta + 1|\mathbf{X}_0 = x, \Delta_0 = \delta) &= E(\|\mathbf{A}_1 x + \mathbf{C}_1\|^\eta + 1|\mathbf{X}_0 = x, \Delta_0 = \delta) \\ &\leq E(\|\mathbf{A}_1\|^\eta|\Delta_0 = \delta)\|x\|^\eta + E(\|\mathbf{C}_1\|^\eta|\Delta_0 = \delta) + 1. \end{aligned}$$

As $E$ is compact and $\epsilon_1 \in L^\eta$, there is a $M > 0$ such that $E(\|\mathbf{C}_1\|^\eta|\Delta_0 = \delta) < M - 1$ for all $\delta \in E$. Hence,

$$E(\|\mathbf{X}_1\|^\eta + 1|\mathbf{X}_0 = x, \Delta_0 = \delta) \;\leq\; c\|x\|^\eta + M.$$

Choose now $\tau > 0$ with $1 - \tau > c$ and then set $R = \left(\frac{M}{1-\tau-c}\right)^{1/\eta}$ and $C = B_R(0)$ (the ball with radius $R$ in $\mathbb{R}^{d(p+q)}$). For all $x \in (B_R(0))^c$ we have that $(1 - \tau - c)\|x\|^\eta \geq M$ and therefore

$$E(\|\mathbf{X}_1\|^\eta + 1|\mathbf{X}_0 = x, \Delta_0 = \delta) \quad \leq \quad c\|x\|^\eta + (1-\tau-c)\|x\|^\eta = (1-\tau)\|x\|^\eta \leq (1-\tau)g(x).$$

Setting $K := C \times E$ we obtain a compact set that together with $g$ and $\tau$ satisfies $(ii)$ of Theorem 2.35. Thus, Theorem 2.35 shows the claimed geometric ergodicity of $(\mathbf{X}_t, \Delta_t)$.

Assume now that $(\Delta_t)_{t\in\mathbb{Z}}$ is stationary and ergodic. The compactness of $E$ and the fact that $\epsilon_1 \in L^\eta$ ensure the finiteness of $E(\|\mathbf{C}_0\|^\eta)$ and thus $E(\log^+\|\mathbf{C}_0\|)$. Likewise, (5.38) gives $E(\|\mathbf{A}_1\|^\eta) < c$, which implies $E(\log^+\|\mathbf{A}_0\|) < \infty$ and $\gamma < 0$. So, all conditions of Theorem 5.2 are satisfied and $(*)$ is now a consequence of $\mathbf{X}_0$ having the series representation given there.

The strong mixing properties stated are now established applying Propositions 2.37 and 2.38. □

Using Meyn and Tweedie (1993, Th. 16.1.2) one should be able to show $V$-uniform ergodicity of $(\mathbf{X}_t, \Delta_t)$ under similar conditions with $V$ being the above considered $g$. Since this would necessitate the introduction of even more technical notions from general Markov chain theory, we refrain from carrying this out in the present thesis.

Naturally, the last theorem raises the question, whether there are some rather easy to check conditions for (5.38) to hold. A straightforward one is the existence of a norm $\|\cdot\|$ and a $c < 1$ such that $\|\mathbf{A}_1\| \leq c$ for all possible states of $\Delta_1$. Again Theorem 5.18 turns out to be helpful, since the following is an immediate consequence:

**Proposition 5.34** *Assume that $E$ is compact and that there is a norm $\|\cdot\|_d$ on $\mathbb{R}^d$ and a $\bar{c} < 1$ such that $\sum_{i=1}^p \|\Phi_{i1}\| \leq \bar{c}$ for all possible states of $\Delta_1$, then there is a norm $\|\cdot\|$ on $\mathbb{R}^{d(p+q)}$ and a $c < 1$ (both explicitly constructed in the proof of Th. 5.18) with $\|\mathbf{A}_1\|_d \leq c$ for all possible states of $\Delta_1$. In particular, (5.38) is satisfied.*

Recall that the same conditions are used in Theorem 5.21 to ensure finiteness of some moments of the MS-ARMA process.

### 5.5.3 Random Coefficient ARMA Processes

Regarding geometric ergodicity and strong mixing, everything becomes easier when leaving the truly Markovian MS-ARMA and turning to random coefficient ARMA processes, as considered in Nicholls and Quinn (1982) or Klüppelberg and Pergamenchtchikov (2004), for instance.

The crucial simplification is that for an i.i.d. sequence $\Delta$ the sequence $(\mathbf{X}_t)$ itself becomes a Markov chain, as already mentioned in the introduction to this section on geometric ergodicity. Originally, geometric ergodicity for random coefficient autoregressions has been studied in Feigin and Tweedie (1985). The following theorem is an adaptation of Basrak (2000, Prop. 3.2.9) to our set-up.

**Theorem 5.35** *Assume that $(\Delta_t)$ is an i.i.d. sequence, $\epsilon_t$ has a strictly positive density w.r.t. $\lambda^d$ (the Lebesgue measure on $\mathbb{R}^d$) and $\Sigma_t$ is invertible for all possible states of $\Delta$. If, moreover, there is an $\eta \in (0, 1]$ such that $\mathbf{C}_1 = \mathbf{m}_1 + \mathbf{\Sigma}_1\epsilon_1 \in L^\eta$ and*

$$E(\|\mathbf{A}_1\|^\eta) < 1 \tag{5.39}$$

*for some norm* $\|\cdot\|$ *on* $\mathbb{R}^{d(p+q)}$*, then* $(\mathbf{X}_t)$ *is geometrically ergodic.*

*Furthermore, when considering doubly infinite* $\Delta$ *and* $\epsilon$*, i.e.* $\Delta = (\Delta_t)_{t\in\mathbb{Z}}$ *and* $\epsilon = (\epsilon_t)_{t\in\mathbb{Z}}$*, all conditions of Theorem 5.2 are fulfilled. Moreover, the higher dimensional representation* $\mathbf{X}$ *of the MS-ARMA process as well as the MS-ARMA process* $X$ *itself are strongly mixing (with geometric rate).*

**Proof:** Let $g : \mathbb{R}^{d(p+q)} \to \mathbb{R}$ be bounded and continuous, then

$$E\left(g(\mathbf{X}_1)|\mathbf{X}_0 = x\right) = E\left(g\left(\mathbf{A}_1 x + \mathbf{m}_1 + \mathbf{\Sigma}_1\epsilon_1\right)\right)$$

is a continuous function of $x \in \mathbb{R}^{d(p+q)}$, as the continuity lemma from standard integration theory (see, for instance, Bauer (1992, Lemma 16.1)) shows. So, $(\mathbf{X}_t)$ is weakly Fellerian. Using the same arguments as in the proof of Proposition 5.31, we see immediately that $(\mathbf{X}_t)$ is $\lambda^{d(p+q)}$-irreducible.

Now we consider the continuous function $g : \mathbb{R}^{d(p+q)} \to \mathbb{R}^+$, $x \mapsto \|x\|^\eta + 1$ and note that we have $\|a+b\|^\eta \le \|a\|^\eta + \|b\|^\eta$ for all $a, b \in \mathbb{R}^{d(p+q)}$ as $0 < \eta \le 1$ (cf. Loève (1977, p. 157)). Thus, for any $x \in \mathbb{R}^{d(p+q)}$

$$\begin{aligned} E(\|\mathbf{X}_1\|^\eta + 1|\mathbf{X}_0 = x) &= E(\|\mathbf{A}_1 x + \mathbf{C}_1\|^\eta + 1) \\ &\le E(\|\mathbf{A}_1\|^\eta)\|x\|^\eta + E(\|\mathbf{C}_1\|^\eta) + 1. \end{aligned}$$

As $\mathbf{C}_1 \in L^\eta$, there is a finite $M > 0$ such that $E(\|\mathbf{C}_1\|^\eta) < M - 1$. Hence,

$$E(\|\mathbf{X}_1\|^\eta + 1|\mathbf{X}_0 = x) \le c\|x\|^\eta + M.$$

Choose now $\delta > 0$ with $1 - \delta > c$ and then set $R = \left(\frac{M}{1-\delta-c}\right)^{1/\eta}$ and $C = B_R(0)$ (the ball with radius $R$ in $\mathbb{R}^{d(p+q)}$). For all $x \in (B_R(0))^c$ we have that $(1-\delta-c)\|x\|^\eta \ge M$ and therefore

$$E(\|\mathbf{X}_1\|^\eta + 1|\mathbf{X}_0 = x) \le c\|x\|^\eta + (1-\delta-c)\|x\|^\eta = (1-\delta)\|x\|^\eta \le (1-\delta)g(x).$$

So, $C$ is a compact set that together with $g$ and $\delta$ satisfies $(ii)$ of Theorem 2.35. Thus Theorem 2.35 shows the claimed geometric ergodicity of $(\mathbf{X}_t)$.

Assume now that we consider doubly infinite sequences. The fact that $\mathbf{C}_1 \in L^\eta$ ensures $E(\log^+\|\mathbf{C}_0\|)$. Likewise, (5.39) gives $E(\log^+\|\mathbf{A}_0\|) < \infty$ and $\gamma < 0$. So, all conditions of Theorem 5.2 are satisfied.

The strong mixing properties stated are now established applying Propositions 2.37 and 2.38. □

## 5.6 Regularly Varying MS-ARMA Processes

In this section we study several cases when the stationary distribution of an MS-ARMA process is regularly varying. As when studying regular variation of the general stochastic recurrence equation $Y_n = A_n Y_{n-1} + C_n$ in Section 4.3, there are basically two different situations in which regular variation appears.

### 5.6.1 Regularly Varying Noise

Assume first that $(\mathbf{C}_t)_{t\in\mathbb{Z}}$ is an i.i.d. sequence independent of $(\mathbf{A}_t)_{t\in\mathbb{Z}}$. This holds, in particular, if the components $\mu_t$ and $\Sigma_t$ evolve independently of the rest of the Markov chain $(\Delta_t)$ as a joint i.i.d. sequence $(\mu_t, \Sigma_t)$. If, moreover, $\mathbf{C}_t$ is regularly varying, one can apply Theorems 4.12, 4.13 and Corollary 4.14. For the details of this case see thus Section 4.3.

In our eyes it is more interesting to study a regularly varying generic noise sequence $\epsilon = (\epsilon_t)$ and a general Markov chain $\Delta_t$, where $\Sigma_t$ is not independent of the other components as assumed above. For the sake of simplicity, we shall, however, assume $\mu_t = 0$ in the following. The results regarding regular variation obtained in the following can be extended to the case with general $\mu_t$ under an appropriate condition ensuring relative light-tailedness of $\sum_{k=0}^{\infty} \mathbf{A}_0\mathbf{A}_{-1}\cdots\mathbf{A}_{-k+1}\mathbf{m}_{-k}$ using Basrak (2000, Remark 2.1.20).

Noting that

$$\mathbf{X}_t = \sum_{k=0}^{\infty} \mathbf{A}_t\mathbf{A}_{t-1}\cdots\mathbf{A}_{t-k+1}\mathbf{\Sigma}_{t-k}\epsilon_{t-k}$$

Theorems 5.2 and 3.19 imply the following result.

**Theorem 5.36** *Let $\mu_t = 0$ for all $t \in \mathbb{Z}$ and $(\epsilon_t)_{t\in\mathbb{Z}}$ be a sequence of i.i.d. regularly varying $\mathbb{R}^d$-valued random variables with index $\alpha$, measure $\nu$ and normalizing sequence $(a_n)$ such that (iv) in Theorem 3.9 holds. Assume further that the Lyapunov exponent satisfies*

$$\gamma = \inf_{t\in\mathbb{N}_0} \left( \frac{1}{t+1} E\left(\log \|\mathbf{A}_0\mathbf{A}_{-1}\cdots\mathbf{A}_{-t}\|\right)\right) < 0.$$

*If $\alpha < 1$, assume that there is an $\eta$ with $0 < \eta < \alpha$ and $\alpha + \eta < 1$ such that $\mathbf{A}_0\cdots\mathbf{A}_{-k+1}\mathbf{\Sigma}_{-k} \in L^{\alpha+\eta}$ for all $k \in \mathbb{N}_0$ and that*

$$\sum_{k=0}^{\infty} E\left(\|\mathbf{A}_0\cdots\mathbf{A}_{-k+1}\mathbf{\Sigma}_{-k}\|^{\alpha+\eta}\right) < \infty \text{ as well as } \sum_{k=0}^{\infty} E\left(\|\mathbf{A}_0\cdots\mathbf{A}_{-k+1}\mathbf{\Sigma}_{-k}\|^{\alpha-\eta}\right) < \infty. \tag{5.40}$$

*If $\alpha \geq 1$, assume that there is an $\eta$ with $0 < \eta < \alpha$ such that $\mathbf{A}_0\cdots\mathbf{A}_{-k+1}\mathbf{\Sigma}_{-k} \in L^{\alpha+\eta}$ for all $k \in \mathbb{N}_0$ and that*

$$\begin{aligned} \sum_{k=0}^{\infty} E\left(\|\mathbf{A}_0\cdots\mathbf{A}_{-k+1}\mathbf{\Sigma}_{-k}\|^{\alpha+\eta}\right)^{1/(\alpha+\eta)} &< \infty \text{ as well as} \\ \sum_{k=0}^{\infty} E\left(\|\mathbf{A}_0\cdots\mathbf{A}_{-k+1}\mathbf{\Sigma}_{-k}\|^{\alpha-\eta}\right)^{1/(\alpha+\eta)} &< \infty. \end{aligned} \tag{5.41}$$

*Then the MS-ARMA$(p, q, \Delta, \epsilon)$ equation (5.3) has a unique stationary and ergodic solution. The unique stationary solution $X = (X_t)$ is formed by the first $d$ coordinates of*

$$\mathbf{X}_t = \sum_{k=0}^{\infty} \mathbf{A}_t\mathbf{A}_{t-1}\cdots\mathbf{A}_{t-k+1}\mathbf{\Sigma}_{t-k}\epsilon_{t-k},$$

*which is the unique stationary and ergodic solution of (5.12). The series defining* $\mathbf{X}$ *converges absolutely a.s.*

*Moreover, the tail behaviour of* $\mathbf{X}_0$ *(and thus of the "one"-dimensional marginal distribution of* $\mathbf{X} = (\mathbf{X}_t)$*, i.e. the higher dimensional representation of the solution to the MS-ARMA*$(p, q, \Delta, \epsilon)$ *equation) is given by*

$$nP(\mathbf{X}_0 \in a_n \cdot) \xrightarrow{v} \tilde{\nu}(\cdot) = \sum_{k=0}^{\infty} E\left(\nu \circ \left(\mathbf{A}_0 \cdots \mathbf{A}_{-k+1}\mathbf{\Sigma}_{-k}\right)^{-1}(\cdot)\right). \tag{5.42}$$

*For* $X_0$ *(and thus for the "one"-dimensional marginal distribution of* $X = (X_t)$*, i.e. the solution to the MS-ARMA*$(p, q, \Delta, \epsilon)$ *equation) it holds that*

$$nP(X_0 \in a_n \cdot) \xrightarrow{v} \bar{\nu}(\cdot) = \sum_{k=0}^{\infty} E\left(\nu \circ \left(\mathbb{I}\mathbf{A}_0 \cdots \mathbf{A}_{-k+1}\mathbf{\Sigma}_{-k}\right)^{-1}(\cdot)\right), \tag{5.43}$$

*where* $\mathbb{I} := (I_d, 0, \ldots, 0) \in M_{d,(p+q)d}(\mathbb{R})$.

*Provided there is a relatively compact* $K \in \overline{\mathbb{R}^{d(p+q)}}$ *with* $E\left(\nu \circ \mathbf{\Sigma}_0^{-1}(K)\right) > 0$, $\mathbf{X}_0$ *and* $X_0$ *are regularly varying with common index* $\alpha$*, normalizing sequence* $(a_n)$ *and measure* $\tilde{\nu}$*, respectively* $\bar{\nu}$.

*Furthermore, if* $\epsilon_0 \in L^\alpha$*, then* $\mathbf{X}_0$ *and* $X_0$ *are in* $L^\alpha$.

**Proof:** All assertions regarding $\mathbf{X}$ follow by combining Theorems 5.2, 3.19 and the arguments given for Theorem 4.12. (5.43) follows by another application of Theorem 3.16 or by considering

$$X_t = \sum_{k=0}^{\infty} \mathbb{I}\mathbf{A}_t\mathbf{A}_{t-1} \cdots \mathbf{A}_{t-k+1}\mathbf{\Sigma}_{t-k}\epsilon_{t-k}.$$

The nondegeneracy of $\bar{\nu}$ under the assumption $E\left(\nu \circ \mathbf{\Sigma}_0^{-1}(K)\right) > 0$ for some relatively compact $K$ is immediate, noting that $E\left(\nu \circ \mathbf{\Sigma}_0^{-1}(A_1 \times A_2 \times \cdots \times A_{p+q})\right) = E(\nu \circ \Sigma_0^{-1}(A_1 \cap A_{p+1}))\epsilon_{0_{\mathbb{R}^{d(p+q-2)}}}(A_2 \times \cdots \times A_{p-1} \times A_{p+2} \times \cdots A_{p+q})$ for $A_i \in \mathcal{B}(\mathbb{R}^d)$. □

Along the lines of reasoning that lead to Theorem 4.13 we obtain:

**Theorem 5.37** *If all conditions of Theorem 5.36 are satisfied, then* $\mathbf{X} = (\mathbf{X}_t)_{t\in\mathbb{Z}}$ *as well as* $X = (X_t)_{t\in\mathbb{Z}}$ *are even regularly varying as a sequence with index* $\alpha$.

Obviously we can again apply Lemmata 3.21 and 4.5 to obtain a version of the above theorem that will suffice in most cases encountered.

**Corollary 5.38** *Let* $\mu_t = 0$ *for all* $t \in \mathbb{Z}$ *and* $(\epsilon_t)_{t\in\mathbb{Z}}$ *be a sequence of i.i.d. regularly varying* $\mathbb{R}^d$*-valued random variables with index* $\alpha$*, measure* $\nu$ *and normalizing sequence* $(a_n)$ *such that (iv) in Theorem 3.9 holds.*

*Assume that there is a* $\beta > \alpha$ *such that* $\mathbf{A}_0 \cdots \mathbf{A}_{-k+1}\mathbf{\Sigma}_{-k} \in L^\beta$ *and* $\mathbf{A}_0 \cdots \mathbf{A}_{-k+1} \in L^\beta$ *for all* $k \in \mathbb{N}_0$ *and that*

$$\limsup_{n\to\infty} E\left(\|\mathbf{A}_0 \cdots \mathbf{A}_{-n+1}\mathbf{\Sigma}_{-n}\|^\beta\right)^{1/(n+1)} < 1 \tag{5.44}$$

*as well as*

$$\limsup_{n\to\infty} E\left(\|\mathbf{A}_0\cdots\mathbf{A}_{-n+1}\|^\beta\right)^{1/n} < 1. \tag{5.45}$$

*Then the MS-ARMA$(p,q,\Delta,\epsilon)$ equation (5.3) has a unique stationary and ergodic solution. The unique stationary solution $X=(X_t)$ is formed by the first $d$ coordinates of*

$$\mathbf{X}_t = \sum_{k=0}^{\infty} \mathbf{A}_t\mathbf{A}_{t-1}\cdots\mathbf{A}_{t-k+1}\boldsymbol{\Sigma}_{t-k}\epsilon_{t-k},$$

*which is the unique stationary and ergodic solution of (5.12). The series defining $\mathbf{X}$ converges absolutely a.s.*

*Moreover, the tail behaviour of $\mathbf{X}_0$ (and thus of the "one"-dimensional marginal distribution of $\mathbf{X}=(\mathbf{X}_t)$, i.e. the higher dimensional representation of the solution to the MS-ARMA$(p,q,\Delta,\epsilon)$ equation) is given by*

$$\tilde{\nu}(\cdot) = nP(\mathbf{X}_0 \in a_n\cdot) \xrightarrow{v} \sum_{k=0}^{\infty} E\left(\nu\circ(\mathbf{A}_0\cdots\mathbf{A}_{-k+1}\boldsymbol{\Sigma}_{-k})^{-1}(\cdot)\right). \tag{5.46}$$

*For $X_0$ (and thus for the "one"-dimensional marginal distribution of $X=(X_t)$, i.e. the solution to the MS-ARMA$(p,q,\Delta,\epsilon)$ equation) it holds that*

$$\bar{\nu}(\cdot) = nP(X_0 \in a_n\cdot) \xrightarrow{v} \sum_{k=0}^{\infty} E\left(\nu\circ(\mathbb{I}\mathbf{A}_0\cdots\mathbf{A}_{-k+1}\boldsymbol{\Sigma}_{-k})^{-1}(\cdot)\right), \tag{5.47}$$

*where $\mathbb{I} := (I_d, 0, \ldots, 0) \in M_{d,(p+q)d}(\mathbb{R})$.*

*Provided there is a relatively compact $K \in \overline{\mathbb{R}^{d(p+q)}}$ with $E\left(\nu\circ\boldsymbol{\Sigma}_0^{-1}(K)\right) > 0$, $\mathbf{X}_0$ and $X_0$ are regularly varying with common index $\alpha$, normalizing sequence $(a_n)$ and measure $\tilde{\nu}$, respectively $\bar{\nu}$. $(\mathbf{X}_t)$ and $(X_t)$ are also regularly varying as a sequence with index $\alpha$.*

*Furthermore, if $\epsilon_0 \in L^\alpha$, then $\mathbf{X}_0$ and $X_0$ are in $L^\alpha$.*

A considerable simplification occurs, if $\boldsymbol{\Sigma}_{-k}$ is independent of $\mathbf{A}_0\cdots\mathbf{A}_{-k+1}$ for all natural $k$. In this case $\boldsymbol{\Sigma}_0 \in L^\beta$ and $\mathbf{A}_0\cdots\mathbf{A}_{-k+1} \in L^\beta$ for all natural $k$ give $\mathbf{A}_0\cdots\mathbf{A}_{-k+1}\boldsymbol{\Sigma}_{-k} \in L^\beta$ $\forall k \in \mathbb{N}_0$ and then (5.45) implies (5.44).

Note that for one-dimensional positive valued random coefficient autoregressive models similar results were already obtained in Resnick and Willekens (1991).

The above results show, in particular, that a Markov switching ARMA process is tail equivalent to its driving noise sequence under appropriate conditions. So, provided the upper tails are nondegenerate, for $d=1$ the distribution of $\epsilon_0$ and $X_0$ both belong to the maximum domain of attraction of the Fréchet distribution $\Phi_\alpha$. Moreover, note that almost all results of Sections 5.2.3 and 5.3 can be used or straightforwardly adapted to verify (5.44) or (5.45). In particular, we immediately obtain the following using the arguments that led to Theorem 5.21.

**Lemma 5.39** *Assume that there are $c<1$, $C, M \in \mathbb{R}^+$ and a norm $\|\cdot\|$ on $\mathbb{R}^d$ such that $\sum_{i=1}^p \|\Phi_{i0}\| \le c$, $\sum_{i=1}^q \|\Theta_{i0}\| \le M$ and $\|\Sigma_0\| \le C$ a.s. Then (5.44) and (5.45) are satisfied.*

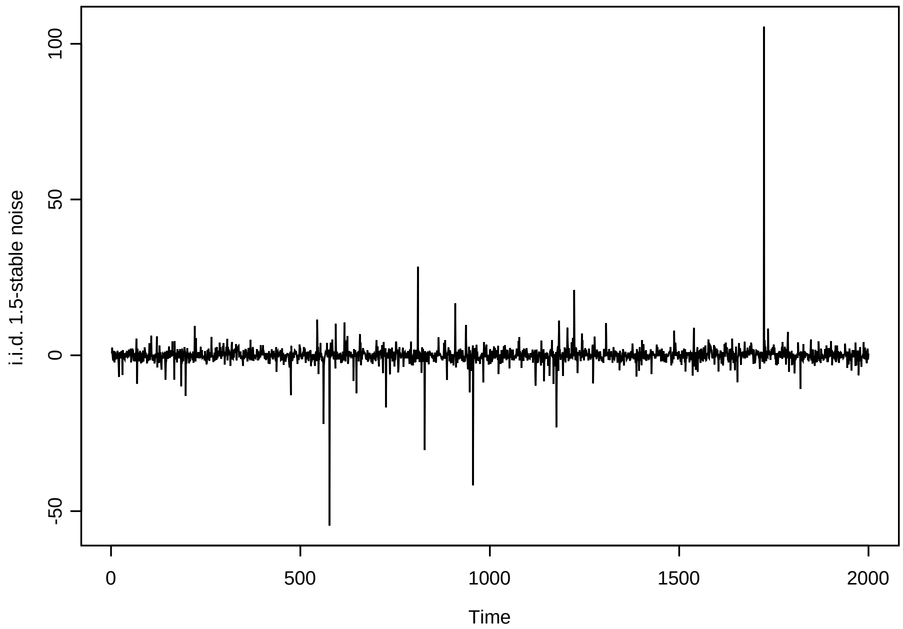


Figure 5.1: Simulation of an i.i.d. symmetric 1.5-stable noise sequence

Note, moreover, that the nondegeneracy condition, viz. that there exists a relative compact $K$ with $E\left(\nu \circ \mathbf{\Sigma}_0^{-1}(K)\right) > 0$, is only a minor technical condition, since it suffices for the latter that $\Sigma_0$ has a strictly positive probability of being invertible (If $\Sigma_0$ is invertible, we have that $\Sigma_0^{-1}(B_0(1)) \subseteq \|\Sigma_0^{-1}\| B_0(1)$, hence $\Sigma_0^{-1}\left((1,\infty]\mathbb{S}^{d-1}\right) = \left(\Sigma_0^{-1}(B_0(1))\right)^c \supseteq (\|\Sigma_0^{-1}\|,\infty]\mathbb{S}^{d-1}$ and thus $\nu \circ \mathbf{\Sigma}_0^{-1}\left((1,\infty]\mathbb{S}^{d-1} \times 0_{\mathbb{R}^{d(p-1)}} \times (1,\infty]\mathbb{S}^{d-1} \times 0_{\mathbb{R}^{d(q-1)}}\right) = \nu \circ \Sigma_0^{-1}\left((1,\infty]\mathbb{S}^{d-1}\right) > 0$ due to the nondegeneracy of $\nu$.).

It is very interesting to compare our above obtained results with those from Brachner (2004) for TAR models in $\mathbb{R}$ with regularly varying noise. For the general TAR(q) O-regular variation is obtained employing considerably stricter conditions than we give in the above lemma to ensure regular variation of MS-ARMA(p,q) processes. Results comparable to ours are, however, obtained for TAR(1) models with only two possible regimes. In this case the sufficient conditions for regular variation and tail equivalence given in Brachner (2004) are comparable to the assumptions of Lemma 5.39. The crucial difference seems to be that due to the regime selection procedure powerful regular variation results like Theorem 3.19 cannot be used for TAR-models.

To conclude this section let us give some simulation examples done with the S-Plus software. We shall consider real-valued MS-AR(1) processes with $\mu_t = 0$ and $\Sigma_t = 1$, i.e. our model is given by $Y_t = \Phi_{1t} Y_{t-1} + \epsilon_t$. As noise we shall consider an i.i.d. sequence $\epsilon_t$ that has a symmetric 1.5-stable distribution, cf. Figure 5.1 for an example of such a sequence. In particular, this noise is nondegenerately regularly varying in both tails and the index of regular variation is 1.5. As in Section 5.4.1 we presume that there are two

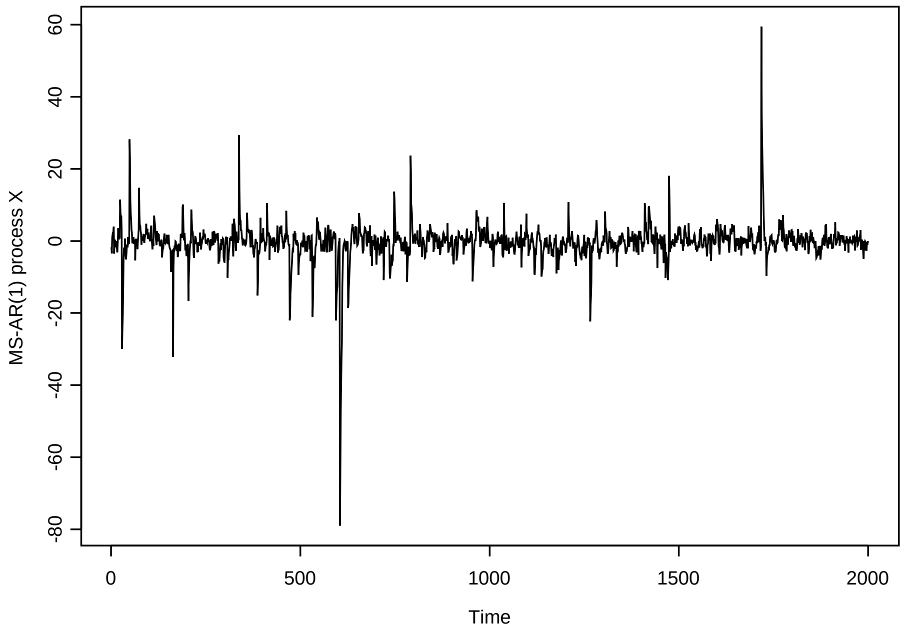


Figure 5.2: Simulation of the MS-AR(1) process in Example 1

possible states $\mathbf{\Phi}^{(1)}$ and $\mathbf{\Phi}^{(2)}$ and that the transition matrix is given by

$$P = \begin{pmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{pmatrix} = \begin{pmatrix} \bar{p} & 1-\bar{p} \\ 1-\bar{p} & \bar{p} \end{pmatrix}$$

for some $\bar{p} \in [0,1)$. Then the stationary distribution is $(\pi^{(1)}, \pi^{(2)}) = (1/2, 1/2)$. The strict stationarity of the following examples has already been established in Section 5.4.1 or is immediate, because $\mathbf{\Phi}^{(1)}$ and $\mathbf{\Phi}^{(2)}$ are both less than one in modulus.
**Example 1:** Take $\bar{p} = 3/4$, $\mathbf{\Phi}^{(1)} = 1/2$ and $\mathbf{\Phi}^{(2)} = 4/5$. Then all conditions of Theorem 5.36 or Corollary 5.38 are obviously satisfied. The simulation in Figure 5.2 illustrates the fact that the MS-AR process is regularly varying (with index 1.5) and the stationary distribution is tail equivalent to the noise, as a comparison with Figure 5.1 shows.
**Example 2:** Take $\bar{p} = 3/4$, $\mathbf{\Phi}^{(1)} = 1/2$ and $\mathbf{\Phi}^{(2)} = 11/10$. As one calculates $\rho(Q_{\mathbf{A}}) = (219 + \sqrt{23761})/400 \approx 0.9328650868$ for the condition from Theorem 5.23, one obtains from this theorem that (5.45) holds with $\beta = 2$. Thus, all conditions of Theorem 5.36 or Corollary 5.38 are obviously satisfied. Again, the simulation in Figure 5.3 illustrates the fact that the MS-AR process is regularly varying (with index 1.5) and the stationary distribution is tail equivalent to the noise, as a comparison with Figure 5.1 shows.

The importance of this example lies in the fact that $\mathbf{\Phi}^{(2)}$ is greater than one. Thus, it shows that Theorem 5.36 or Corollary 5.38 are not only applicable in the case of Lemma 5.39 in practice, but even in the presence of an explosive regime.

Finally, observe that in both above examples one deducts immediately from (5.46) that both tails of the stationary distribution of the MS-AR(1) process are nondegenerately

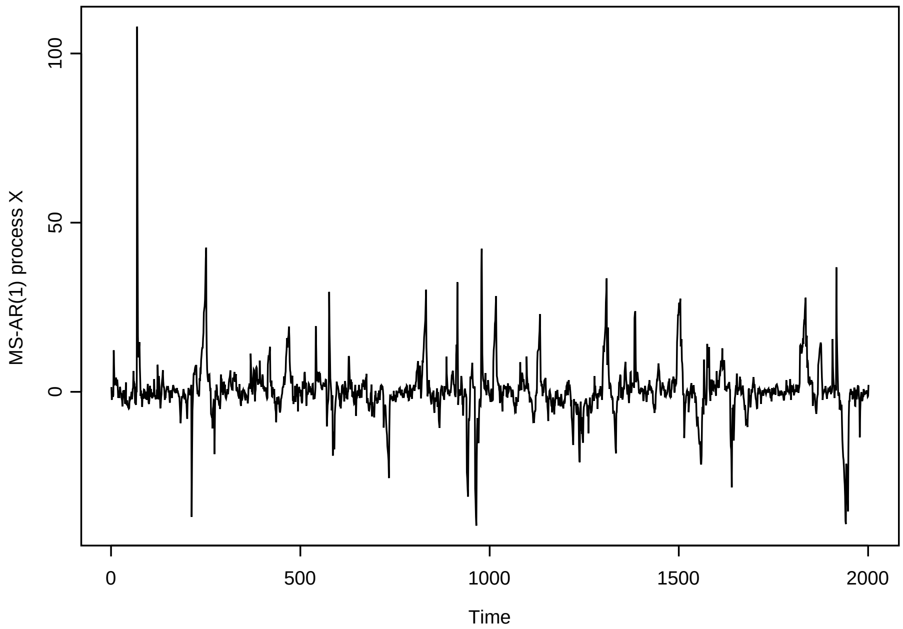


Figure 5.3: Simulation of the MS-AR(1) process in Example 2

regularly varying, since this holds for the noise $\epsilon_t$.

### 5.6.2 Light-Tailed Noise

As in the general case of Section 4.3.2, light-tailed noise may lead to regularly varying MS-ARMA processes provided the sequence $(\mathbf{A}_t)$ satisfies appropriate technical conditions ensuring the appearance of consecutive "large" values. Observe that again the noise can be arbitrarily light-tailed in all cases below. In the following we summarize the results of Klüppelberg and Pergamenchtchikov (2004) for random coefficient AR(p) models ($\mathbb{R}$-valued MS-ARMA(p,0) processes with $\Sigma_t = 1$, $\mu_t = 0$ and $(\mathbf{A}_t) = (\mathbf{\Phi}_t)$ being an i.i.d. sequence) and of Saporta (2004b) for one-dimensional Markov-switching AR(1) processes ($\mathbb{R}$-valued MS-ARMA(1,0) processes with $\Sigma_t = 1$, $\mu_t = 0$ and a finite state space of $(\mathbf{A}_t)$). Observe that it is also possible to employ the results of Section 4.3.2 directly on $\mathbf{X}_t = \mathbf{A}_t\mathbf{X}_{t-1} + \mathbf{C}_t$.

All these results indicate that under appropriately adjusted technical conditions light-tailed noise may well also lead to regularly varying MS-ARMA processes in the general MS-ARMA(p,0) case with coefficients that follow a general Markov chain. However, proving this appears to be very involved. In particular, neither the approach by Klüppelberg and Pergamenchtchikov (2004) seems to be rather straightforwardly adaptable to a Markovian dependence structure nor the one by Saporta (2004b) to orders $p > 1$ or a non-finite state space of $(\mathbf{A}_t)$, as in all cases powerful new renewal theorems appear to be necessary.

**Random Coefficient AR(p) Processes**

Klüppelberg and Pergamenchtchikov (2004) considered the one-dimensional, i.e. $\mathbb{R}$-valued, special case of our general MS-ARMA(p,0) model, where, moreover, the autoregressive coefficients $(\Phi_{it})_{i=1,\ldots,p}$, resp. $(\mathbf{A}_t) = (\mathbf{\Phi}_t)$, are an i.i.d. sequence, $\Sigma_t = 1$ and $\mu_t = 0$. The coefficients $\Phi_{1t}, \Phi_{2t}, \ldots, \Phi_{pt}$ are in the following assumed to be given by

$$\Phi_{it} = \bar{\Phi}_{it} + \sigma_i \eta_{it} \text{ for } i = 1, \ldots, p,$$

where $(\bar{\Phi}_{1t}, \ldots, \bar{\Phi}_{pt})^\mathsf{T} \in \mathbb{R}^p$, $(\sigma_1, \ldots, \sigma_p)^\mathsf{T} \in (\mathbb{R}^+)^p$ are constants and $\eta_t = (\eta_{1t}, \ldots, \eta_{pt})^\mathsf{T}$ is an $\mathbb{R}^p$-valued i.i.d. sequence. Moreover, we presume that the sequences $\eta = (\eta_t)_{t\in\mathbb{Z}}$ and $\epsilon = (\epsilon_t)_{t\in\mathbb{Z}}$ are independent. Furthermore they are centred and of unital variance, i.e. $E(\epsilon_0) = E(\eta_{i0}) = 0$ and $E(\epsilon_0^2) = E(\eta_{i0}^2) = 1$. The following is a reformulation of the results given in Klüppelberg and Pergamenchtchikov (2004) applying to stationary random coefficient AR(p) processes.

**Theorem 5.40 (Klüppelberg and Pergamenchtchikov (2004, Th. 2.4))** *Let a random coefficient AR(p) model with the above described properties be given and assume that moreover the following conditions are satisfied:*

*(i)* $\rho\left(E(\mathbf{A}_0 \otimes \mathbf{A}_0)\right) < 1$.

*(ii) The random variables* $\{\eta_{it}, i = 1, 2, \ldots, p \,;\, t \in \mathbb{Z}\}$ *are i.i.d., have a symmetric (around 0), continuous and positive density* $\phi(\cdot)$*, which is non-increasing on* $\mathbb{R}^+$*, and moments of all orders exist, i.e.* $\eta_{it} \in L^m \; \forall\, m \in \mathbb{N}$.

*(iii) There is an* $m \in \mathbb{N}$ *such that* $E\left((\Phi_{1,0} - \bar{\Phi}_1)^{2m}\right) = \sigma_1^{2m} E\left(\eta_{1,0}^{2m}\right) \in (1, \infty)$.

*(iv)* $\epsilon_0 \in L^m$ *for all* $m \in \mathbb{N}$.

*(v) For any non-zero real sequence* $(c_k)_{k\in\mathbb{N}} \in \ell_1$*, i.e.* $0 < \sum_{k=1}^{\infty} |c_k| < \infty$*, the random variable* $\tau = \sum_{k=1}^{\infty} c_k \epsilon_k$ *has a symmetric density that is non-increasing on* $\mathbb{R}^+$.

*(vi)* $\bar{\Phi}_p^2 + \sigma_p^2 > 0$.

*Then there exists a unique stationary solution to the random coefficient AR(p) equation (5.3), which is given by (5.13) employing the usual higher dimensional representation* $\mathbf{X}_t$*. Furthermore, there exists a unique positive solution* $\lambda_0$ *to the equation*

$$\kappa(\lambda) := \lim_{n\to\infty} \left(E\left(\|\mathbf{A}_1 \cdots \mathbf{A}_n\|_t^\lambda\right)\right)^{1/n} = 1$$

*and some probability measure* $\nu$ *on* $\mathbb{S}^{p-1}$ *(w.r.t. to the Euclidean norm* $\|\cdot\|_2$*) such that*

$$\kappa(\lambda) = \int_{\mathbb{S}^{p-1}} E(\|x^\mathsf{T}\mathbf{A}_1\|_2^\lambda)\nu(dx).$$

*It holds that* $\lambda_0 > 2$ *and with some strictly positive and continuous function* $w : \mathbb{S}^{p-1} \to \mathbb{R}$ *the random variable* $\mathbf{X}_0$ *has the regular variation property*

$$\lim_{t\to\infty} t^{\lambda_0} P(\langle x, \mathbf{X}_0\rangle > t) = w(x) \;\forall\, x \in \mathbb{S}^{p-1},$$

*i.e.* $\mathbf{X}_0$ *is regularly varying in the sense of Kesten. In particular, the univariate marginal distribution of the stationary random coefficient AR(p) process* $X$ *is regularly varying with index* $\lambda_0$.

All densities above are understood to be with respect to the Lebesgue measure and recall from Section 5.3.2 that $\|\cdot\|$ is the Froebenius norm given by $\|A\|_t = \sqrt{tr(AA^\intercal)}$. Condition $(i)$ ensures strict negativity of the Lyapunov exponent $\gamma$, $(iii)$ implies, in particular, that $\sigma_1 > 0$ needs to hold and the very last one gives that $\mathbf{A}_t$ is a.s. invertible. As condition $(v)$ appears to be rather complicated to verify, the following Lemma is very useful.

**Lemma 5.41** *If* $\epsilon_1$ *has a bounded, symmetric and continuously differentiable density* $f$ *and* $f'$ *is bounded and non-positive on* $[0, \infty)$*, then condition* $(v)$ *of the last theorem is satisfied.*

Provided the conditions of Theorem 3.11 (ii), (iii) or (iv) are satisfied, the above theorem implies the regular variation of $\mathbf{X}_0$ and one may deduct again using Basrak, Davis and Mikosch (2002b, Cor. 2.7) that the process $(\mathbf{X}_t)_{t\in\mathbb{Z}}$ is regularly varying as a sequence. The results of Klüppelberg and Pergamenchtchikov (2004) were also briefly discussed in Saporta (2004a).

### MS-AR(1) Processes

Naturally the above result raises the question whether something similar is obtainable under a Markovian dependence structure of the AR coefficients. The only paper on this subject is, as far as we know, Saporta (2004b) where the $\mathbb{R}$-valued AR(1) case is studied. The renewal theory necessary to prove the results has been developed in Saporta (2003).

In this section we now consider the MS-AR(1) model given by:

$$X_t = A_t X_{t-1} + \epsilon_t \tag{5.48}$$

where $(A_t)_{t\in\mathbb{Z}}$ is an $\mathbb{R}$-valued finite state space Markov chain and $\epsilon = (\epsilon_t)_{t\in\mathbb{Z}}$ an i.i.d. sequence of real random variables independent of $A = (A_t)$. The number of states, which $A_t$ can assume, is denoted by $m$ and $A^{(1)}, A^{(2)}, \ldots, A^{(m)}$ are as previously the individual possible states, which are all assumed to be different from zero. The state space is $E := \{A^{(1)}, A^{(2)}, \ldots, A^{(m)}\}$ and $P$ the transition probability matrix. As the necessary assumptions for a stationary solution with regularly varying tails are different dependent on whether $E \subset \mathbb{R}^+$ or not, we now state the results as two separate theorems. One of the main conditions is formulated in terms of the matrix

$$P_\lambda := diag\left(\left|A^{(i)}\right|^\lambda\right) P^\intercal = \begin{pmatrix} |A^{(1)}|^\lambda p_{11} & |A^{(1)}|^\lambda p_{21} & \cdots & |A^{(1)}|^\lambda p_{m1} \\ |A^{(2)}|^\lambda p_{12} & |A^{(2)}|^\lambda p_{22} & \cdots & |A^{(2)}|^\lambda p_{m2} \\ \vdots & \ddots & \ddots & \vdots \\ |A^{(m)}|^\lambda p_{1m} & |A^{(m)}|^\lambda p_{2m} & \cdots & |A^{(m)}|^\lambda p_{mm} \end{pmatrix}$$

with $\lambda \in \mathbb{R}^+$.

**Theorem 5.42 (Saporta (2004b, Th. 1))** *Consider the above given set-up and let the Markov chain* $(A_t)$ *be irreducible, aperiodic and stationary with state space* $E \subset \mathbb{R}^+\backslash\{0\}$. *Assume that the following conditions are satisfied:*

*(i)* $E(\log A_0) < 0$ *and* $E(\log^+ |\epsilon_0|) < \infty$.

*(ii)* *There is a* $\lambda_0 > 0$ *such that* $\rho(P_{\lambda_0}) = 1$.

*(iii)* *The numbers* $\log(A^{(i)})$, $i = 1, 2, \ldots, m$, *are not integral numbers of the same number, i.e. there is no* $r \in \mathbb{R}$ *such that* $\{\log(A^{(i)})\}_{i=1,2,\ldots,m} \subset r\mathbb{Z}$.

*(iv)* *There is a* $\delta > 0$ *such that* $\epsilon_0 \in L^{\lambda_0+\delta}$.

*Then the unique stationary solution of (5.48) is formed by*

$$X_t = \sum_{k=0}^{\infty} A_t A_{t-1} \cdots A_{t-k+1} \epsilon_{t-k}$$

*and the following regular variation property holds for all* $s \in \mathbb{S}^0 = \{-1, 1\}$:

$$\lim_{t\to\infty} t^{\lambda_0} P(sX_0 > t) = L(s)$$

*where* $L(1), L(-1) \geq 0$ *and* $L(1) + L(-1) > 0$. *In particular, at least one tail of the stationary distribution of the MS-AR(1) process is regularly varying with index* $\lambda_0$.

*If* $\epsilon_0 \geq 0$ *a.s., then* $L(-1) = 0$, $L(1) > 0$ *and, vice versa, if* $\epsilon_0 \leq 0$ *a.s., then* $L(1) = 0$, $L(-1) > 0$.

Observe that

$$\|P_\lambda\|_1 = \max_{j=1,2,\ldots,m} \left( \sum_{i=1}^{m} \left(A^{(i)}\right)^\lambda p_{ji} \right) \leq \sum_{i=1}^{m} p_{ji} \max_{i=1,2,\ldots,m} \left( \left(A^{(i)}\right)^\lambda \right) = \max_{i=1,2,\ldots,m} \left( \left(A^{(i)}\right)^\lambda \right)$$

and, since $\rho(P_\lambda) \leq \|P_\lambda\|_1$, *(ii)* and $E(\log A_0) < 0$ thus imply that there is at least one state $A^{(i_0)} > 1$, i.e. one state that is explosive.

In order to study the general case a concept called $\ell$-irreduciblibility is introduced.

**Definition 5.43 (Saporta (2004b, Def. 3))** *Let* $A = (a_{ij}) \in M_d(\mathbb{R})$ *be a matrix with non-negative entries* $a_{ij}$ *and* $0 \leq \ell \leq d-1$ *be an integer number. A is said to be* $\ell$-reducible, *if there is a (possibly trivial) partition* $(I, J)$ *of the set* $\{1, 2, \ldots, d\}$ *such that:*

- *For all* $1 \leq i \leq \ell$: $\begin{array}{l} i \in I \Rightarrow a_{ij} = 0 \,\forall\, j \in J \\ i \in J \Rightarrow a_{ij} = 0 \,\forall\, j \in I \end{array}$

- *For all* $\ell + 1 \leq i \leq d$: $\begin{array}{l} i \in I \Rightarrow a_{ij} = 0 \,\forall\, j \in I \\ i \in J \Rightarrow a_{ij} = 0 \,\forall\, j \in J \end{array}$.

*If A is not* $\ell$*-reducible, then it is called* $\ell$-irreducible.

For a result linking the above concept to the standard one of irreducibility of matrices see Saporta (2004b, Proposition 6).

**Theorem 5.44 (Saporta (2004b, Th. 1))** *Consider the above MS-AR(1) set-up and let the Markov chain $(A_t)$ be irreducible, aperiodic and stationary with state space $E = \{A^{(1)}, A^{(2)}, \ldots, A^{(m)}\} \subset \mathbb{R}$. Let the elements of $E$ be sorted such that there is an $\ell \in \{0, 1, \ldots, m-1\}$ with $A^{(1)}, A^{(2)}, \ldots, A^{(\ell)} > 0$ and $A^{(\ell+1)}, A^{(\ell+2)}, \ldots, A^{(m)} < 0$ ($\ell = 0$ means that all possible states of $A_t$ are negative). Assume that the following conditions are satisfied:*

*(i)* $E(\log |A_0|) < 0$ *and* $E(\log^+ |\epsilon_0|) < \infty$.

*(ii)* *There is a* $\lambda_0 > 0$ *such that* $\rho(P_{\lambda_0}) = 1$.

*(iii)* *The numbers* $\log\left(\left|A^{(i)}\right|\right)$, $i = 1, 2, \ldots, m$, *are not integral numbers of the same number, i.e. there is no* $r \in \mathbb{R}$ *such that* $\{\log\left(\left|A^{(i)}\right|\right)\}_{i=1,2,\ldots,m} \subset r\mathbb{Z}$.

*(iv)* *There is a* $\delta > 0$ *such that* $\epsilon_0 \in L^{\lambda_0+\delta}$.

*Then the unique stationary solution of (5.48) is formed by*

$$X_t = \sum_{k=0}^{\infty} A_t A_{t-1} \cdots A_{t-k+1} \epsilon_{t-k}$$

*and the following regular variation property holds for all* $s \in \mathbb{S}^0 = \{-1, 1\}$:

$$\lim_{t\to\infty} t^{\lambda_0} P(sX_0 > t) = L(s)$$

*where* $L(1), L(-1) \geq 0$ *and* $L(1) + L(-1) > 0$. *In particular, at least one tail of the stationary distribution of the MS-AR(1) process is regularly varying with index* $\lambda_0$.

*If* $P^\top$ *is* $\ell$*-irreducible, then* $L(1) = L(-1) > 0$ *and thus both tails are regularly varying with index* $\lambda_0$.

Observe that again condition $(ii)$ implies that there has to be at least one explosive state $|A^{(i_0)}| > 1$. Moreover, it is noteworthy that the $\ell$-irreducibility is only needed to ensure non-degenerate regular variation in both tails.

Unfortunately, we have not been able to extend the above results to regular variation of $(X_t)$ as a sequence. An approach motivated by Basrak, Davis and Mikosch (2002b, Cor. 2.7) fails, since we lack the necessary independence, and one can neither use similar arguments as in the proof of Theorem 4.13, as this would mean that one has to leave the one-dimensional setting. The crucial point, why it appears to be hard to extend the results of Saporta (2004b) to higher orders, resp. the multivariate case, seems to be that $\mathbb{S}^0$ is a finite set, whereas $\mathbb{S}^{d-1}$ is uncountable for all $d \geq 2$.

Let us conclude this section on regular variation in the presence of (relatively) light-tailed noise with some simulations. As previously, we presume that there are only two possible states $A^{(1)}$ and $A^{(2)}$ and that the transition matrix is given by

$$P = \begin{pmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{pmatrix} = \begin{pmatrix} \bar{p} & 1-\bar{p} \\ 1-\bar{p} & \bar{p} \end{pmatrix}$$

for some $\bar{p} \in (0, 1)$. Then the stationary distribution is $(\pi^{(1)}, \pi^{(2)}) = (1/2, 1/2)$ and the Markov chain is irreducible and aperiodic. $E(\log |A_0|) < 0$ and thus the strict stationarity

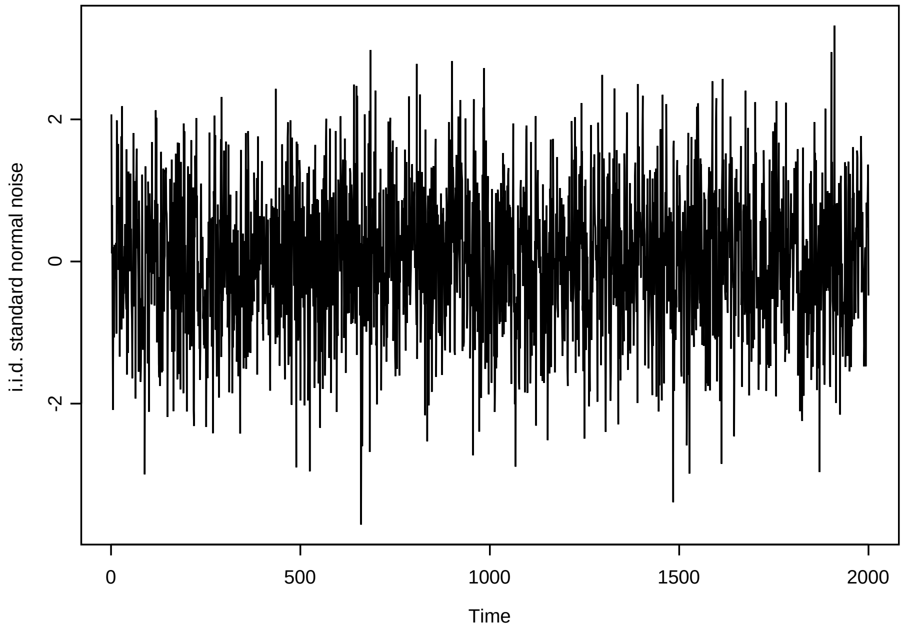


Figure 5.4: Simulation of an i.i.d. standard normal noise sequence

of the following examples has already been established in Section 5.4.1 or is immediate, because $A^{(1)}$ and $A^{(2)}$ are both less than one in modulus. First we shall consider a standard normal noise sequence $\epsilon$, as then all conditions of Theorems 5.42 and 5.44 on $\epsilon$ are always satisfied. Figure 5.4 depicts a simulation of such a noise sequence.

**Example 1:** Take $\bar{p} = 3/4$, $A^{(1)} = 1/2$ and $A^{(2)} = 4/5$. Then the conditions of Theorem 5.42 cannot be satisfied, as there is no explosive regime. The simulation in Figure 5.5 also shows no signs of regularly varying tails.

**Example 2:** Take $\bar{p} = 3/4$, $\mathbf{\Phi}^{(1)} = 1/2$ and $\mathbf{\Phi}^{(2)} = 11/10$. For condition $(ii)$ in Theorem 5.42 numerical calculations show that it holds with $\lambda_0 = 2.88775$. So, Theorem 5.42 gives that the MS-AR(1) process is regularly varying with index $\lambda_0$ provided $(iii)$ is also satisfied. Unfortunately, we have not been able to find a feasible way of actually checking $(iii)$. Yet, the simulation in Figure 5.6 seems to indicate that the MS-AR(1) process is indeed heavy-tailed and probably regularly varying. Recall from Section 5.4.1 that the sufficient second-order stationarity condition from Francq and Zakoïan (2001) was satisfied and so the regular variation (if present) has to be of index two or larger.

**Example 3:** Let us look at what happens, when increasing the probability of remaining in the current regime. Take $\bar{p} = 49/50$, $\mathbf{\Phi}^{(1)} = 1/2$ and $\mathbf{\Phi}^{(2)} = 11/10$. For condition $(ii)$ in Theorem 5.42 numerical calculations show that it holds with $\lambda_0 = 0.1846475$. So, Theorem 5.42 gives that the MS-AR(1) process is regularly varying with index $\lambda_0$ provided $(iii)$ is also satisfied. Yet, the simulation in Figure 5.7 seems to indicate that the MS-AR(1) process is indeed heavy-tailed and probably regularly varying. Recall from Section 5.4.1 that the sufficient and in this case necessary second-order stationarity condition from

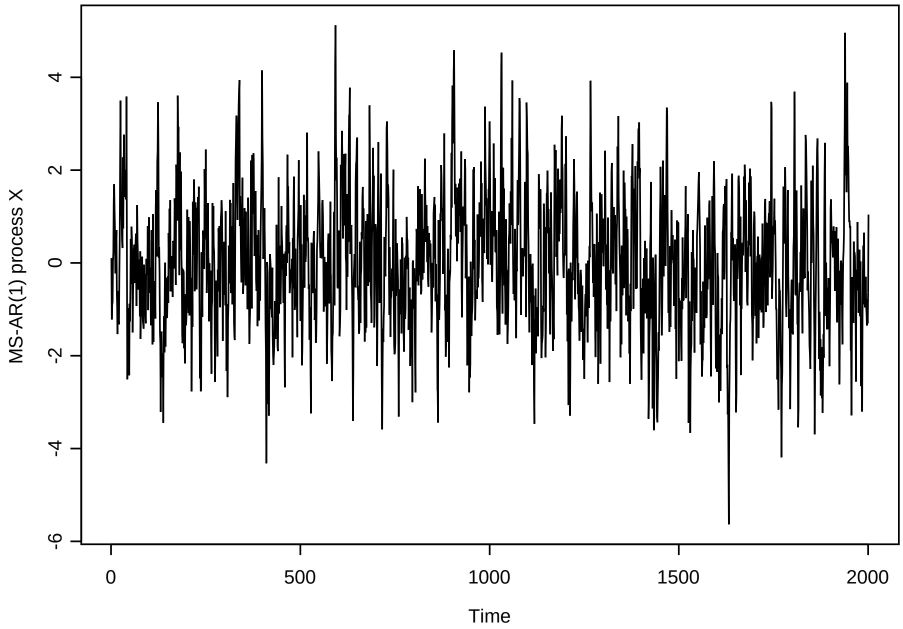


Figure 5.5: Simulation of the MS-AR(1) process in Example 1

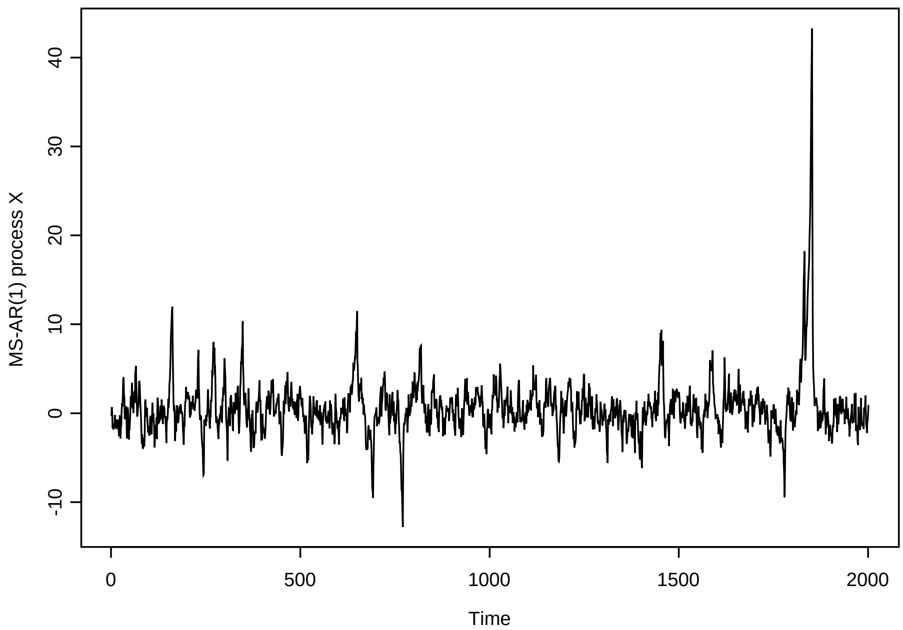


Figure 5.6: Simulation of the MS-AR(1) process in Example 2

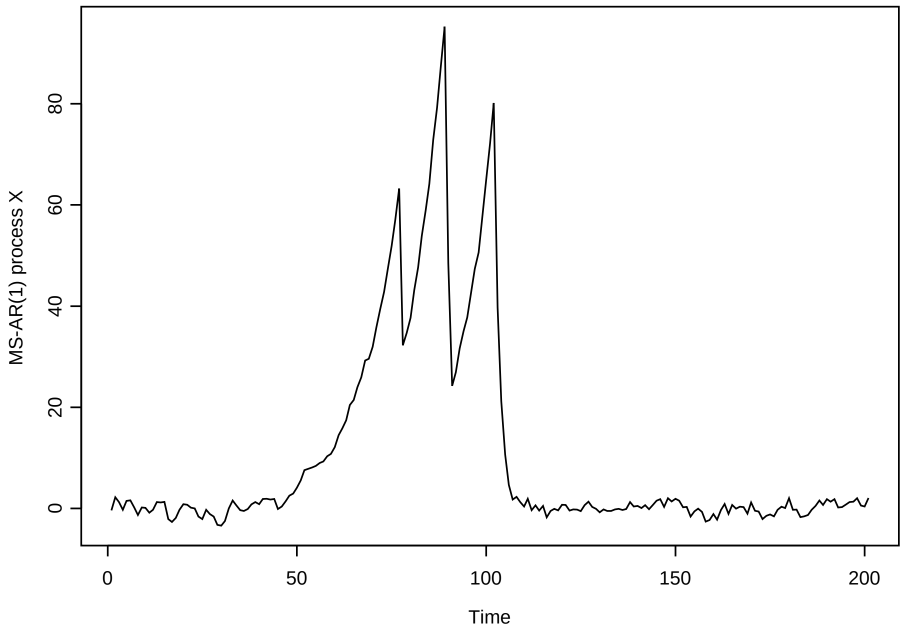


Figure 5.7: Simulation of the MS-AR(1) process in Example 3

Francq and Zakoïan (2001) failed, which fits in with regular variation of order less than two.

**Example 4:** Let us look at what happens, when increasing the explosiveness of the second regime. Take $\bar{p} = 3/4$, $\mathbf{\Phi}^{(1)} = 1/2$ and $\mathbf{\Phi}^{(2)} = 3/2$. For condition (*ii*) in Theorem 5.42 numerical calculations show that it holds with $\lambda_0 = 0.34028$. So, Theorem 5.42 gives that the MS-AR(1) process is regularly varying with index $\lambda_0$ provided (*iii*) is also satisfied. Yet, the simulation in Figure 5.8 seems to indicate that the MS-AR(1) process is indeed heavy-tailed and probably regularly varying. Recall from Section 5.4.1 that the sufficient and in this case necessary second-order stationarity condition from Francq and Zakoïan (2001) failed, which fits in with regular variation of order less than two.

**Example 5:** Let us reconsider Example 4 in the presence of an i.i.d. symmetric 1.5-stable sequence. As this noise has finite moments of orders smaller than 1.5 and $1.5 > \lambda_0 = 0.34028$, Theorem 5.42 gives that the MS-AR(1) process is regularly varying with index $\lambda_0$ provided (*iii*) is also satisfied. Yet, the simulation in Figure 5.9 also seems to indicate that the MS-AR(1) process is indeed heavy-tailed and probably regularly varying. Moreover, it appears to be qualitatively rather similar to Figure 5.8.

Actually, we conjecture that (*iii*) of Theorem 5.42 was satisfied in the above examples and thus the MS-AR(1) processes of Examples 2-5 are indeed regularly varying. A comparison of Examples 2 and 3 then shows that the finiteness of moments does not only depend on the possible states and stationary distribution of the AR parameter, but also on the transition probabilities.

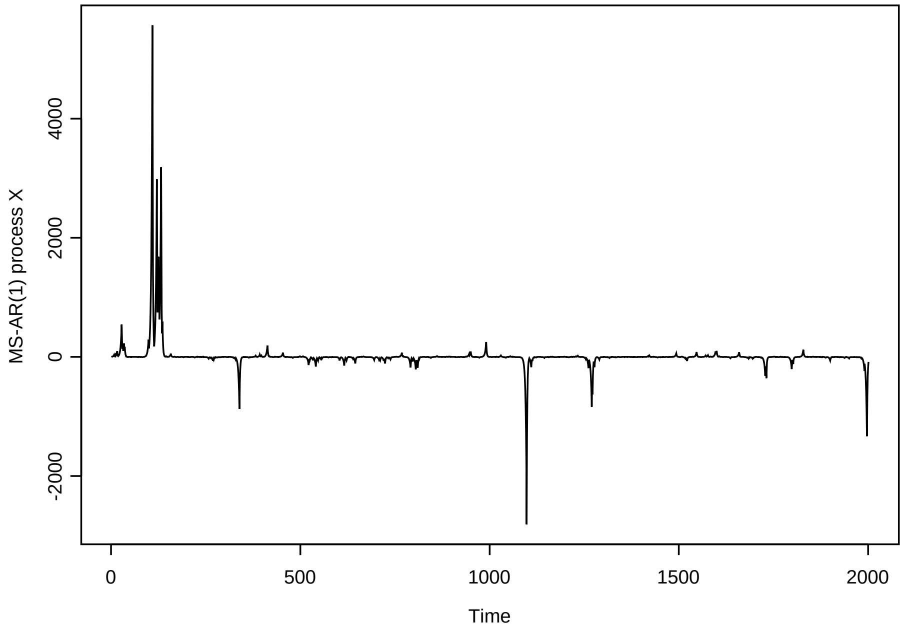


Figure 5.8: Simulation of the MS-AR(1) process in Example 4

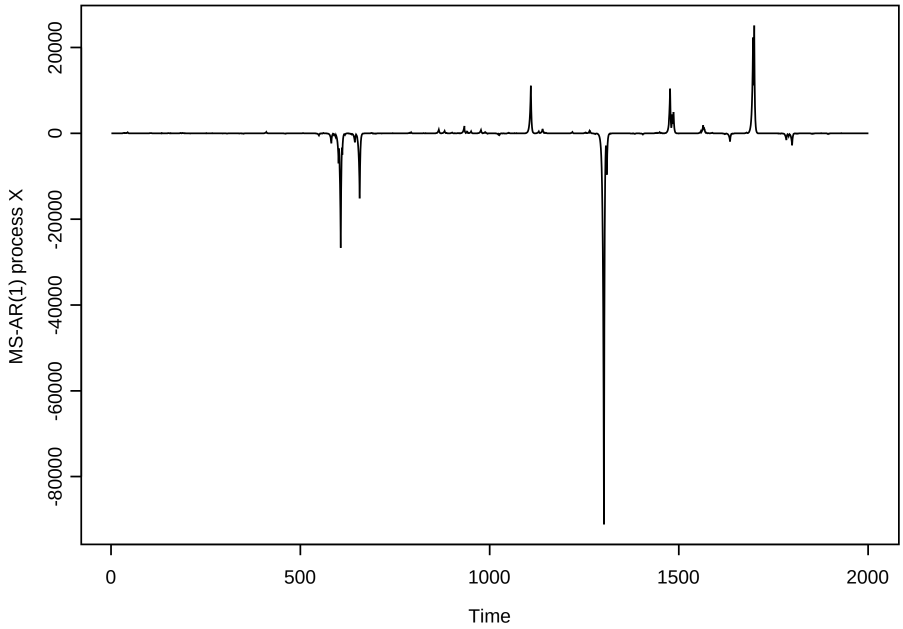


Figure 5.9: Simulation of the MS-AR(1) process in Example 5

# Chapter 6

# Markov-Switching GARCH Models

Let us now turn to the analysis of Markov-switching GARCH models in this chapter. We will briefly state two possible transformations into a stochastic recurrence equation of the type $Y_n = A_n Y_{n-1} + C_n$ and stationarity and moment existence conditions based upon the results of Chapter 4.

## 6.1 Definition

The autoregressive conditional heteroskedasticity (ARCH) model has been introduced in the seminal work of Engle (1982) and extended to generalized autoregressive conditional heteroskedasticity by Bollerslev (1986). Strict stationarity issues have been studied in Bougerol and Picard (1992a) and for results regarding the tail behaviour and sample autocorrelations see Basrak, Davis and Mikosch (2002b). The tail behaviour of an autoregressive process with ARCH(1) errors is analysed in Borkovec and Klüppelberg (2001) and for some existence of moments results see Chen and An (1998) or Carrasco and Chen (2002), who also consider some mixing properties.

Recall that the GARCH(p,q) model with $p, q \in \mathbb{N}_0, p + q > 0$, is defined by a set of positive parameters $\alpha_0, \alpha_1, \ldots, \alpha_p, \beta_1, \beta_2, \ldots, \beta_q$ and an i.i.d. sequence $(Z_t)_{t\in\mathbb{Z}}$ of $\mathbb{R}$-valued random variables. The solution $X_t$ to the system of equations

$$X_t = \sqrt{\sigma_t^2} Z_t \tag{6.1}$$

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{p} \alpha_i X_{t-i}^2 + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^2 \tag{6.2}$$

is then called a *GARCH(p,q)* process. In the case $q = 0$ the term *ARCH(p)* is used. As $\sigma_t^2$ determines the conditional variance of $X_t^2$ given $\sigma_t^2$, provided $Z_t \in L^2$, we refer to $(\sigma_t^2)$ as the *variance process*. In most papers one presumes $E(Z_t) = 0$ and $E(Z_t^2) = 1$. In this set-up it was shown in the original paper by Bollerslev that the GARCH equations have a second-order stationary solution, iff

$$\sum_{i=1}^{p} \alpha_i + \sum_{j=1}^{q} \beta_j < 1.$$

Using a stochastic recurrence equation representation, which we shall consider in detail later when studying the Markov-switching GARCH, Bougerol and Picard (1992a) showed that the GARCH equations (6.1), (6.2) have a strictly stationary solution, if and, provided $\alpha_0 > 0$, only if the associated top Lyapunov exponent is strictly negative, and that this is satisfied in the case analysed in Bollerslev (1986). Observe that for $\alpha_0 = 0$ one has that $X_t = 0, \sigma_t^2 \, \forall \, t \in \mathbb{Z}$ is a trivial solution.

By letting the parameters change over time as a Markov chain we now define Markov-switching GARCH processes. Again we allow the parameter chain to have arbitrary state space, whereas previous papers restricted the state space to be finite.

**Definition 6.1 (MS-GARCH($p, q$) process)** *Let $p, q \in \mathbb{N}_0$, $p + q \geq 1$ and*

$$\Delta = (\alpha_{0t}, \alpha_{1t}, \dots, \alpha_{pt}, \beta_{1t}, \beta_{2t}, \dots, \beta_{qt}, \tau_t)_{t \in \mathbb{Z}} \tag{6.3}$$

*be a stationary and ergodic Markov chain with some (measurable) subset $E$ of the cone $(\mathbb{R}^+)^{p+q+2}$ as state space. Moreover, let $\epsilon = (\epsilon_t)_{t \in \mathbb{Z}}$ be an i.i.d. sequence of $\mathbb{R}^d$-valued random variables independent of $\Delta$. A stationary process $(X_t)_{t \in \mathbb{Z}}$ in $\mathbb{R}^d$ is called a Markov-switching GARCH,* MS-GARCH$(p, q, \Delta, \epsilon)$*, process, if*

$$\begin{aligned} X_t &= \sqrt{\sigma_t^2} \tau_t \epsilon_t & (6.4) \\ \sigma_t^2 &= \alpha_{0t} + \sum_{i=1}^{p} \alpha_{it} X_{t-i}^2 + \sum_{j=1}^{q} \beta_{jt} \sigma_{t-j}^2 & (6.5) \end{aligned}$$

*holds for all $t \in \mathbb{Z}$.*

*Furthermore, a stationary process $(X_t)_{t \in \mathbb{Z}}$ is said to be an MS-GARCH($p, q$) process, if it is an* MS-GARCH$(p, q, \Delta, \epsilon)$ *process for some $\Delta$ and $\epsilon$ satisfying the above conditions.*

The above definition is basically the same as for the MS-GARCH models driven by a finite state-space chain analysed in Francq, Roussignol and Zakoïan (2001) and Francq and Zakoïan (2004). In our opinion this appears to be the most natural formulation of MS-GARCH processes. We included also $\tau_t$ into the driving Markov chain and (6.4) to allow specifications somewhat similar to the MS-ARCH process of Hamilton and Susmel (1994). Some other extensions of GARCH to Markov-switching models like the one of Haas, Mittnik and Paolella (2004) do, however, not fit into our above framework.

## 6.2 Stationarity of Markov-Switching GARCH processes

Having defined MS-GARCH processes above, we now turn to higher dimensional representations leading to easy-to-handle first order stochastic difference equations in order to analyse stationarity properties of MS-GARCH processes.

Results regarding (strict) stationarity of MS-GARCH processes are to the best of our knowledge only contained in Francq, Roussignol and Zakoïan (2001) and Francq and Zakoïan (2004) as far as the existing literature is regarded. However, both papers only consider finitely many states of the Markov chain and focus on the $L^2$ structure. The notion of the top Lyapunov coefficient is only used in Francq, Roussignol and Zakoïan (2001) and restricted to establishing that $(\sigma_t^2, \sigma_{t-1}^2, \ldots, \sigma_{t-\max(p,q)+1}^2)_{t\in\mathbb{Z}}^\mathsf{T}$ is a stationary random sequence.

The following multidimensional representation is the analogue of the one employed e.g. in Bougerol and Picard (1992b) or Basrak, Davis and Mikosch (2002b) for the standard GARCH.

Define

$$
\begin{aligned}
\mathbf{X}_t &= \left(\sigma_{t+1}^2, \sigma_t^2, \ldots, \sigma_{t-q+2}^2, X_t^2, X_{t-1}^2, \ldots, X_{t-p+2}^2\right)^\mathsf{T} \in (\mathbb{R}^+)^{p+q-1} \\
\mathbf{C}_t &= (\alpha_{0,t+1}, 0, 0, \ldots, 0)^\mathsf{T} \in (\mathbb{R}^+)^{p+q-1} \\
\mathbf{A}_t &= \begin{pmatrix}
\alpha_{1,t+1}\tau_t^2\epsilon_t^2 + \beta_{1,t+1} & \beta_{2,t+1} & \cdots & \beta_{q-1,t+1} & \beta_{q,t+1} & \alpha_{2,t+1} & \alpha_{3,t+1} & \cdots & \alpha_{p,t+1} \\
1 & 0 & \cdots & 0 & 0 & 0 & 0 & \cdots & 0 \\
0 & 1 & \ddots & 0 & 0 & 0 & 0 & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\
0 & \cdots & 0 & 1 & 0 & 0 & \cdots & \cdots & 0 \\
\tau_t^2\epsilon_t^2 & 0 & \cdots & 0 & 0 & 0 & 0 & \cdots & 0 \\
0 & 0 & \cdots & 0 & 0 & 1 & 0 & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \ddots & \vdots \\
0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 1 & 0
\end{pmatrix}, \\
\mathbf{A}_t &\in M_{p+q-1}(\mathbb{R}^+).
\end{aligned}
$$

To avoid any degeneracies we presume w.l.o.g. that $p, q \geq 2$. This presumption shall always be valid, whenever we use the above representation later, and can be ensured by simply including higher order terms with GARCH coefficients equal to zero.

Using the same arguments as for MS-ARMA equations, one then immediately obtains that

$$\mathbf{X}_t = \mathbf{A}_t\mathbf{X}_{t-1} + \mathbf{C}_t \tag{6.6}$$

has a stationary and ergodic solution, iff the squared system of the MS-GARCH equations (6.4), (6.5), viz.

$$
\begin{aligned}
X_t^2 &= \sigma_t^2\tau_t^2\epsilon_t^2 &(6.7)\\
\sigma_t^2 &= \alpha_{0t} + \sum_{i=1}^{p}\alpha_{it}X_{t-i}^2 + \sum_{j=1}^{q}\beta_{jt}\sigma_{t-j}^2, &(6.8)
\end{aligned}
$$

has one. Moreover, the solutions can be transformed into one another by the above formulae.

In the following we shall use the above representation. Another possibility is to define

$$
\begin{aligned}
\widetilde{\mathbf{X}}_t &= \left(X_t^2, X_{t-1}^2, \ldots, X_{t-p+1}^2, \sigma_t^2, \sigma_{t-1}^2, \ldots, \sigma_{t-q+1}^2\right)^\mathsf{T} \in (\mathbb{R}^+)^{p+q} \\
\widetilde{\mathbf{C}}_t &= (\alpha_{0t}\tau_t^2\epsilon_t^2, \underbrace{0, \ldots, 0}_{p-1}, \alpha_{0t}, \underbrace{0, \ldots, 0}_{q-1})^\mathsf{T} \in (\mathbb{R}^+)^{p+q}
\end{aligned}
$$

$$\begin{aligned}\widetilde{\mathbf{A}}_t &= \begin{pmatrix} \alpha_{1t}\tau_t^2\epsilon_t^2 & \cdots & \alpha_{p-1,t}\tau_t^2\epsilon_t^2 & \alpha_{pt}\tau_t^2\epsilon_t^2 & \beta_{1t}\tau_t^2\epsilon_t^2 & \cdots & \beta_{q-1,t}\tau_t^2\epsilon_t^2 & \beta_{qt}\tau_t^2\epsilon_t^2 \\ 1 & \cdots & 0 & 0 & 0 & \cdots & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & \cdots & 1 & 0 & 0 & \cdots & 0 & 0 \\ \alpha_{1t} & \cdots & \alpha_{p-1,t} & \alpha_{pt} & \beta_{1t} & \cdots & \beta_{q-1,t} & \beta_{qt} \\ 0 & \cdots & 0 & 0 & 1 & \cdots & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & \cdots & 0 & 0 & 0 & \cdots & 1 & 0 \end{pmatrix} \\ \widetilde{\mathbf{A}}_t &\in \in M_{p+q}(\mathbb{R}^+)\end{aligned}$$

as in Basrak (2000) or Francq and Zakoïan (2004) and to consider the stochastic difference equation

$$\widetilde{\mathbf{X}}_t = \widetilde{\mathbf{A}}_t\widetilde{\mathbf{X}}_{t-1} + \widetilde{\mathbf{C}}_t, \tag{6.9}$$

which again is equivalent to (6.7), (6.8). The upcoming theorems can also straightforwardly be formulated using this set-up. Note that the first representation is only $p+q-1$-dimensional, whereas the second is in $\mathbb{R}^{p+q}$. Moreover, the "noise" $\mathbf{C}_t$ does not depend on $\epsilon$ in the first one. On the other hand the second representation has the advantage that no time $t+1$ variables are involved.

As in the MS-ARMA case Proposition 2.23 and Theorem 2.24 imply that the joint random sequence $(\Delta, \epsilon) = (\Delta_t, \epsilon_t)_{t\in\mathbb{Z}}$ is stationary and ergodic and thus an obvious application of Lemma 2.25 shows that the transformed sequence $(\mathbf{A}, \mathbf{C}) = (\mathbf{A}_t, \mathbf{C}_t)_{t\in\mathbb{Z}}$ is stationary and ergodic. Hence, we obtain the following result from Theorem 4.1 stating sufficient conditions for (6.7) and (6.8) to have a solution.

**Theorem 6.2** *The (squared) MS-GARCH($p, q, \Delta, \epsilon$) equations (6.7) and (6.8) have a unique stationary and ergodic solution, if $E(\log^+\|\mathbf{A}_0\|)$, $E(\log^+\|\mathbf{C}_0\|)$ are finite and the Lyapunov exponent satisfies*

$$\gamma = \inf_{t\in\mathbb{N}_0} \frac{1}{t+1}\left(E\left(\log\|\mathbf{A}_0\mathbf{A}_{-1}\cdots\mathbf{A}_{-t}\|\right)\right) < 0.$$

*The unique stationary solution $(X^2, \sigma^2) = (X_t^2, \sigma_t^2)_{t\in\mathbb{Z}}$ is formed by the (q+1)th and the second coordinate of*

$$\mathbf{X}_t = \sum_{k=0}^{\infty} \mathbf{A}_t\mathbf{A}_{t-1}\cdots\mathbf{A}_{t-k+1}\mathbf{C}_{t-k}, \tag{6.10}$$

*which is the unique stationary and ergodic solution of (6.6). The series defining $\mathbf{X}$ converges absolutely a.s.*

*Let $\mathbf{V}_0$ be an arbitrary $(\mathbb{R}^+)^{p+q-1}$-valued random variable defined on the same probability space as $(\Delta_t, \epsilon_t)_{t\in\mathbb{Z}}$ and define $(\mathbf{V}_t)_{t\in\mathbb{N}}$ recursively via (6.6). Then*

$$\|\mathbf{X}_t - \mathbf{V}_t\| \overset{\text{a.s.}}{\to} 0 \text{ as } t\to\infty \tag{6.11}$$

*and, in particular,*

$$\mathbf{V}_t \overset{\mathscr{D}}{\to} \mathbf{X}_0 \text{ as } t\to\infty, \tag{6.12}$$

*i.e. the distribution of $\mathbf{V}_t$ converges to the stationary distribution of $\mathbf{X}_t$.*

The above theorem constructs a unique stationary and ergodic solution $(X_t^2, \sigma_t^2)$ to equations (6.7) and (6.8), whereas we are actually interested in a stationary and ergodic solution $(X_t, \sigma_t^2)$ to the system of equations (6.4) and (6.5). It is obvious that a solution of (6.4) leads to a solution of (6.7) by simply taking squares. Thus instead of searching for a solution for (6.4) and (6.5), we can equivalently consider a solution of the three equations

$$
\begin{aligned}
X_t &= \sqrt{\sigma_t^2}\tau_t\epsilon_t & (6.13)\\
X_t^2 &= \sigma_t^2\tau_t^2\epsilon_t^2 & (6.14)\\
\sigma_t^2 &= \alpha_{0t} + \sum_{i=1}^{p}\alpha_{it}X_{t-i}^2 + \sum_{j=1}^{q}\beta_{jt}\sigma_{t-j}^2. & (6.15)
\end{aligned}
$$

Assume the assumptions of the above theorem are satisfied. From the representation of the unique stationary and ergodic solution $\mathbf{X}_t = \left(\sigma_{t+1}^2, \sigma_t^2, \ldots, \sigma_{t-q+2}^2, X_t^2, X_{t-1}^2, \ldots, X_{t-p+2}^2\right)^\mathsf{T}$ to the equations (6.14) and (6.15) given in (6.10) and Lemma 2.25 we can conclude that $(\mathbf{X}_t, \tau_t, \epsilon_t) = \left(\left(\sigma_{t+1}^2, \sigma_t^2, \ldots, \sigma_{t-q+2}^2, X_t^2, X_{t-1}^2, \ldots, X_{t-p+2}^2\right)^\mathsf{T}, \tau_t, \epsilon_t\right)$ is a stationary and ergodic sequence, as $(\Delta_t, \epsilon_t)$ is stationary and ergodic. But setting $X_t = \sqrt{\sigma_t^2}\tau_t\epsilon_t$ also solves (6.13) and another application of Lemma 2.25 then gives that $(X_t, \sigma_t^2)$ is a stationary and ergodic sequence. That this is a unique solution to the original MS-GARCH equations (6.4) and (6.5) is clear in view of the uniqueness of $(X_t^2, \sigma_t^2)$ ensured by the last theorem. Let us summarize these conclusions in the following theorem:

**Theorem 6.3** *The MS-GARCH($p, q, \Delta, \epsilon$) equations (6.4) and (6.5) have a unique stationary and ergodic solution, if $E(\log^+\|\mathbf{A}_0\|)$, $E(\log^+\|\mathbf{C}_0\|)$ are finite and the Lyapunov exponent satisfies*

$$\gamma = \inf_{t\in\mathbb{N}_0}\left(\frac{1}{t+1}E\left(\log\|\mathbf{A}_0\mathbf{A}_{-1}\cdots\mathbf{A}_{-t}\|\right)\right) < 0.$$

*The unique stationary and ergodic solution $(X, \sigma^2) = (X_t, \sigma_t^2)_{t\in\mathbb{Z}}$ is formed by $X_t = \sqrt{\sigma_t^2}\tau_t\epsilon_t$ and the second coordinate $\sigma_t^2$ of*

$$\mathbf{X}_t = \sum_{k=0}^{\infty}\mathbf{A}_t\mathbf{A}_{t-1}\cdots\mathbf{A}_{t-k+1}\mathbf{C}_{t-k}, \tag{6.16}$$

*which is the unique stationary and ergodic solution of (6.6). The series defining $\mathbf{X}$ converges absolutely a.s.*

It is even possible to show that the strict negativity of the top Lyapunov coefficient $\gamma$ is necessary for the existence of a stationary solution to (6.6) under some technical conditions. The crucial difference to the MS-ARMA case is that all involved matrices and vectors are non-negative. The proof of the following result is a straightforward adaptation of the proof given in Bougerol and Picard (1992a) to the Markov-switching case. A similar result on a stochastic difference equation for $(\sigma_t^2, \sigma_{t-1}^2, \ldots, \sigma_{t-\max\{p,q\}+1}^2)^\mathsf{T}$ can be found in Francq, Roussignol and Zakoïan (2001).

**Theorem 6.4** *Assume that* $E(\log^+ \|\mathbf{A}_0\|) < \infty$ *and that there are constants* $a_0, a_2, a_3, \ldots, a_p, b_1, b_2, \ldots, b_q > 0$ *such that* $\alpha_{0t} \geq a_0, \alpha_{2t} \leq a_2, \alpha_{3t} \leq a_3, \ldots, \alpha_{pt} \leq a_p, \beta_{1t} \leq b_1, \beta_{2t} \leq b_2, \ldots, \beta_q \leq b_q$ *a.s. for* $t \in \mathbb{Z}$. *Then a necessary condition for equation (6.6) to have a stationary solution is strict negativity of the Lyapunov coefficient*

$$\gamma = \inf_{t \in \mathbb{N}_0} \left( \frac{1}{t+1} E\left(\log \|\mathbf{A}_0 \mathbf{A}_{-1} \cdots \mathbf{A}_{-t}\|\right) \right).$$

Observe that no restriction is imposed upon $\alpha_{1t}$ and that as usual $\|\cdot\|$ may be any norm on $\mathbb{R}^{p+q-1}$.

**Proof:** Assume that $(\mathbf{Y}_t)_{t \in \mathbb{Z}}$ is a stationary solution of (6.6) and recall that all coefficients of $\mathbf{A}_t, \mathbf{C}_t, \mathbf{Y}_t$ are non-negative. We have for any $n \in \mathbb{N}$

$$\begin{aligned} \mathbf{Y}_0 &= \mathbf{A}_0 \mathbf{Y}_{-1} + \mathbf{C}_0 \\ &= \mathbf{A}_0 \mathbf{A}_{-1} \mathbf{Y}_{-2} + \mathbf{C}_0 + \mathbf{A}_0 \mathbf{C}_{-1} \\ &= \mathbf{A}_0 \mathbf{A}_{-1} \cdots \mathbf{A}_{-n} \mathbf{Y}_{-n-1} + \mathbf{C}_0 + \sum_{k=0}^{n-1} \mathbf{A}_0 \mathbf{A}_{-1} \cdots \mathbf{A}_{-k} \mathbf{C}_{-k-1} \end{aligned}$$

and so the non-negativity gives

$$\mathbf{Y}_0 \geq \sum_{k=0}^{n-1} \mathbf{A}_0 \mathbf{A}_{-1} \cdots \mathbf{A}_{-k} \mathbf{C}_{-k-1}$$

for all $n \in \mathbb{N}$, where $\geq$ is to be understood componentwise (as in the remainder of this proof). Hence, the series

$$\sum_{k=0}^{n-1} \mathbf{A}_0 \mathbf{A}_{-1} \cdots \mathbf{A}_{-k} \mathbf{C}_{-k-1}$$

converges a.s. and thus

$$\lim_{n \to \infty} \mathbf{A}_0 \mathbf{A}_{-1} \cdots \mathbf{A}_{-n} \mathbf{C}_{-n-1} = 0 \text{ a.s.} \tag{6.17}$$

Let now $\{e_i\}_{i=1,\ldots,p+q-1}$ by the canonical basis vectors of $\mathbb{R}^{p+q-1}$, i.e. $e_1 = (1, 0, \ldots, 0)^\mathsf{T}$, $e_2 = (0, 1, 0, \ldots, 0)^\mathsf{T}$, etc. We shall now subsequently show that

$$\lim_{n \to \infty} \mathbf{A}_0 \mathbf{A}_{-1} \cdots \mathbf{A}_{-n} e_i = 0 \tag{6.18}$$

for all $i \in \{1, 2, \ldots, p+q-1\}$. This implies that

$$\|\mathbf{A}_0 \mathbf{A}_{-1} \cdots \mathbf{A}_{-n}\| \to 0 \text{ as } n \to \infty \tag{6.19}$$

and therefore an application of Lemma 4.4 concludes the proof.

So let us turn to establishing (6.18). As $\mathbf{C}_{-k-1} = \alpha_{0,-k} e_1 \geq a_0 e_1$, (6.17) gives

$$\lim_{n \to \infty} \mathbf{A}_0 \mathbf{A}_{-1} \cdots \mathbf{A}_{-n} e_1 = 0 \text{ a.s.} \tag{6.20}$$

Moreover, $\mathbf{A}_{-n}e_q = \beta_{q,-n+1}e_1 \leq b_q e_1$ and thus

$$\lim_{n\to\infty} \mathbf{A}_0\mathbf{A}_{-1}\cdots\mathbf{A}_{-n}e_q = 0 \text{ a.s.},$$

as $\mathbf{A}_{-n}e_q = \beta_{q,-n+1}e_1 \leq b_q e_1$ implies $\mathbf{A}_0\mathbf{A}_{-1}\cdots\mathbf{A}_{-n}e_q \leq \mathbf{A}_0\mathbf{A}_{-1}\cdots\mathbf{A}_{-n+1}b_q e_1 \to 0$ and $\mathbf{A}_0\mathbf{A}_{-1}\cdots\mathbf{A}_{-n}e_q \geq 0$ due to the non-negativity of all involved vectors and matrices. Using $\mathbf{A}_{-n}e_{j-1} = \beta_{j-1,-n+1}e_1 + e_j \leq b_{j-1}e_1 + e_j$ for all $2 < j \leq q$, backwards induction starting with $j = q$ and arguments analogous to the above give

$$\lim_{n\to\infty} \mathbf{A}_0\mathbf{A}_{-1}\cdots\mathbf{A}_{-n}e_i = 0 \text{ a.s.}$$

for all $2 \leq i < q$.

The very same line of argumentation based on the relations

$$\mathbf{A}_{-n}e_{p+q-1} = \alpha_{p,-n+1}e_1 \leq a_p e_1$$

and

$$\mathbf{A}_{-n}e_{q+j-1} = \alpha_{j,-n+1}e_1 + e_{q+j} \leq a_j e_1 + e_{q+j}$$

for all $2 \leq j < p$ shows

$$\lim_{n\to\infty} \mathbf{A}_0\mathbf{A}_{-1}\cdots\mathbf{A}_{-n}e_i = 0 \text{ a.s.}$$

for all $q+1 \leq i \leq p+q-1$. Thus, (6.18) is established.

To conclude the proof let us give precise arguments for (6.19). Obviously it suffices to show this for any particular norm on $\mathbb{R}^{p+q-1}$. We take w.l.o.g. $\|\cdot\|_\infty$. Now we have

$$\begin{aligned}
\lim_{n\to\infty} \|\mathbf{A}_0\mathbf{A}_{-1}\cdots\mathbf{A}_{-n}\|_\infty &= \lim_{n\to\infty} \sup_{\substack{\|x\|_\infty = 1, \\ x \in \mathbb{R}^{p+q-1}}} \left(\|\mathbf{A}_0\mathbf{A}_{-1}\cdots\mathbf{A}_{-n}x\|_\infty\right) \\
&\leq \lim_{n\to\infty} \sum_{i=1}^{p+q-1} \|\mathbf{A}_0\mathbf{A}_{-1}\cdots\mathbf{A}_{-n}e_i\|_\infty = 0 \text{ a.s.}
\end{aligned}$$

□

In order to check the strict negativity of the Lyapunov coefficient for a given model the simplest approach appears to be to try to get $E\left(\log\|\mathbf{A}_0\|\right) < 0$ for some algebra norm $\|\cdot\|$. Using the alternative stochastic recurrence equation representation (6.9) Francq and Zakoïan (2004) have again given a spectral radius condition ensuring strict stationarity and finite second moments in the finite state space case.

## 6.3 Existence of Moments

To study the existence of moments of MS-GARCH models the results of Section 4.2 can be applied to the stochastic recurrence equations of the previous section. For the reader's convenience we summarize the results below, but do not give any detailed proofs as they can either be found in Section 4.2 or are analogous to the ones in Section 5.2.3. Again we use $\tilde{p}$ to denote orders of moments, as $p$ is already employed to denote the ARCH order.

It is immediate that Lemmata 4.5 and 4.6 can be applied to the sequence $(\mathbf{A}_t)$.

The general Theorem 4.7 becomes for MS-GARCH processes:

**Theorem 6.5** *Assume the conditions of Theorem 6.2 are fulfilled. If, moreover, for some $\tilde{p} \in (1, \infty]$*

$$\sum_{k=0}^{\infty} \|\mathbf{A}_0 \cdots \mathbf{A}_{-k+1} \mathbf{C}_{-k}\|_{L^{\tilde{p}}} \tag{6.21}$$

*or for some $\tilde{p} \in (0, 1)$*

$$\sum_{k=0}^{\infty} E\left(\|\mathbf{A}_0 \cdots \mathbf{A}_{-k+1} \mathbf{C}_{-k}\|^{\tilde{p}}\right) \tag{6.22}$$

*converges, then the solution $(X_t^2, \sigma_t^2)$ of the (squared) MS-GARCH equations (6.7), (6.8) and its higher dimensional representation $\mathbf{X}_t$ are in $L^{\tilde{p}}$. Moreover, the series defining $\mathbf{X}_t$ (as given in Theorem 6.2) converges in $L^{\tilde{p}}$.*

Observe that $X_t^2 \in L^{\tilde{p}}$, of course, implies $X_t \in L^{2\tilde{p}}$ for the solution of the MS-GARCH equations (6.4), (6.5).

It is clear that, if (6.21) holds for some $\tilde{p} \geq 1$, it holds for all $r \in [1, \tilde{p}]$ as well. However, if (6.22) holds for some $\tilde{p} \in (0, 1)$ this does not imply that it holds for all smaller values of $\tilde{p}$ as well. Yet, of course, if $\mathbf{X} \in L^{\tilde{p}}$ for some $\tilde{p} > 0$ then $\mathbf{X} \in L^r$ for all $r \in (0, \tilde{p}]$. Note, however, that the asymptotic conditions given in the next lemmata are much better behaved. If there is one $\tilde{p} \in (0, \infty]$ that fulfils the asymptotic condition, then the asymptotic conditions for all $s \in (0, \tilde{p}]$ are satisfied as well (use Jensen's inequality as in the proof of Lemma 3.20).

**Proof:** Combine Theorems 6.2 and 4.7 to obtain the results on $\mathbf{X}_t$. $(X_t^2, \sigma_t^2) \in L^{\tilde{p}}$ is now a consequence of $\mathbf{X}_t \in L^{\tilde{p}}$ and Corollary 2.15. □

We restate also Lemma 4.8 and Propositions 4.9, 4.10 for the special case of MS-GARCH processes. For all the following results one should keep in mind that $\mathbf{A}_t$ is formed by components of both $\Delta$ and $\epsilon$, whereas $\mathbf{C}_t$ is solely determined by $\Delta$. One important effect of this is that $\mathbf{A}_t$ can, apart from degenerate cases, only be in $L^\infty$, if $\epsilon_0 \in L^\infty$.

**Lemma 6.6** *Let $1 \leq \tilde{p} \leq \infty$, resp. $0 < \tilde{p} < 1$, and assume that*

$$\limsup_{k \to \infty} \|\mathbf{A}_0 \cdots \mathbf{A}_{-k+1} \mathbf{C}_{-k}\|_{L^{\tilde{p}}}^{1/k} < 1,$$

*resp.*

$$\limsup_{k \to \infty} \left(E(\|\mathbf{A}_0 \cdots \mathbf{A}_{-k+1} \mathbf{C}_{-k}\|^{\tilde{p}})\right)^{1/k} < 1,$$

*holds. Then (6.21), resp. (6.22), is fulfilled.*

**Proposition 6.7** *Let $\tilde{p} \in (0, \infty)$. If there exist $r, s \geq 1$ with $1/r + 1/s = 1$, such that $\mathbf{A}_0 \cdots \mathbf{A}_{-k+1} \in L^{\tilde{p}r}$ $\forall k \in \mathbb{N}$, $\limsup_{k \to \infty} E(\|\mathbf{A}_0 \cdots \mathbf{A}_{-k+1}\|^{\tilde{p}r})^{1/k} < 1$ for $0 < \tilde{p}r < \infty$, resp. $\lim_{k \to \infty} \|\mathbf{A}_0 \cdots \mathbf{A}_{-k+1}\|_{L^\infty}^{1/k} < 1$ for $\tilde{p}r = \infty$, and $\mathbf{C}_0 \in L^{\tilde{p}s}$, then $\gamma < 0$ and (6.21) for $\tilde{p} \geq 1$, resp. (6.22) for $0 < \tilde{p} < 1$, hold.*

Again one especially obtains that, provided $\mathbf{A}_0 \in L^\infty$ (and thus $\mathbf{A}_0 \cdots \mathbf{A}_{-k+1} \in L^\infty$) and $\lim_{k \to \infty} \|\mathbf{A}_0 \cdots \mathbf{A}_{-k+1}\|_{L^\infty}^{1/k} < 1$, the squared MS-GARCH process $(X_t^2, \sigma_t^2)$ and its higher dimensional representation $\mathbf{X}_t$ are in $L^{\tilde{p}}$, if $\mathbf{C}_0 \in L^{\tilde{p}}$. The latter is equivalent to $\alpha_{0t} \in L^{\tilde{p}}$.

**Proposition 6.8** *If* $\mathbf{A}_0 \in L^\infty \forall k \in \mathbb{N}$, $\lim_{k\to\infty} \|\mathbf{A}_0 \cdots \mathbf{A}_{-k+1}\|_{L^\infty}^{1/k} < 1$ *and* $\mathbf{C}_0 \in L^\infty$, *then* $\gamma < 0$ *and (6.21) with* $p = \infty$ *hold.*

If we have that $\mathbf{A} = (\mathbf{A}_t)$ is a sequence of independent random variables (which is the case, if $(\tau_t)$ and $(\alpha_{1t}, \ldots, \alpha_{pt}, \beta_{1t}, \ldots, \beta_{qt})$ are i.i.d. sequences independent of one another), $E(\|\mathbf{A}_0\|^{\tilde{p}}) < 1$ ensures $\lim_{k\to\infty} E(\|\mathbf{A}_0 \cdots \mathbf{A}_{-k+1}\|^{\tilde{p}})^{1/k} < 1$, since $\|\cdot\|$ is submultiplicative.

The most straightforward moment conditions are obtainable under the assumption that $\mathbf{A} = (\mathbf{A}_t)$ and $\mathbf{C} = (\mathbf{C}_t)$ are independent. This happens, if $\alpha_{0t}$ is a constant or at least independent from the other components of the Markov chain $\Delta$. In this case one obtains the following simplification of Proposition 6.7.

**Proposition 6.9** *Let* $\mathbf{A}_0\mathbf{A}_{-1} \cdots \mathbf{A}_{-k+1}$ *be independent of* $\mathbf{C}_{-k}$ *for all* $k \in \mathbb{N}$ *and* $\tilde{p} \in (0, \infty)$. *If* $\mathbf{A}_0 \cdots \mathbf{A}_{-k+1} \in L^{\tilde{p}} \,\forall\, k$, $\mathbf{C}_0 \in L^{\tilde{p}}$ *and* $\limsup_{k\to\infty} E(\|\mathbf{A}_0 \cdots \mathbf{A}_{-k+1}\|^{\tilde{p}})^{1/k} < 1$, *then (6.21) for* $\tilde{p} \geq 1$, *resp. (6.22) for* $0 < \tilde{p} < 1$, *holds.*

The prerequisite independence is in particular satisfied, if $\mathbf{A}$ and $\mathbf{C}$ are independent or $(\mathbf{A}_k, \mathbf{C}_k)_{k\in\mathbb{Z}}$ is an i.i.d. sequence.
**Proof:** Proceed along the lines of the proof of Proposition 4.9, but instead of the Hölder inequality use the independence, which gives $E(\|\mathbf{A}_0 \cdots \mathbf{A}_{-k+1}\mathbf{C}_{-k}\|^{\tilde{p}}) \leq E(\|\mathbf{A}_0 \cdots \mathbf{A}_{-k+1}\|^{\tilde{p}})$ $E(\|\mathbf{C}_0\|^{\tilde{p}})$. $\square$

## 6.4 A Note on the Tail Behaviour

Regarding the tail behaviour there is one fact that causes considerable problems, namely that $\mathbf{A}_t$ as well as $\widetilde{\mathbf{A}}_t$ are built form both the driving chain $\Delta$ and the noise sequence $\epsilon$. Moreover, $\mathbf{C}_t$ is formed solely by the driving Markov chain and $\Delta$ is also a main ingredient in $\widetilde{\mathbf{C}}_t$. So there is no “regularly varying” noise case to be considered, at least none that appears natural. (To assume $\alpha_{0t}$ is regularly varying and independent of the other components of $\Delta_t$ would be possible, but this seems to be a rather artificial case.) If $\Delta$ actually is an i.i.d. sequence one can use Theorem 4.15, to study the tails. Such an analysis should be rather similar to the one of the standard GARCH in Basrak, Davis and Mikosch (2002b), which is also based on the results of Kesten. As we focus on “truly” Markovian parameters, we refrain from discussing this in its details.

# Bibliography


An, H. Z. and Huang, F. C.: 1996, The geometrical ergodicity of nonlinear autoregressive models, *Statistica Sinica* **6**, 943–956.

Ash, R. B. and Gardner, M. F.: 1975, *Topics in Stochastic Processes*, Vol. 27 of *Probability and Mathematical Statistics*, Academic Press, New York.

Asmussen, S.: 1999, Stochastic simulation with a view towards stochastic processes, *MaPhySto Lecture Notes 2*, MaPhySto, University of Århus, Denmark.

Asmussen, S.: 2003, *Applied Probability and Queues*, Vol. 51 of *Stochastic Modelling and Applied Probability*, 2nd edn, Springer, New York.

Basrak, B.: 2000, *The Sample Autocorrelation Function of Non-Linear Time Series*, PhD-Thesis, Rijksuniversiteit Groningen, The Netherlands.

Basrak, B., Davis, R. A. and Mikosch, T.: 2002a, A characterization of multivariate regular variation, *Annals of Applied Probability* **12**, 908–920.

Basrak, B., Davis, R. A. and Mikosch, T.: 2002b, Regular variation of GARCH processes, *Stochastic Processes and their Applications* **99**, 95–115.

Bauer, H.: 1992, *Maß- und Integrationstheorie*, 2nd edn, Walter de Gruyter, Berlin.

Bauer, H.: 2001, *Wahrscheinlichkeitstheorie*, 5th edn, Walter de Gruyter, Berlin.

Bingham, G. H., Goldie, C. M. and Teugels, J. L.: 1989, *Regular Variation*, Vol. 27 of *Encyclopedia of Mathematics and its Applications*, 1st paperback edn, Cambridge University Press, Cambridge.

Bollerslev, T.: 1986, Generalized autoregeressive conditional heteroskedasticity, *Journal of Econometrics* **31**, 307–327.

Borkovec, M. and Klüppelberg, C.: 2001, The tail of the stationary distribution of an autoregressive process with ARCH(1) errors, *Annals of Applied Probability* **11**, 1220–1241.

Bougerol, P. and Picard, N.: 1992a, Stationarity of GARCH processes and of some non-negative time series, *Journal of Econometrics* **52**, 115–127.

Bougerol, P. and Picard, N.: 1992b, Strict stationarity of generalized autoregressive processes, *Annals of Probability* **20**, 1714–1730.

Brachner, C.: 2004, *Tailverhalten autoregressiver Thresholdmodelle*, Diplomarbeit, Centre of Mathematical Sciences, Munich University of Technology, Garching, Germany.

Brandt, A.: 1986, The stochastic equation $Y_{n+1} = A_n Y_n + B_n$ with stationary coefficients, *Advances in Applied Probability* **18**, 211–220.

Brandt, A., Franken, P. and Lisek, B.: 1990, *Stationary Stochastic Models*, Wiley, Chichester.

Breiman, L.: 1965, On some limit theorems similar to the arc-sin law, *Theory of Probability and Its Applications* **10**, 323–331.

Brémaud, P.: 1999, *Markov Chains - Gibbs Fields, Monte Carlo Simulation and Queues*, Vol. 31 of *Texts in Applied Mathematics*, Springer, New York.

Brockwell, P. J. and Davis, R. A.: 1991, *Time Series: Theory and Methods*, Springer Series in Statistics, 2nd edn, Springer, New York.

Brown, J. R.: 1976, *Ergodic Theory and Topological Dynamics*, Academic Press, New York.

Cai, J.: 1994, A Markov model of switching-regime ARCH, *Journal of Business & Economic Statistics* **12**, 309–316.

Carrasco, M. and Chen, X.: 2002, Mixing and moment properties of various GARCH and stochastic volatility models, *Econometric Theory* **18**, 17–39.

Chen, M. and An, H. Z.: 1998, A note on the stationarity and the existence of moments of the GARCH model, *Statistica Sinica* **8**, 505–510.

Ciarlet, P. G.: 1989, *Introduction to numerical linear algebra and optimisation*, Cambridge Texts in Applied Mathematics, Cambridge University Press, Cambridge.

Davis, R. A. and Mikosch, T.: 1998, The sample autocorrelation of heavy-tailed processes with applications to ARCH, *Annals of Statistics* **26**, 2049–2080.

Davis, R. A. and Resnick, S. I.: 1996, Limit theory for bilinear processes with heavy-tailed noise, *Annals of Applied Probability* **6**, 1191–1210.

Douc, R., Moulines, E. and Rydén, T.: 2004, Asymptotic properties of the maximum likelihood estimator in autoregressive models with Markov regime, *Annals of Statistics* **32**, 2254–2304.

Doucet, A., Logothetis, A. and Krishnamurthy: 2000, Stochastic sampling algorithms for state estimation of jump Markov linear systems, *IEEE Transactions on Automatic Control* **45**, 188–202.

Doukhan, P.: 1994, *Mixing*, Vol. 85 of *Lecture Notes in Statistics*, Springer, New York.

Dueker, M. J.: 1997, Markov switching in GARCH processes and mean-reverting stock-market volatility, *Journal of Business & Economic Statistics* **15**, 26–34.

Duflo, M.: 1997, *Random Iterative Models*, Vol. 34 of *Stochastic Modelling and Applied Probability*, Springer, Berlin.

Dunford, N. and Schwartz, J. T.: 1958, *Linear Operators - Part I: General Theory*, Interscience Publishers, New York.

Embrechts, P., Klüppelberg, C. and Mikosch, T.: 1997, *Modelling Extremal Events for Insurance and Finance*, Vol. 33 of *Stochastic Modelling and Applied Probability*, Springer, Berlin.

Engle, R. F.: 1982, Autoregressive conditional heteroskedasticity with estimates of the variance of United Kingdom inflation, *Econometrica* **50**, 987–1008.

Fasen, V., Klüppelberg, C. and Lindner, A.: 2004, Extremal behavior of stochastic volatility models, *Preprint*, Centre for Mathematical Sciences, Munich University of Technology, Garching, Germany. available from: www.ma.tum.de/stat/Papers.

Feigin, P. D. and Tweedie, R. L.: 1985, Random coefficient autoregressive processes: A Markov chain analysis of stationarity and finiteness of moments, *Journal of Time Series Analysis* **6**, 1–14.

Francq, C. and Roussignol, M.: 1998, Ergodicity of autoregressive processes with Markov switching and consistency of the maximum likelihood estimator, *Statistics* **32**, 151–173.

Francq, C., Roussignol, M. and Zakoïan, J.-M.: 2001, Conditional heteroskedasticity driven by hidden Markov chains, *Journal of Time Series Analysis* **22**, 197–220.

Francq, C. and Zakoïan, J.-M.: 2001, Stationarity of multivariate Markov-switching ARMA models, *Journal of Econometrics* **102**, 339–364.

Francq, C. and Zakoïan, J.-M.: 2002, Comments on the paper: "Some properties of vector autoregressive processes with Markov-switching coefficients", *Econometric Theory* **18**, 815–818.

Francq, C. and Zakoïan, J.-M.: 2004, $L^2$- structures of standard and switching-regime GARCH models and their implications for statistical inference, *Discussion paper*, Université du Littoral-Côte d'Opale, LMPA J. Liouville, Calais, France.

Fuh, C.-D.: 2004, Asymptotic operating characteristics of an optimal change point detection in hidden Markov models, *Annals of Statistics* **32**, 2305–2339.

Furstenberg, H. and Kesten, H.: 1960, Products of random matrices, *Annals of Mathematical Statistics* **31**, 457–469.

Goldie, C. M.: 1991, Implicit renewal theory and tails of solutions of random equations, *Annals of Applied Probability* **1**, 126–166.

Goldie, C. M. and Grübel, R.: 1996, Perpetuities with thin tails, *Advances in Applied Probability* **28**, 463–480.

Goldie, C. M. and Maller, R. A.: 2000, Stability of perpetuities, *Annals of Probability* **28**, 1195–1218.

Goldsheid, I. Y.: 1991, Lyapunov exponents and asymptotic behaviour of the product of random matrices, *in* L. Arnold, H. Crauel and J.-P. Eckmann (eds), *Lyapunov Exponents - Proceedings of a Conference held in Oberwolfach, May 28 - June 2, 1990*, Vol. 1486 of *Lecture Notes in Mathematics*, Springer, Berlin, pp. 23–37.

Gray, S. F.: 1996, Modeling the conditional distributions of interest rates as a regime-switching process, *Journal of Financial Economics* **42**, 27–62.

Grey, D. R.: 1994, Regular variation in the tail behaviour of random difference equations, *Annals of Applied Probability* **4**, 169–183.

Haas, M., Mittnik, S. and Paolella, M. S.: 2004, A new approach to Markov-switching GARCH models, *Journal of Financial Econometrics* **2**, 493–530.

Hamilton, J. D.: 1989, A new approach to the economic analysis of nonstationary time series and the business cycle, *Econometrica* **57**, 357–384.

Hamilton, J. D.: 1990, Analysis of time series subject to changes in regime, *Journal of Econometrics* **45**, 39–70.

Hamilton, J. D.: 1994, *Time Series Analysis*, Princeton University Press, Princeton.

Hamilton, J. D. and Raj, B. (eds): 2002, *Advances in Markov-switching Models - Applications in Business Cycle Research and Finance*, Studies in Empirical Economics, Physica-Verlag, Heidelberg.

Hamilton, J. D. and Susmel, R.: 1994, Autoregressive conditional heteroskedasticity and changes in regime, *Journal of Econometrics* **64**, 307–333.

Heuser, H.: 1992, *Funktionalanalysis*, Mathematische Leitfäden, 3rd edn, Teubner, Stuttgart.

Hille, E. and Phillips, R. S.: 1957, *Functional Analysis and Semi-Groups*, Vol. 31 of *Colloqium Publications*, revised edn, American Mathematical Society, Providence.

Hult, H. and Lindskog, F.: 2002, Multivariate extremes, aggregation and dependence in elliptical distributions, *Advances in Applied Probability* **34**, 587–608.

Hult, H. and Lindskog, F.: 2004, Extremal behaviour of regularly varying stochastic processes, *Stochastic Processes and their applications* . forthcoming.

Karlsen, H. A.: 1990a, Doubly stochastic vector AR(1) processes, *Technical Report*, Department of Mathematics, University of Bergen, Bergen, Norway.

Karlsen, H. A.: 1990b, Existence of moments in a stationary stochastic difference equation, *Advances in Applied Probability* **22**, 129–146.

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

Kesten, H.: 1974, Renewal theory for functionals of a Markov chain with general state space, *Annals of Probability* **2**, 355–386.

Kingman, J. F. C.: 1968, The ergodic theory of subadditive stochastic processes, *Journal of the Royal Statistical Society, Series B (Methodological)* **30**, 499–510.

Kingman, J. F. C.: 1973, Subadditive ergodic theory (with discussion), *Annals of Probability* **1**, 883–909.

Klaassen, F.: 2002, Improving GARCH volatility forecasts with regime-switching GARCH, *Empirical Economics* **27**, 263–394.

Klüppelberg, C. and Pergamenchtchikov, S.: 2003, Renewal theory for functionals of a Markov chain with compact state space, *Annals of Probability* **31**, 2270–2300.

Klüppelberg, C. and Pergamenchtchikov, S.: 2004, The tail of the stationary distribution of a random coefficient AR(q) model, *Annals of Applied Probability* **14**, 971–1005.

Konstantinides, D. G. and Mikosch, T.: 2004, Large deviations and ruin probabilities for solutions to stochastic recurrence equations with heavy-tailed innovations, *Annals of Probability* . to appear.

Krengel, U.: 1985, *Ergodic Theorems*, Vol. 6 of *de Gruyter Studies in Mathematics*, Walter de Gruyter, Berlin.

Krishnamurthy, V. and Rydén, T.: 1998, Consistent estimation of linear and non-linear autoregressive models with Markov regime, *Journal of Time Series Analysis* **19**, 291–307.

Krolzig, H.-M.: 1997, *Markov-Switching Vector Autoregressions*, Vol. 454 of *Lecture Notes in Economics and Mathematical Systems*, Springer, Berlin.

Leadbetter, M. R.: 1974, On extreme values in stationary sequences, *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete* **28**, 289–303.

Leadbetter, M. R., Lindgren, G. and Rootzén, H.: 1983, *Extremes and Related Properties of Random Sequences and Processes*, Springer Series in Statistics, Springer, New York.

LePage, E.: 1983, Théorèmes de renouvellement pour les produits de matrices aléatoires. Équations aux différences aléatoires, *Publ. sém. math.*, Univ. Rennes I, Rennes, France.

Lindner, A. and Maller, R.: 2004, Lévy integrals and the stationarity of generalised Ornstein-Uhlenbeck processes, *Preprint*, Centre for Mathematical Sciences, Munich University of Technology, Garching, Germany. available from: www.ma.tum.de/stat/Papers.

Lindskog, F.: 2004, *Multivariate Extremes and Regular Variation for Stochastic Processes*, Dissertation, Department of Mathematics, ETH Zürich, Zürich, Switzerland. available from: `e-collection.ethbib.ethz.ch/diss/`.

Loève, M.: 1977, *Probability Theory I*, Vol. 45 of *Graduate Texts in Mathematics*, 4th edn, Springer, New York.

Loève, M.: 1978, *Probability Theory II*, Vol. 46 of *Graduate Texts in Mathematics*, 4th edn, Springer, New York.

Loynes, R. M.: 1965, Extreme values in uniformly mixing stationary stochastic processes, *Annals of Mathematical Statistics* **36**, 993–999.

Meyn, S. P. and Tweedie, R. L.: 1993, *Markov Chains and Stochastic Stability*, Communications and Control Engineering Series, Springer, London.

Mikosch, T.: 1999, Regular variation, subexponentiatility and their applications in probability theory, *Lecture notes for the workshop "Heavy Tails and Queues"*, EURANDOM Institute, Eindhoven, The Netherlands. available from `www.math.ku.dk/~mikosch`.

Mikosch, T.: 2003, Modeling dependence and tails of financial time series, *in* B. Finkenstädt and H. Rootzén (eds), *Extreme Values in Finance, Telecommunications, and the Environment*, Chapman and Hall, London, pp. 185–286.

Mikosch, T. and Stărică, C.: 2000, Limit theory for the sample autocorrelations and extremes of a GARCH(1,1) process, *Annals of Statistics* **28**, 1427–1451.

Nicholls, D. F. and Quinn, B. G.: 1982, *Random Coefficient Autoregressive Models: An Introduction*, Vol. 11 of *Lecture Notes in Statistics*, Springer, New York.

Poskitt, D. S. and Chung, S.-H.: 1996, Markov chain models, time series analysis and extreme value theory, *Advances in Applied Probability* **28**, 405–425.

Resnick, S. I.: 1987, *Extreme Values, Regular Variation and Point Processes*, Vol. 4 of *Applied Probability*, Springer, New York.

Resnick, S. I.: 1992, *Adventures in Stochastic Processes*, Birkhäuser, Boston.

Resnick, S. I. and Willekens, E.: 1991, Moving averages with random coefficients and random coefficient autoregressive models, *Communications in Statistics - Stochastic Models* **7**, 511–525.

Rossi, A. and Gallo, G. M.: 2003, Volatility estimation via hidden Markov models, *Proceedings of the CFS Workshop "New Directions in Financial Risk Management", Frankfurt, Nov 3-4,2003*, Center for Finacial Studies, Frankfurt, Germany.

Rudin, W.: 1976, *Principles of Mathematical Analysis*, 3rd edn, McGraw-Hill, Auckland.

Saporta, B. d.: 2003, Renewal theorem for a system of renewal equations, *Annales de l'Institut Henri Poincaré (B) Probabilitiés et Statistiques* **39**, 823–838.

Saporta, B. d.: 2004a, *Etude de la solution stationnaire de l'équation* $Y_{n+1} = a_n Y_n + b_n$ *à coefficients aléatoires*, PhD thesis, Institut de Recherche Mathématique de Rennes, Université de Rennes I, Rennes, France.

Saporta, B. d.: 2004b, Tail of the stationary solution of the stochastic equation $Y_{n+1} = a_n Y_n + b_n$ with Markovian coefficients, *Stochastic Processes and their Applications* . forthcoming.

Saporta, B. d., Guivarc'h, Y. and LePage, E.: 2004, On the multidimensional stochastic equation $Y_{n+1} = A_n Y_n + B_n$, *Comptes Rendus Mathematique* **339**, 499–502.

Shiryaev, A. N.: 1996, *Probability*, Vol. 95 of *Graduate Texts in Mathematics*, 2nd edn, Springer, New York.

Timmermann, A.: 2000, Moments of Markov switching models, *Journal of Econometrics* **96**, 75–111.

Tugnait, J. K.: 1982, Adaptive estimation and identification for discrete systems with Markov jump parameters, *IEEE Transactions on Automatic Control* **27**, 1054–1065. (correction: vol. 29, 1984, p. 286).

Vervaat, W.: 1979, On a stochastic difference equation and a representation of non-negative infinitely divisible random variables, *Advances in Applied Probability* **11**, 750–783.

Werner, D.: 2002, *Funktionalanalysis*, 4th edn, Springer, Berlin.

Wong, C. S. and Li, W. K.: 2001, On a mixture autoregressive conditional heteroscedastic model, *Journal of the American Statistical Association* **96**, 982–995.

Yang, M.: 2000, Some properties of vector autoregressive processes with Markov-switching coefficients, *Econometric Theory* **16**, 23–43.

Yao, J.: 2001, On square-integrability of an AR process with Markov switching, *Statistics & Probability Letters* **52**, 265–270.

Yao, J. F. and Attali, J. G.: 2000, On stability of nonlinear AR processes with Markov switching, *Advances in Applied Probability* **32**, 394–407.

Zhang, J. and Stine, R. A.: 2001, Autocovariance structure of Markov regime switching models and model selection, *Journal of Time Series Analysis* **22**, 107–124.