ON MARKOV CHAINS OF FINITE RANK

PROEFSCHRIFT

TER VERKRIJGING VAN DE GRAAD VAN DOCTOR IN DE TECHNISCHE WETENSCHAPPEN AAN DE TECHNISCHE HOGESCHOOL EINDHOVEN, OP GEZAG VAN DE RECTOR MAGNIFICUS, PROF. DR. S.T.M. ACKERMANS, VOOR EEN COMMISSIE AANGEMELEN DOOR HET COLLEGE VAN DEKANEN IN HET OPENBAAR TE VERDENIGEN OP DINSDAG 1 FEBRUARI 1983 TE 16.00 UUR

DOOR

ÆGLE HJALMAR HOEKSTRA

geboren te Haarlem

1983

MATHEMATICUM CENTRUM, AMSTERDAM
Dit proefschrift is goedgekeurd
door de promotoren

Prof.dr. F.W. Steutel
en
Prof.dr. J.Th. Runnenburg.
Aan mijn ouders
CONTENTS

0. INTRODUCTION
0.1. Motivation 1
0.2. Summary 2
0.3. Notations 3

1. THE MODEL
1.1. Definition, basic properties and linear transformations 4
1.2. The kernel matrix 10
1.3. Interpretation 16
1.4. Two special cases 19
1.4.1. Markov operators 19
1.4.2. Hergenrother's bivariate distribution function 20
1.5. Examples 22

2. EIGENVALUES
2.1. General results 26
2.2. The eigenvalue \lambda; classification of states 27
2.3. The eigenvalues of modulus \lambda; periodicity 31
2.4. The location of eigenvalues 43
2.5. Examples 47

3. LIMIT DISTRIBUTIONS
3.1. The invariant distribution 58
3.2. The central limit theorem 58
3.3. Expectation and variance of \Sn 63
3.4. Extreme values 72
3.5. Renewal theory 77
4. GENERALIZATIONS AND APPLICABILITY

4.1. Generalizations

4.1.1. Multidimensional central limit theorem

4.1.2. Continuous-time analogue

4.2. Approximation by kernels of lower rank

4.2.1. Approximation and invariant distribution

4.2.2. Correlation coefficients

4.3. Applicability

APPENDIX

A.1. The spectral decomposition of a matrix

A.2. Miscellaneous

REFERENCES

SAMENVATTING

CURRICULUM VITAE
CHAPTER 0. INTRODUCTION

0.1. Motivation

There exists a vast literature on Markov chains. A Markov chain describes a process which moves from state to state in discrete steps. The transition to a next state is stochastic, but depends only on the present state and not on "the past".

The best known and most widely studied case is the (time-homogeneous) finite Markov chain, where the state space $S$, i.e. the set of possible states, is finite, e.g. $S = \{1, 2, \ldots, s\}$. The probability of a transition from state $j$ to state $k$ is denoted by $p_{jk}$, and the matrix $P = (p_{jk})_{j,k=1}^s$ is called the transition matrix of the chain. In general, any square matrix with nonnegative elements and row sums 1 is called a transition matrix. An essential property of the finite Markov chains is that the $n$-step transitions ($n \geq 1$) are given by the matrix $P^n$. An extension of the finite Markov chain is the denumerable chain, where the state space $S$ is countably infinite.

If one wants a more general state space, e.g. an interval of the real line, there is no theoretical problem. Let $(S,S)$ be the measurable space. The probability of a transition from $x \in S$ to $y \in S$ is denoted by $P(\cdot|x)$, and now $P(\cdot|x)$ is for each $x \in S$ a probability measure on $(S,S)$. It is called the transition kernel of the chain. The $n$-step transition probabilities $P^{(n)}(\cdot|x)$ are defined recursively by

$$p^{(n)}(\cdot|x) = \int_S p^{(n-1)}(\cdot|y)P(dy|x) \quad \text{for } n \geq 2.$$  

There is, however, the practical problem that in general $P^{(n)}(\cdot|x)$ cannot easily be computed. Hence it is difficult to obtain detailed results about these general Markov chains.

The same practical problem of computational difficulties often necessitates the assumption of independence, when a sequence of random variables $X_0, X_1, X_2, \ldots$ is considered. For instance, in waiting-time and renewal theory it usually is assumed that the intervals between successive
renewals are independent and that they have the same distribution. RUNNENBURG [28] and [29] considered renewal theory for Markov-dependent random variables. In RUNNENBURG [28], as an example the following type of transition kernel was introduced:

\[ P(A|x) = \sum_{j=1}^{r} a_j(x) B_j(A), \]

where \( r \) is finite. This example arises as a natural extension of the independent case. For \( r = 1 \) the kernel in (0.2) describes a sequence of independent random variables, while for \( r \geq 2 \), \( a_j(x) \geq 0 \) and the \( B_j \) probability measures, the kernel simply is a convex combination of distributions. One may, however, allow the \( a_j \) and \( B_j \) to be arbitrary functions (measures), as long as \( P(A|x) \) is a transition kernel for each \( x \).

The kernels of type (0.2) are called of finite rank. This term was in the present context first used by KINGMAN [19]. A justification for this expression is given at the end of section 1.1. Markov chains of finite rank have the advantage of being more general than finite Markov chains (which are included as a special case) but having comparable computational accessibility; their transition probabilities are also governed by powers of finite matrices.

0.2. Summary

In chapter 1 the chains of finite rank are formally introduced and the kernel matrix, which takes over the position of the transition matrix of a finite Markov chain, is defined. The eigenvalues of kernel matrices are studied in chapter 2; just as in the case of finite Markov chains this leads to a classification of the state space. In section 4 of this chapter some numerical results about the eigenvalues, which were obtained with the generous help of Dr. J. de Jong and R. Kool, are presented. Limit theorems are the topic of chapter 3; the (existence of an) invariant distribution is studied, a central limit theorem for chains of finite rank is proved, two extreme-value theorems are derived, and finally some elementary renewal theory is considered. All proofs are essentially based on simple matrix theory, in particular on spectral decompositions. Chapter 4 contains some generalizations and possible approximations and applications. Several results in matrix theory, especially on spectral behaviour, which are used in the chapters 2, 3 and 4, are proved in an appendix.
0.3. Notations

Random variables are denoted by capitals. The expectation of a random variable $X$ is denoted by $\mathbb{E}X$ (in the spectral decomposition of a matrix we write $E_k$ for the idempotent matrix corresponding to the eigenvalue $\lambda_k$; confusion seems unlikely).

Underlined symbols denote column vectors; dimensions, when not indicated, are clear from the context. If $\mathbf{y} \in \mathbb{F}^r$, then $y_j$ is the $j$-th component of $\mathbf{y}$ ($j = 1, \ldots, r$) and $\mathbf{y}^T$ is the transpose of $\mathbf{y}$, i.e., the row vector with $j$-th component $y_j$. Hence, if $u, \mathbf{v} \in \mathbb{F}^r$, then $\mathbf{y} \cdot \mathbf{v} = \sum_{j=1}^{r} y_j v_j$ is the inner product of the vectors, whereas $\mathbf{y}^T \mathbf{v}$ is an $r \times r$ matrix. Special vectors that often occur are $\mathbf{0}$ (with all components equal to 0) and $\mathbf{1}$ (with all components equal to 1). If $\mathbf{v}$ is an eigenvector of a matrix $C$ corresponding to the eigenvalue $\lambda$, then $\mathbf{v}$ is said to be a $\lambda$-eigenvector of $C$ (see further the beginning of chapter 2).

$I_r$ is the $r \times r$ unit matrix (the index $r$ may be omitted. The matrix $0$ is the matrix with all elements equal to 0.

The symbols $O$ and $o$ have their usual meaning: if $g(x) > 0$, then

$$\varepsilon(x) = O(g(x)) \ (x \to a) \iff \frac{|f(x)|}{g(x)} \text{ is bounded for } x \to a,$$

$$o(x) = o(g(x)) \ (x \to a) \iff \frac{|f(x)|}{g(x)} \to 0 \text{ as } x \to a.$$

Statements about vectors and matrices, such as e.g. $|y| \leq 1$ and $c_j \leq o(\varepsilon)$ ($a \to -\infty$), are always supposed to hold elementwise; i.e., in the given examples we would have $|y_j| \leq 1$ for all $j$ and $c_j^{(n)} \leq o(\varepsilon)$ ($n \to -\infty$) for all $j, k$.

The open square □ marks the end of a definition, an assumption, a remark, an example, or the statements of a lemma or a theorem. The closed square ■ marks the end of a proof.

Concerning our notation for vectors we make the following remark. Several mathematicians in The Netherlands use underlined symbols to denote random variables. Efforts have been made in the past to propagate this notation, which had been introduced by Van Danzig, to other countries. Unfortunately, these efforts have not been very successful. Our use of underlined symbols for vectors may certainly not be taken for a dislike of the "Dutch notation", which has indeed many advantages. Yet we have chosen here to follow the common usage in English literature on probability theory, denoting random variables by (non-underlined) capitals. Thereupon the present vector notation was adopted as being the most practical in our situation.
CHAPTER 1. THE MODEL

The motivation for our model has already been given in the previous introductory chapter. In section 1.1 we shall properly define the model and prove some of its basic properties. In section 1.2 the kernel matrix \( C \) is introduced, which plays a central role in the rest of our investigations, and some elementary properties of this matrix are determined. The possibility of an interpretation is considered in section 1.3. Two special cases are discussed in section 1.4, and finally some simple numerical examples are given in section 1.5.

Many of the results in the sections 1.1, 1.2 and 1.3 can also be found in \( \text{HUHNEBURG AND STEFEL} \{30\} \).

1.1. Definition, basic properties and linear transformations

**DEFINITION 1.1.1.** Let \((\Omega, \mathcal{S}, \mathbb{P})\) be a measurable space. A time-homogeneous Markov chain \( X_0, X_1, \ldots \), taking values in \( S \), is said to be of finite rank if there are complex-valued \( S \)-measurable functions \( a_1, \ldots, a_r \) and complex-valued measures \( B_1, \ldots, B_r \) on \( S \), such that the transition kernel \( P(\cdot|\cdot) \) has the form

\[
P(A|x) = \sum_{j=1}^{r} a_j(x) B_j(A) \quad \text{for all } x \in S, A \in \mathcal{S}.
\]

Also the kernel itself is then said to be of finite rank.

In a Markov chain of finite rank the conditional probabilities

\[
P(A|x) = P(X_{n+1} \in A | X_n = x)
\]

are defined for all \( x \in S \), see (1.1.1). In a general Markov chain this does not have to be so.

We shall often be concerned with the special case where \( S = \mathbb{R} \) and \( S = \mathbb{R}(\mathbb{R}) \). We then write for \( x, y \in \mathbb{R} \):

\[
P(y|x) := P((-\infty, y] | x)\]

\[
B_j(y) := B_j((-\infty, y])
\]
P(y|x) is now a distribution function and so is \( B_j(y) \) if \( B_j \) is a probability measure.

Consistent with our notations listed on page 3 we write \( \mathbf{a} \) and \( \mathbf{B} \) for the (column) vector with elements \( a_j \) and \( B_j \), respectively. In this vector notation (1.1.1) becomes

\[
(1.1.3) \quad P(A|x) = \mathbf{a}^T \mathbf{B}(A),
\]

where \( \mathbf{a}^T \) denotes the transpose of \( \mathbf{a} \). Dimensions, when not indicated, will be clear from the context.

We now further investigate (1.1.1), and we shall show that no generality is lost by assuming that the \( B_j \) are probability measures and that the \( a_j(x) \) are real-valued. First we give one more definition.

**Definition 1.1.2.** The representation (1.1.1) of \( P(A|x) \) is called **minimal** if \( P(A|x) \) cannot be represented as a sum of less than \( r \) terms. In that case \( r \) is called the **rank** of the chain.

Clearly, if the \( a_j \) or the \( B_j \) are linearly dependent, then \( P(A|x) \) can be written as a sum of less than \( r \) terms, and the representation is not minimal. For the moment we assume linear independence, and in Theorem 1.1.6 we shall prove that linear independence is equivalent to minimality of the representation.

**Remark 1.1.3.** The measures \( B_j \) are said to be **linearly independent** if from

\[
(1.1.4) \quad \sum_{j=1}^{r} B_j(A) = 0 \quad \text{for all} \ A \in S
\]

it follows that \( B_1 = \ldots = B_r = 0 \).

Here \( S \) can sometimes be replaced by a smaller class of measurable sets. For instance, if \( S = \mathbb{R} \), \( S = \mathcal{B}(\mathbb{R}) \) and if \( B_j(\mathbb{R}) = - \) for all \( j \), then (1.1.4) follows from

\[
(1.1.5) \quad \sum_{j=1}^{r} B_j(y) = 0 \quad \text{for all} \ y \in \mathbb{R},
\]

since the \( \sigma \)-algebra \( \mathcal{B}(\mathbb{R}) \) is generated by the sets of the form \( (-\infty, y] \). So the linear independence of the measures \( B_j \) on \( \mathcal{B}(\mathbb{R}) \) is then equivalent to the linear independence of the functions \( B_j(y) \) on \( \mathbb{R} \).
THEOREM 1.1.4. It is no restriction to assume that the $a_j(x)$ are real-valued and that the $B_j$ are probability measures. In that case the $a_j(x)$ satisfy

\begin{equation}
(1.1.6) \quad \sum_{j=1}^{r} a_j(x) = 1 .
\end{equation}

PROOF. By the assumed linear independence of the $a_j(x)$ there exists at least one $r$-tuple $x_1, \ldots, x_r$ such that $\det[a_k(x_k)] \neq 0$ (this is proved in the appendix as lemma A.2.4).

If we define the matrix $T = (t_{kj})$ by $t_{kj} = a_k(x_j)$, then $T$ is a nonsingular $r \times r$ matrix. Write

\begin{align}
(1.1.7) \quad & T^*(x) := T(x) T^{-1} , \\
& B^*(A) := TB(A) ,
\end{align}

then

\begin{equation}
(1.1.8) \quad P(A|x) = T^*(x) B^*(A) ,
\end{equation}

where

\begin{equation}
(1.1.9) \quad B_k^* = \sum_{j=1}^{r} t_{kj} B_j = \sum_{j=1}^{r} a_k(x_j) B_j = P(x_k | x)
\end{equation}

is a probability measure for each $k$. By the assumed linear independence of the $B_j$ it follows that the $B_k^*$ are linearly independent, and hence that the $a_k^*(x)$ are real-valued (consider the imaginary parts).

Relation (1.1.6) follows by substituting $A = \phi$ in (1.1.8).

□

LEMMA 1.1.5. In every representation (1.1.1) where the $a_j$ and the $B_j$ are linearly independent, the functions $a_j(x)$ are bounded and the measures $B_j$ are finite.

PROOF. We have seen in the proof of theorem 1.1.4 that a linear transformation $T$ exists such that $B = T^{-1} B^*$ with $B^*$ a probability (vector) measure. Hence the $B_j$ are finite.

As an analogue to lemma A.2.4 it can be shown that there exist $A_1, \ldots, A_r$ such that $\det[B_j(A_k)] \neq 0$. Now replace $T^{-1}$ by the transformation $T$ with elements $u_{jk} = B_j(A_k)$. Then we obtain functions $a_j^*(x)$ and measures $B_j^*$, where as a counterpart of (1.1.9) we get
(1.1.10) \[ a^*_k(x) = \sum_{j=1}^{r} a_j(x) u_{jk} = \sum_{j=1}^{r} a_j(x) B_j^* (A_k) = F(A_k | x) . \]

So the \( a^*_k \) are bounded, and hence the \( a_j \) are bounded. \( \square \)

**Theorem 1.1.6.** The representation (1.1.1) is minimal if and only if both the \( a_j(x) \) and the \( B_j \) are linearly independent.

**Proof.** The "only if"-part has already been dealt with.

Now suppose both the \( a_j(x) \) and the \( B_j \) are linearly independent, and that

(1.1.11) \[ \sum_{j=1}^{r} a_j(x) B_j = \sum_{j=1}^{r} a^*_j(x) B_j^* \quad \text{for all } x \in S . \]

By lemma 1.1.3 there exist \( x_1, \ldots, x_r \) such that \( \det(a_j(x_k)) \neq 0 \). Consider the \( r \) vectors \( a^*_1(x_1), \ldots, a^*_r(x_r) \in \mathbb{R}^{r-1} \). Even if the functions \( a^*_j \) are linearly independent, these vectors are linearly dependent, so there exist \( c_1, \ldots, c_r \), not all zero, such that \( \sum_{k=1}^{r} c_k a^*_j(x_k) = 0 \) for \( j = 1, \ldots, r-1 \). But then we have

(1.1.12) \[ \sum_{j=1}^{r} \left( \sum_{k=1}^{r} c_k a_j(x_k) \right) B_j = \sum_{k=1}^{r} c_k \sum_{j=1}^{r} a_j(x_k) B_j^* = \sum_{j=1}^{r} \left( \sum_{k=1}^{r} c_k a^*_j(x_k) \right) B_j^* = 0 . \]

By the linear independence of the \( B_j \), we conclude that \( \sum_{k=1}^{r} c_k a_j(x_k) = 0 \) for \( j = 1, \ldots, r \), hence \( c_k = 0 \) for \( k = 1, \ldots, r \). Thus a contradiction is obtained, which means that there are no \( a^*_j \) and \( B_j^* \) satisfying (1.1.11). \( \square \)

**Lemma 1.1.7.** Suppose

(1.1.13) \[ P(A | x) = \sum_{j=1}^{r} a_j(x) B_j^* (A) = \sum_{j=1}^{r} a^*_j(x) B_j^* (A) \quad \text{for all } A \in S , \]

where \( r \) is minimal. Then there is a nonsingular \( r \times r \) matrix \( T \), such that

(1.1.14) \[ T^* (x) = T^* (x) T^{-1} , \quad B^* = TB . \]
PROOF. Choose any $r$-tuple $(x_1, \ldots, x_r)$, such that the matrix $D$ with elements $a_{jk} = a_{jk}(x_j)$ is nonsingular, and let $D^*$ be the matrix with elements $a_{jk}^* = a_{jk}(x_j)$. Further, choose an $r$-tuple $(A_1, \ldots, A_r)$, such that the matrix $E$ with elements $e_{jk} = E_{jk}(A_k)$ is nonsingular, and let $E^*$ be the matrix with elements $e_{jk}^* = E_{jk}(A_k)$. Let $T := E^*E^{-1}$. From $T_a(x)x = T_a^*(x)x^*$ it follows that $DE = D^*E^*$. As $D$ and $E$ are nonsingular, it follows that $D^*$ and $E^*$, and hence $T$, are nonsingular. For all $x$ we have $T_a(x)x = T_a^*(x)x^*$, or $T_a^*(x) = T_a(x)T^{-1}$. For all $y$ we have $DB = D^*B^*$, or $B^* = (B^*)^{-1} DB - TB$. □

**Definition 1.1.8.** Representation (1.1.1) is called standard if it is minimal and the $R_j$ are probability measures, □

**Assumption.** Unless otherwise stated, the representation of the kernel is assumed to be standard.

It should be noted, however, that even standard representations are nonunique. To a given standard representation one can always apply a linear transformation $T$, as in (1.1.7), where $T$ is any nonsingular $r \times r$ transition matrix. The resulting new representation is then again standard.

The set $S$, appearing in definition 1.1.1, is usually called the state space of the chain. The process may in fact be concentrated on a subset $S_0$ of $S$, i.e. we may have

\[(1.1.15) \quad P(S_0|x) = 1 \quad \text{for all} \ x \not\in S .\]

The following lemma states that then every $R_j$ is concentrated on $S_0$.

**Lemma 1.1.9.** If (1.1.15) holds, then

\[(1.1.16) \quad R(S_0) = 1 .\]

**Proof.** For all $x \in S$ we have, with $\bar{S} := S \setminus S_0$,

\[(1.1.17) \quad P(\bar{S}|x) = T_a(x)E(\bar{S}) = 0 ,\]

hence $R(\bar{S}) = 0$, by the linear independence of the $\lambda_j$. □
If the process is concentrated on \( S_0 \), the values of \( a_j(x) \) for \( x \not \in S_0 \) are unimportant. Therefore we shall sometimes specify the \( a_j(x) \) only for \( x \in S_0 \). In such a case one may assign arbitrary values to \( a_j(x) \) for \( x \in S \setminus S_0 \).

The Markov chains of finite rank include all finite Markov chains as a special case. Let \( P = (p_{jk}) \) be the transition matrix of a Markov chain with state space \( S = \{s_1, \ldots, s_N \} \). We simply take

\[
\begin{align*}
  a_j(s_k) &= p_{kj}, \\
  b_j(A) &= 1_A(a_j),
\end{align*}
\]

(1.1.18) where \( 1_A \) is the indicator function of the set \( A \). The representation thus obtained is standard if and only if \( P \) is nonsingular.

If \( P \) has rank \( r \) we obtain a standard representation as follows. Select \( r \) linearly independent rows of \( P \) and let \( B \) be the \( r \times N \) matrix consisting of these rows. Suppose for simplicity that the rows are the first \( r \) ones of \( P \), so that \( B \) is equal to the upper \( r \times N \) part of \( P \). Let \( A = (a_{jk}) \) be the \( N \times r \) matrix given by

\[
A = \begin{bmatrix}
1 & 0 & 0 \\
0 & 1 & 0 \\
\vdots & \vdots & \ddots \\
0 & 0 & 1 \\
\end{bmatrix}
\]

(1.1.19)

where the \( a_{jk} \) are such that \( \sum_{k=1}^{r} a_{jk} b_{kt} = p_{jt} \) for \( j \geq r+1 \) and all \( t \). Then

(1.1.20) \hspace{1cm} P = AB,

and if we take

\[
\begin{align*}
  a_j(x_k) &= a_{kj}, \\
  b_j(A) &= \sum_{k} 1_{\{k \in \mathbb{E} \}} b_{jk}
\end{align*}
\]

(1.1.21)

we obtain a standard representation.
This construction is easily generalized for infinite transition matrices \( P \) of finite rank. This justifies the term "Markov chain of finite rank" for the model under consideration.

**Remark 1.1.10.** On the measurable space \((S,\mathcal{S})\) a measure \( m \) can be defined such that each measure \( B_j \) is absolutely continuous with respect to \( m \). If, for example, we take

\[
(1.1.22) \quad m := \frac{1}{r} \sum_{j=1}^{r} B_j ,
\]

then \( m \) is even a probability measure on \( S \).

Obviously \( P(\cdot|x) \) also is absolutely continuous with respect to \( m \) for all \( x \in S \). The measure \( m \) will play a role in section 4 of this chapter and in some other places.

The definition (1.1.22) of \( m \) depends on the choice of the (standard) representation of the kernel. However, any such \( m \) will do for all standard representations. To see this, suppose \( \mathcal{A}(x) \mathcal{B}(y) \) is such a representation; if \( m(A) = 0 \), then \( 0 = P(A|x) = \mathcal{A}(x) \mathcal{B}(A) \) for all \( x \in S \), hence \( \mathcal{B}(A) = 0 \).

1.2. The kernel matrix

Now consider the \( n \)-step transition distribution

\[
(1.2.1) \quad P^{(n)}(A|x) = \mathbb{P}(X_n \in A \mid X_0 = x) ,
\]

defined by

\[
(1.2.2) \quad P^{(n+1)}(A|x) := \int_S P^{(n)}(A|z) P(\mathcal{D}|x) \quad \text{for all } n \geq 1 .
\]

For \( P^{(2)}(A|x) \) we find

\[
(1.2.3) \quad P^{(2)}(A|x) = \int_S P(A|z) P(\mathcal{D}|x) =
\]

\[
= \sum_{k=1}^{r} \sum_{j=1}^{r} \sigma_k(x) \int_S \sigma_k(x) B_j(\mathcal{D}) B_k(A) .
\]
If we write

\[(1.2.4) \quad c_{jk} = \int_a b_j (s) d \alpha (s) \quad \text{for} \quad j, k = 1, \ldots, r, \]

relation (1.2.3) becomes, in vector notation,

\[(1.2.5) \quad p^{(2)} (A| x) = T (x) C^{(2)} (A) ,\]

where \( C \) is the \( r \times r \) matrix with elements \( c_{jk} \). By iteration we find

**Theorem 1.2.1.** The \( n \)-step transition probability can be written in the form

\[(1.2.6) \quad p^{(n)} (A| x) = T (x) C^{(n-1)} \bar{B} (A) ,\]

where the \( r \times r \) matrix \( C = (c_{jk}) \) is given by (1.2.4), and \( C^0 = I \), the unit matrix.

**Definition 1.2.2.** The matrix \( C \) with elements given by (1.2.4) is called the *kernel matrix* corresponding to the kernel with representation (1.1.1). \( C \) is called *minimal* or *standard* if the representation of the corresponding kernel is minimal or standard.

Theorem 1.2.1 represents the essential feature of Markov chains of finite rank: The behaviour of the \( n \)-step transition functions is governed by the behaviour of the \( n \)-th power of the finite kernel matrix \( C \), just as in the case of a finite Markov chain, where the \( n \)-th power of the transition matrix provides the \( n \)-step probabilities.

For finite Markov chains, theorem 1.2.1 follows directly from (1.1.20), since

\[(1.2.7) \quad P^n = (AB)^n = A (B A)^{n-1} B = AC^{n-1} \bar{B} .\]

The kernel matrix now takes the simple form \( C = B \bar{A} \).

For the analysis of a Markov chain with \( N \) states there is an advantage in our approach, using \( C^n \) rather than \( P^n \), if the rank \( r \) of the \( N \times N \) matrix \( P \) is less than \( N \).

The kernel matrix of a Markov chain of finite rank is no more unique than the representation of its kernel. Even if we restrict to minimal representations there is a whole set of corresponding kernel matrices. This set is given by
Theorem 1.2.3. All minimal kernel matrices of a fixed kernel are similar, i.e., for any two such matrices C and C' there exists a nonsingular T, such that \( C' = TCT^{-1} \).

Proof. If C corresponds to \( \{a(x)\} \) and \( C'' \) to \( \{a'(x)\} \), then by Lemma 1.1.7 \( \{a(x)\} = \{a(x)T\} \) and \( T = TB \). It follows from (1.2.5) that
\[
\{a(x)\}_{C''} = \{a(x)\}_{TCT^{-1}} \quad C'' \quad TB.
\]
Hence \( C'' = TCT^{-1} \) by the linear independence of the \( a_j \) and \( b_j \).

If C is a standard kernel matrix of a kernel \( F \), then by definition C is minimal for \( F \). Further we have

Lemma 1.2.4. A standard kernel matrix has row sums 1.

Proof. Apply (1.1.6) of Theorem 1.1.4 to definition (1.2.4) of \( c_{jk} \).

The set of all \( r \times r \) matrices that can occur as kernel matrices for Markov chains of rank \( r \) or less will be denoted by \( K(r) \). We have

Lemma 1.2.5. Every \( r \times r \) transition matrix \( F \) is an element of \( K(r) \).

Proof. Consider the representation given by (1.1.18). We obtain \( C = F \), hence \( F \in K(r) \).

In general, a matrix \( C \in K(r) \) is the kernel matrix of many different transition kernels. C may be minimal for one kernel, nonminimal for another. Consider, for example, the matrix \( C = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \). For the finite (2-state) Markov chain with \( \begin{bmatrix} 0 & 1 \\ 0 & 1 \end{bmatrix} \) as transition matrix, C is kernel matrix, but not minimal; the minimal kernel matrix in this instance is the \( 1 \times 1 \) matrix (1). But the finite (3-state) Markov chain with transition matrix
\[
(1.2.8) \quad P = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{bmatrix} \quad = \quad \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 \end{bmatrix}
\]
is of rank 2 and has \( C \) as a minimal (even standard) kernel matrix.

A matrix \( C \in K(r) \) will be called minimal if it is minimal for at least one kernel.

In view of (1.2.6) and (1.2.7) we want to compare kernel matrices, in particular standard ones, with transition matrices. With this in mind we derive some elementary properties.
Lemma 1.2.6. \( c^n \) is bounded for \( n \in \mathbb{N} \).

**Proof.** For the \( n \)-step transition kernel in standard representation we have

\[
P^{(n)}(A|x) = \frac{T_{\mathbb{R}}(x)}{S} \int_S p^{(n-1)}(A|y)B(dy) = T_{\mathbb{R}}(x)B^{(n)}(A),
\]

where the \( B_j^{(n)} \) are probability measures. Hence we obtain the relation

\[
T_{\mathbb{R}}(x) c^{n-1} \mathbb{R} = T_{\mathbb{R}}(x) B^{(n)} \quad \text{for all } x \in S.
\]

From the left hand side of (1.2.10) it is seen that \( c^{n-1} \mathbb{R} = c^n \) is a kernel matrix corresponding to the kernel \( P^{(n)}(\cdot|x) \), and that \( c^{n-1} \mathbb{R} = B^{(n)} \). It follows that

\[
|c_{jk}^{(n)}| = \left| \int_S a_k(x)B_j^{(n)}(dx) \right| \leq \int_S |a_k(x)|B_j^{(n)}(dx) \leq \sup_{x \in S} |a_k(x)| \quad \text{for all } n \geq 1,
\]

since by lemma 1.1.5 the \( a_j(x) \)'s are bounded. \( \square \)

Lemma 1.2.7. If \( C \in \mathbb{R}(x) \), \( n \geq 1 \), \( a_k \geq 0 \) \( i = 1, 2, \ldots, n \) and \( \sum_{i=1}^n a_i = 1 \), then \( \sum_{i=1}^n a_i C^i \in \mathbb{R}(x) \).

**Proof.** Suppose \( C \) is a kernel matrix corresponding to the kernel \( P(\cdot|x) = T_{\mathbb{R}}(x)B \). Define \( \bar{C} \) by \( \bar{T}_{\mathbb{R}}(x) := T_{\mathbb{R}}(x)c^{n-1} \), and consider \( P^{(n)}(\cdot|x) \) as a new kernel \( \bar{P}(\cdot|x) \) with

\[
\bar{P}(\cdot|x) = T_{\mathbb{R}}(x) c^{n-1} \mathbb{R} = T_{\mathbb{R}}(x) B.
\]

We then obtain the kernel matrix

\[
\bar{C} = \left( \int_S a_k(x)B_j^{(n)}(dx) \right) = \mathbb{C}^{n-1} \left( \int_S a_k(x)B_j^{(n)}(dx) \right) = \mathbb{C}^n.
\]

We can do this for \( i = 1, 2, \ldots, n \), and one now easily verifies that the matrix \( \sum_{i=1}^n a_i \mathbb{C}^i \) is a kernel matrix for the chain with kernel \( \sum_{i=1}^n a_i P^{(i)}(\cdot|x) \). \( \square \)
Lemma 1.2.7 might suggest that the set $K(r)$ is convex. However, in chapter 2 it will be shown by an example that $K(3)$ is not convex. In this respect kernel matrices differ from transition matrices.

**Theorem 1.2.8.** Every minimal $C$ in $K(r)$ is the limit of a sequence of matrices $C^{(k)}$ ($k = 1, 2, \ldots$), where the $C^{(k)}$ are kernel matrices for finite Markov chains of rank $r$.

**Proof.** There is no loss of generality in assuming that $C$ is standard, as we can apply the same linear transformation to all $C^{(k)}$. Suppose $C$ corresponds to the kernel $\sum_{j \neq i} a_j(x)b_j$. By Lemma 1.1.5 the $a_j$ are bounded, say $|a_j(x)| < \epsilon \in \mathbb{N}$ for all $x$ and all $j$.

For each fixed $k \in \mathbb{N}$ and $j \in \{1, \ldots, r\}$ the sets

$$A_{jk}^{(k)} = \{x : \frac{k}{\epsilon} < a_j(x) < \frac{k+1}{\epsilon}\} \quad (k = 1, \ldots, L+1, \ldots, L+L\}$$

form a measurable partition of $S$.

Keep $k$ fixed and let $A_1^{(k)}, \ldots, A_r^{(k)}$ be a measurable partition containing the $A_{jk}^{(k)}$. Take an arbitrary, fixed $n^{(k)}_j \in A^{(k)}_n$ for each $n \in \{1, \ldots, N^{(k)}\}$, for each $j$ define the function $a_j^{(k)}$ by

$$a_j^{(k)}(x) = a_j(x) \quad \text{for } x \in A_n^{(k)}, \quad n = 1, \ldots, N^{(k)}.$$ 

and let $B_j^{(k)}$ be the probability measure with mass $B_j^{(k)}(A_n^{(k)})$ at $a_j^{(k)}$. The $a_j^{(k)}$ are measurable stepfunctions, and for all $x$ and $j$ we have

$$|a_j^{(k)}(x) - a_j(x)| \leq \frac{1}{\epsilon}.$$ 

Furthermore the measure $P_j^{(k)}(\cdot|x)$, defined by

$$P_j^{(k)}(\cdot|x) = \frac{1}{\sum_{j=1}^{N^{(k)}} a_j^{(k)}(x)B_j^{(k)}},$$

is a transition kernel concentrated on the finite set $\{x^{(k)}_1, \ldots, x^{(k)}_{N^{(k)}}\}$, corresponding therefore to a finite Markov chain. Let $C^{(k)}$ be the corresponding kernel matrix. The elements $c_{jk}^{(k)}$ of $C^{(k)}$ satisfy

$$c_{jk}^{(k)} = \frac{1}{\epsilon} \int a_j^{(k)}(x)B_j^{(k)}(dx) = \frac{1}{\epsilon} \int a_j^{(k)}(x)B_j^{(k)}(dx).$$
Using Lebesgue's dominated convergence theorem \(|d_j^{(k)}(x)| < L_x\), for \(L > 0\) together with (1.2.16) we find for all \(j\) and \(k\)

\[
\lim_{n \to \infty} c_{jk}^{(n)} = \lim_{n \to \infty} \int \frac{a_j^{(n)}(x)b_j(x)dx}{\int a_j(x)b_j(x)dx} = \int \frac{a_j(x)b_j(x)dx}{\int a_j(x)b_j(x)dx} = c_{jk},
\]

in other words: \(\lim_{n \to \infty} C^{(n)} = C\).

The following simple property will prove to be useful.

**Lemma 1.2.9.** The trace of a kernel matrix is nonnegative, i.e.

\[
(1.2.20) \quad \text{tr}(C) = \sum_{j=1}^{r} c_{jj} \geq 0 \quad \text{for all } C \in K(r).
\]

**Proof.** A transition matrix has a nonnegative trace, since all its elements are nonnegative. Now (1.2.20) follows from theorem 1.2.8.

The product of two transition matrices is again a transition matrix. The set \(K(r)\), however, is not closed under matrix multiplication, as the following example shows. The two matrices

\[
(1.2.21) \quad C_1 = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}, \quad C_2 = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 1 & -1 & 1 \end{bmatrix}
\]

are both elements of \(K(3)\); \(C_1\) is a transition matrix (see lemma 1.2.5), and for \(C_2\) we refer to example 1.5.3. The product matrix

\[
(1.2.22) \quad C_1C_2 = \begin{bmatrix} 1 & 1 \\ -1 & 1 \\ 1 & 0 \end{bmatrix}
\]

however, has a negative trace, so \(C_1C_2 \not\in K(3)\) by lemma 1.2.9.
1.3. Interpretation

A Markov chain of rank 1 is simply a sequence of independent random variables: we have $a_1(x) \equiv 1$ and $B \equiv (1)$, hence

\[(1.3.1) \quad P^{(n)}(A|X_n) = B(\lambda) \quad \text{for all } x \in S \text{ and } n \geq 1,
\]

so that e.g., with $1 \leq k \leq 1$ and $F_0$ the distribution of $X_0$,

\[(1.3.2) \quad P(X_0 \in A_0, X_1 \in A_1, \ldots, X_k \in A_k) =
\]

\[= \int \left( \int \left( \int (z-k)(\lambda_z|y)\pi(k)(dy|z) \right) F_0(dx) \right) =
\]

\[= F_0(A_0)B_1(A_1)B_k(A_k) .
\]

Here $X, X_2, \ldots$ are identically distributed with distribution $B_1$.

Now consider a chain of rank $r \geq 2$. In a standard representation of the kernel the $a_j(x)$ need not be nonnegative (see example 1.5.1), but suppose that they are. In that case the $a_j(x)$ can be considered as probabilities, while the kernel matrix $B$ is a transition matrix. We introduce random variables $J_0, J_1, \ldots, \text{taking values in } S \times \{1, \ldots, r\}$, such that

\[(1.3.3) \quad P(\{X_k = A \mid J_k = j\} = B_j(A)
\]

\[P(\{J_{k+1} = j \mid X_k = x\} = a_j(x)
\]

for formal correctness one can take $\tilde{S} = (S \times \{0\}) \cup (\{0\} \times S(r))$ as state space for this Markov chain. Considering only the $X_n$ we recover our original chain of rank $r$, whereas the $J_n$ form a finite Markov chain with

\[(1.3.4) \quad P(\{J_{n+1} = k \mid J_n = j\} = \int_S B_j(dx)a_k(x) = c_{jk} .
\]

I.e. with transition matrix $B$. Furthermore it follows that

\[(1.3.5) \quad P(X_n \in A, J_{n+1} = k, J_0, X_1, \ldots, J_{n-1}, X_{n-1}; J_n = j) =
\]

\[= \int_A P(X_n \in A, J_{n+1} = k \mid J_n = j) = \int_A a_k(x)B_j(dx) .
\]
This means that we actually have an example of a semi-Markov process (cf. Pyke [26]). If \( S = \{0, w^* \} \), then \( X_n \) can be considered as the sojourn time of the process in state \( 0 \).

If the \( a_j(x) \) can take on negative values, they have no probabilistic interpretation. But the Markov chain as such can still be described in terms of a semi-Markov process as done above, if there exists at least one other representation of the kernel with nonnegative \( a_j(x) \) and probability measures \( B_j \). In section 2.4 it will be seen that for chains of rank \( r \geq 3 \) a standard representation with nonnegative \( a_j(x) \) does not necessarily exist. For rank 2 we have

**Theorem 1.3.1.** For every chain of rank 2 there is a standard representation of \( P(A|x) \) with nonnegative functions \( a_1(x) \) and \( a_2(x) \).

**Proof.** Let

\[
P(A|x) = a_1^*(x)B_1^*(A) + a_2^*(x)B_2^*(A)
\]

be an arbitrary standard representation of \( P(A|x) \). Define

\[
\ell := \inf a_1^*(x), \quad L := \sup a_1^*(x).
\]

An \( a_1^*(x) \) is bounded, \( \ell \) and \( L \) are finite; moreover \( \ell \neq L \), since otherwise the chain would be of rank 1.

It is easily verified that the measures \( B_1 \) and \( B_2 \), defined by

\[
B_1 := \ell B_1^* + (1-L)B_2^*,
\]

\[
B_2 := LB_1^* + (1-\ell)B_2^*,
\]

are probability measures. Applying the transformation

\[
T = \begin{pmatrix}
L & 1-L \\
\ell & 1-\ell
\end{pmatrix}, \quad T^{-1} = \frac{1}{L-\ell} \begin{pmatrix}
1-\ell & L-1 \\
\ell & L
\end{pmatrix}
\]

we find \( T B_1^* = \mathbb{1} \) and

\[
(a_1(x), a_2(x)) := (a_1^*(x), a_2^*(x)) T^{-1} = \frac{1}{L-\ell} ((1-L)a_1^*(x) - \ell a_2^*(x), (L-1)a_1^*(x) + L a_2^*(x)) = \]


\[ = \frac{1}{L - \ell_0} \left( a_1^*(x) - x, L - a_1^*(x) \right). \]

There is, of course, in this situation no reason for requiring a standard (which means minimal) representation. Whether a (possibly non-minimal) representation with nonnegative \( a_j \) always exists if \( r \geq 3 \), we do not know. We remark that it is always possible by a linear transformation to obtain \( a_j(x) \) between 0 and 1; if in the proof of theorem 1.1.4 we replace \( T^{-1} \) by \( (1.3, 11) \quad U = (u_{jk}) = (B_j(A_k)) \),

where the \( A_k \in S \) are chosen such that \( U \) is nonsingular, then the resulting \( a_j^*(x) \) is a vector of probabilities. However, the resulting \( B_j^* \) are now not necessarily probability measures.

The relations \( P = AB \) and \( C = BA \) for finite (or denumerable) Markov chains remind one of the technique of lumping of states, as described in Kemeny and Snell [161]. One might suspect that lumpability is equivalent to having small rank. The following example shows that this is not true.

**Example 1.3.2.** Consider the transition matrices

\[
P_1 = \begin{pmatrix} \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} & 0 \end{pmatrix} \quad \text{and} \quad P_2 = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \end{pmatrix}.
\]

\( P_1 \) has full rank 3, but the first two states can be lumped together, yielding

\[ P^* = \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix} \]

as the transition matrix for the lumped chain. \( P_2 \), on the other hand, has rank 2 and kernel matrix

\[ C = \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}, \]

but no two states can be lumped together in this case.
1.4. Two special cases

1.4.1. Markov operators

For Markov chains of finite rank the theory of Markov operators becomes very simple and rather attractive. Here we exhibit the special form the operator and its adjoint take. For general definitions, proofs and details we refer to FELLER [9].

Let $\overline{U}$ be the set of (equivalence classes of) nonnegative measurable functions on $(S, S)$. For a fixed kernel $P(\lambda|x)$ let $m$ be a measure on $(S, S)$, such that $P(\cdot|x)$ is absolutely continuous with respect to $m$ for every $x$. Now $P$ is considered as an operator on $\overline{U}$, i.e. as a mapping $P: \overline{U} \rightarrow \overline{U}$ defined by

$$ (Pf)(x) := \int_S f(y) P(dy|x) \quad \text{for all } f \in \overline{U}^+ .$$

(1.4.1)

For fixed $x$, let $p(\cdot|x)$ be the Radon-Nikodym derivative of $P$. Then the adjoint operator of $Pf$, denoted by $P^*$, can be defined by

$$ (P^*f)(y) = \int_S f(x)p(y|x)m(dx) \quad \text{for all } f \in \overline{U}^+ .$$

(1.4.2)

Now if $P(\cdot|x) = \sum a(y) \delta(y)$ is the kernel of a Markov chain of rank $r$, then for $m$ we may take the measure defined by (1.1.22). We obtain

$$ (Pf)(x) = \int_S f(y) \sum a(y) \delta(dy) = \sum a(y) \delta(f) ,$$

(1.4.3)

with $\delta(f) := \int_S f(y) \delta(dy)$, i.e. $Pf$ is a linear combination of the $a_j$.

For the adjoint operator we find, putting $b_j := db_j/dm$,

$$ (P^*f)(x) = \int_S f(x) \sum a(y) b(y)m(dx) = a\{f\} b(y) ,$$

(1.4.4)

with $a\{f\} := \int_S f(x) a(y)m(dx)$, i.e. $P^*f$ is a linear combination of the $b_j$. 

1.4.2. Morgenstern's bivariate distribution function

As a simple example of a two-dimensional distribution function MORGENSENM [24] considered

\[
P_\rho(x, y) = G_1(x)G_2(y)\frac{1 + \rho(1 - G_1(x))(1 - G_2(y))}{1 - \rho^2}, \quad |\rho| < 1,
\]

having \(G_1(x), G_2(y)\) as marginal distribution functions, see also GUMBEL [13]. We note that \(\rho\) is the corresponding coefficient of correlation see the end of this section.

We assume that all distributions are absolutely continuous with respect to Lebesgue measure. The bivariate density is then given by

\[
f_\rho(x, y) = g_1(x)g_2(y)(1 + \rho(1 - G_1(x))(1 - G_2(y)) - 1)\frac{1}{1 - \rho^2}.
\]

It follows that

\[
f_\rho(y|x) = g_2(y)(1 + \rho(1 - G_1(x))(1 - G_2(y)) - 1)\frac{1}{1 - \rho^2},
\]

and hence the transition distribution function is

\[
P_{\rho}(y|x) = G_2(y) + \rho(2G_1(x) - 1)(G_2(y) - G_2(y)^2 - G_2(y)) = (1 - \rho(1 - G_1(x))(1 - G_2(y)) + \rho(1 - G_1(x))(1 - G_2(y))^2.
\]

Unless \(\rho = 0\), \(P_{\rho}(y|x)\) is a transition kernel of rank 2 for all \(\rho\) with \(|\rho| < 1\). An easy computation shows that, with

\[
a_1 := \int_{\mathbb{R}} G_1(y)G_2(dy), \quad a_2 := \int_{\mathbb{R}} G_1(y)G_2^2(dy),
\]

the kernel matrix \(C_\rho\) is given by

\[
C_\rho = \begin{bmatrix}
1 + 2\rho a_1 & 2\rho a_2 - \rho a_1 \\
1 + 2\rho a_2 & 2a_2 - \rho a_2
\end{bmatrix} = \frac{1}{1 + 2\rho (1 - a_1 - a_2)} \begin{bmatrix}
1 + 2\rho a_1 & \rho a_1 - 2\rho a_2 \\
1 + 2\rho a_2 & \rho a_2 - 2\rho a_1
\end{bmatrix} + \frac{1}{1 + 2\rho (1 - a_1 - a_2)} \begin{bmatrix}
1 - 2\rho a_1 & 2\rho a_1 - \rho \\
2\rho a_2 - 1 - \rho & 1 + 2\rho a_2
\end{bmatrix}
\]

for all \(a > 1\).
If \( G_1 = G_2 = G \), then \( a_1 = 1/2 \) and \( a_2 = 2/3 \); we find

\[
C_\rho^n = \begin{bmatrix} 1 & 0 \\
1 & 0 \end{bmatrix} + \begin{bmatrix} \frac{\beta}{2} \\
-1 \end{bmatrix}
\]

and

\[
p^{(n)}_\rho (y|x) = G(y) + \left( \frac{\beta}{2} \right)^n \rho (2G(x) - 1)(G^2(y) - G(y)).
\]

Define

\[
u := \int xG(dx), \quad \delta := \int (1 - G(x))G(x)dx.
\]

Then

\[
\int xG^2(dx) = \int x^2G(dx) + \int (1 - G(x))dx = \int \{1 - G(x)\}G(x) + G(x)\}dx + \int (1 + G(x))(1 - G(x))dx = \nu + \delta.
\]

Using (1.2.6) and (1.4.11) we find, putting \( \psi(x) := 2G(x) - 1 \),

\[
\int xG(dx) \psi^{(n)}(dy|x) = \int xG(dx) \psi^{(n)}(dy|x) = \int \int xG(dx) \psi^{(n)}(dy|x) = \int \int xG(dx)(1 - \psi(x), \psi(x))\rho^{n-1} \psi^{(n)}(dy|x) = \int \int xG(dx)(1 - \psi(x), \psi(x))\rho^{n-1} \psi^{(n)}(dy|x) = \int \int xG(dx)(1 - \psi(x), \psi(x))\rho^{n-1} \psi^{(n)}(dy|x) = \int \int xG(dx) + \mu \beta \rho^{n-1} \psi^{(n)}(dy|x) = \nu \beta + \frac{\beta}{2} \nu ^2,
\]
It follows that

\[(1.4.16) \quad \text{cov}(X_0, X_1) = 3B^2 \left( \frac{1}{2} \right)^n, \]

in agreement with theorem 3 of LAI [20], and it follows that the correlation function \( \rho(X_0, X_1) = \rho/3 \) (we return to this example in chapter 4).

1.5. Examples

We give some simple explicit examples of kernels of finite rank. Some of the examples will be referred to in the following chapters.

**Example 1.5.1.** The chain with state space \( S = \{0, 1\} \subset \mathbb{R} \) and kernel

\[(1.5.1) \quad P(y|x) = (1 + x)y - xy^2 \quad \text{for } x, y \in \{0, 1\}, \]

is of rank \( \varepsilon = 2 \). The representation is standard and the kernel matrix is

\[(1.5.2) \quad C = \begin{bmatrix} \frac{1}{2} & -\frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} \end{bmatrix}. \]

Another representation, also standard, for the same kernel is

\[(1.5.3) \quad P(y|x) = (1 - x)y + x(2y - y^2), \]

and now the kernel matrix is a transition matrix:

\[(1.5.4) \quad C'' = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix} = TC^{-1}, \]

where \( T = \begin{bmatrix} 1 & 0 \\ 2 & -1 \end{bmatrix}. \)

**Example 1.5.2.** Let the infinite transition matrix \( P = (p_{jk})_{j,k=1}^{\infty} \) be given by

\[(1.5.5) \quad p_{jk} = \frac{1}{j} z^{-j} = \frac{j-1}{j} 5^k, \]

i.e.,
Each row of $P$ is a linear combination of $(\frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \frac{1}{16}, \ldots)$ and $(1, 0, 0, 0, \ldots)$, hence $P$ is of rank 2. With

\begin{align*}
B &= \begin{bmatrix}
\frac{1}{2} & \frac{1}{4} & \frac{1}{8} & \frac{1}{16} & \ldots \\
1 & 0 & 0 & 0 & \ldots
\end{bmatrix}, \\
S^T &= \begin{bmatrix}
1 & \frac{1}{2} & \frac{1}{3} & \frac{1}{4} & \ldots \\
0 & \frac{1}{2} & \frac{2}{3} & \frac{3}{4} & \ldots
\end{bmatrix},
\end{align*}

the relation $P = AB$ holds, and the kernel matrix $C = BA$ is standard:

\begin{align*}
C &= \begin{bmatrix}
\log 2 & 1 - \log 2 \\
1 & 0
\end{bmatrix}.
\end{align*}

**EXAMPLE 1.5.3.** The transition matrix

\begin{align*}
P &= \begin{bmatrix}
\frac{1}{2} & \frac{1}{4} & 0 \\
0 & \frac{1}{2} & \frac{1}{4} \\
0 & 0 & \frac{1}{2}
\end{bmatrix} = \begin{bmatrix}
1 & 0 & 0 \\
0 & 1 & 0 \\
0 & 0 & 1
\end{bmatrix} \begin{bmatrix}
\frac{1}{2} & \frac{1}{4} & 0 \\
0 & \frac{1}{2} & \frac{1}{4} \\
0 & 0 & \frac{1}{2}
\end{bmatrix},
\end{align*}

has rank 3, and a standard kernel matrix is

\begin{align*}
C &= \begin{bmatrix}
\frac{1}{2} & \frac{1}{4} & 0 \\
0 & \frac{1}{2} & \frac{1}{4} \\
\frac{1}{2} & 0 & -\frac{1}{2}
\end{bmatrix}.
\end{align*}

As we shall see in the next chapter no standard kernel matrix with non-negative elements exists in this case.
EXAMPLE 1.5.4. Let the distribution functions $b_1, \ldots, b_5$ be given by their respective densities (which are "0 elsewhere"):

$$b_1(y) = e^{1-y} \quad \text{for } y \geq 1,$$
$$b_2(y) = 2e^{2-2y} \quad \text{for } y \geq 1,$$
$$b_3(y) = 2y \quad \text{for } y \leq 0,$$
$$b_4(y) = 8y^2 \quad \text{for } y \leq 0,$$
$$b_5(y) = 1 \quad \text{for } 0 < y < 1.$$

Let the function $a_j$ be given by

$$a_1(x) = \begin{cases} 
0 & \text{for } x < 0 \\
|x| & \text{for } 0 \leq x < 1 \\
e^{-x} & \text{for } x \geq 1 
\end{cases}$$
$$a_2(x) = \begin{cases} 
0 & \text{for } x < 1 \\
1-e^{-x} & \text{for } x \geq 1 
\end{cases}$$
$$a_3(x) = \begin{cases} 
0 & \text{for } x < 0 \\
e^x & \text{for } 0 \leq x < 1 \\
1-\frac{1}{x} & \text{for } x \geq 1 
\end{cases}$$
$$a_4(x) = \begin{cases} 
1-e^{-x} & \text{for } x < 0 \\
0 & \text{for } x \geq 0 
\end{cases}$$
$$a_5(x) = \begin{cases} 
1 & \text{for } 0 \leq x \leq 1 \\
0 & \text{otherwise.} 
\end{cases}$$

These functions define a chain of rank 5 in standard representation. The kernel matrix is
\[
\begin{pmatrix}
\frac{1}{2} e^{-1} & \frac{1}{2} e^{-1} & 0 & 0 & 0 \\
\frac{2}{3} e^{-1} & \frac{1}{3} e^{-1} & 0 & 0 & 0 \\
0 & 0 & \frac{1}{2} & \frac{1}{2} & 0 \\
0 & 0 & \frac{2}{3} & \frac{1}{3} & 0 \\
\frac{1}{4} & 0 & \frac{1}{4} & 0 & \frac{1}{2}
\end{pmatrix}
\]

We here have a chain with two "absorbing" sets: \((-\infty,0]\) and \(\{1,\infty]\), and \(C\) is a "reducible" transition matrix. We shall discuss this situation in the next chapter.

\[\text{EXAMPLE 1.5.5.} \text{ The finite Markov chain given by the transition matrix}
\]

\[
\begin{pmatrix}
\frac{1}{2} & 0 & \frac{1}{2} \\
0 & 0 & 1 \\
0 & 1 & 0 \\
0 & 0 & 1 \\
1 & 0 & 0 \\
1 & 0 & 0
\end{pmatrix}
\]

is of rank 3 with kernel matrix

\[
\begin{pmatrix}
0 & \frac{1}{2} & \frac{1}{2} \\
1 & 0 & 0 \\
0 & 0 & 0
\end{pmatrix}
\]

This chain is cyclic with period 2.
CHAPTER 2. EIGENVALUES

In chapter 1 we have seen that to each Markov chain of finite rank there corresponds an equivalence class of similar kernel matrices. Since all matrices in such an equivalence class have the same eigenvalues, it is quite natural to study the eigenvalues of kernel matrices. It turns out that in many respects kernel matrices behave like the well-studied transition matrices of finite Markov chains. For the eigenvalue structure of the latter we refer to Iosifescu [16] and to Fritz, Huppert and Willems [12]. The eigenvectors of the kernel matrix have natural analogues in the eigenfunctions of the kernel $P(\cdot|\cdot)$; their close relationship plays an essential role in our considerations.

General results are given in the first section. In section 2 we consider the eigenvalue $\lambda_0 = 1$, which leads to a classification of (sets of) states, similar to that for finite Markov chains. The eigenvalues of modulus 1, but unequal to 1, determine "cyclically moving subsets"; this is the topic of section 3. Section 4 is devoted to a further comparison of transition and kernel matrices, and to the location of eigenvalues in the complex plane. In the final section some examples are given.

We use the following notation and terminology: A vector $y \in \mathbb{R}^r$ with $y \neq 0$, satisfying $Cy = \lambda y$ for some $\lambda \in \mathbb{C}$, will be called a right $\lambda$-eigenvector of $C$; the word right is sometimes omitted. A vector $y \in \mathbb{R}^r$ with $y \neq 0$, satisfying $y^tC = \lambda y^t$ for some $\lambda \in \mathbb{C}$, will be called a left $\lambda$-eigenvector of $C$.
Throughout the chapter we assume a standard representation. Integrals are over the whole state space $S$, unless otherwise indicated.
2.1. General results

**Definition 2.1.1.** A function \( v \) is called a \( \lambda \)-eigenfunction of the kernel \( P(\lambda|\cdot) \) if for all \( \lambda \neq 0 \)

\[
(2.1.1) \quad \int v(y) P(dy|x) = \lambda v(x) \quad \text{for all} \quad x \in S \quad \lambda \in \mathbb{C}.
\]

**Lemma 2.1.2.** If \( v \) is a \( \lambda \)-eigenfunction of the kernel \( P(\lambda|\cdot) = T(\lambda)|\cdot \) and \( \lambda \neq 0 \), then

\[
(2.1.2) \quad v(x) = \sum_{j=1}^{F} v_j \alpha_j(x) = \sum_{j=1}^{F} \alpha_j(x) v_j
\]

with \( v \) a right \( \lambda \)-eigenfunction of the kernel matrix \( C \).

**Proof.** We have

\[
(2.1.3) \quad v(x) = \frac{1}{\lambda} \int v(y) P(dy|x) = \sum_{j=1}^{F} \alpha_j(x) \int \frac{v(y)}{\lambda} B_j(dy),
\]

hence \( v(x) \) is of the form (2.1.2). Substituting (2.1.2) in (2.1.1) we find

\[
(2.1.4) \quad \sum_{j=1}^{F} \alpha_j(x) v_j = \lambda v(x),
\]

and by the linear independence of the \( \alpha_j(x) \) it follows that \( v \) is a \( \lambda \)-eigenfunction of \( C \).

**Lemma 2.1.3.** The eigenfunctions \( v_1(x), \ldots, v_F(x) \) are linearly independent if and only if the corresponding eigenvectors \( v_1, \ldots, v_F \) are linearly independent.

**Proof.** Follows immediately from the linear independence of the \( \alpha_j(x) \).

**Theorem 2.1.4.** Every kernel matrix \( C \) satisfies

\[
(2.1.5) \quad C1 = 1 ;
\]

\[
(2.1.6) \quad Cv = \lambda v, \quad v \neq 0 \Rightarrow |\lambda| \leq 1 .
\]

**Proof.** Relation (2.1.5) only restates that \( C \) has unit row sums. Now let \( v(x) \) be given by (2.1.2), then \( v(x) \) satisfies (2.1.1). As the \( \alpha_j(x) \) are bounded (see Lemma 1.1.5), \( v(x) \) is bounded. We have
\[(2.1.7) \quad |\lambda| \cdot |v(x)| \leq \int |v(y)| P(dy \mid x) \leq \sup_{y \in S} |v(y)| \quad \text{for all } x \in S. \]

since \( \sup_{y \in S} |v(y)| \neq 0 \), it follows that \(|\lambda| \leq 1 \).

An eigenvalue \( \lambda \) is called nondegenerate, if its algebraic multiplicity equals its geometric multiplicity, i.e. if the multiplicity of \( \lambda \) as a root of the characteristic polynomial of the matrix is equal to the number of linearly independent \( \lambda \)-eigenvectors; the geometric multiplicity cannot exceed the algebraic multiplicity, but it may be smaller.

For the eigenvalues of a kernel matrix we have

**Lemma 2.1.3.** If \(|\lambda| = 1\), then \( \lambda \) is nondegenerate. \( \Box \)

This result is a simple generalization of the corresponding result for finite Markov chains; it is proved in the appendix as lemma A.2.3. By lemma 1.2.6 and theorem 2.1.4 the conditions of lemma A.2.3 are satisfied here.

**Remark 2.1.6.** Although, as we shall see in section 4, a kernel matrix \( C \) is not necessarily similar to a nonnegative matrix, it has a Perron-Frobenius eigenvalue as we have shown by theorem 2.1.4. This is related to the fact that a convex cone is left invariant by \( C \) (compare KINGMAN [19]). However, in our case there is not necessarily a nonnegative left \( 1 \)-eigenvector, as is seen from the example

\[(2.1.8) \quad \begin{bmatrix} \frac{2}{3} & 0 & \frac{1}{6} & \frac{1}{3} \\ \frac{2}{3} & 0 & \frac{1}{3} & 0 \\ \frac{2}{3} & 0 & 0 & \frac{1}{3} \\ \frac{2}{3} & 0 & 0 & \frac{1}{3} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 2 & -1 \\ 2 & -1 \end{bmatrix} \begin{bmatrix} \frac{2}{3} & 0 & \frac{1}{6} & \frac{1}{6} \\ \frac{2}{3} & 0 & \frac{1}{3} & 0 \\ \frac{2}{3} & 0 & \frac{1}{3} & 0 \end{bmatrix}, \]

where

\[(2.1.9) \quad C = \begin{bmatrix} \frac{2}{3} & -\frac{1}{3} \\ \frac{1}{3} & -\frac{1}{3} \\ \frac{1}{3} & -\frac{1}{3} \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} \frac{4}{3} & -\frac{1}{3} \\ \frac{1}{3} & -\frac{1}{3} \end{bmatrix} \]

has left \( 1 \)-eigenvector \((4/3, -1/3)\). \( \Box \)
In this and in the following chapters we often use the so-called spectral decomposition of a matrix. Details will be given in the appendix, theorem A.1.4. The matrix theory result that we exploit is

**LEMMA 2.1.7.** Let $M$ be a complex $r \times r$ matrix with distinct eigenvalues $\lambda_0, \lambda_1, \ldots, \lambda_d$. Then

\[
(2.1.10) \quad M = \sum_{k=0}^{d} (\lambda_k E_k + N_k),
\]

where $E_k, N_k, E_k - E_k, N_k - N_k, E_k - N_k, E_k^2 - E_k$ and $N_k^2 = 0$ for all $n \geq r$. Furthermore, if $\lambda_k$ is a nondegenerate eigenvalue, then $U_k = 0$.

If $\lambda_k$ has algebraic multiplicity 1 and if $u_k$ and $v_k$ are left and right $\lambda_k$-eigenvectors of $M$ with $u_k^T v_k = 1$, then

\[
(2.1.11) \quad E_k = u_k^T u_k. \quad \Box
\]

We apply lemma 2.1.7 to a kernel matrix $C$. Let $1 = \lambda_0, \lambda_1, \ldots, \lambda_d$. denote the distinct eigenvalues of modulus 1. From (2.1.10) we find, using lemma 2.1.5,

\[
(2.1.12) \quad C = \sum_{k=0}^{d} \lambda_k E_k + \sum_{k=0}^{d} (\lambda_k E_k + N_k),
\]

and more generally for all $n \in \mathbb{N}$

\[
(2.1.13) \quad C = \sum_{k=0}^{d} \lambda_k E_k + \sum_{k=0}^{d} (\lambda_k E_k + \sum_{k=0}^{n} \lambda_k^m E_k^{n-m}).
\]

We note that the inner sum has less that $r$ terms, since $N_k^r = 0$. Hence only the first sum matters if $n$ is large.

**LEMMA 2.1.8.** There exists an $\varepsilon$, $0 < \varepsilon < 1$, such that

\[
(2.1.14) \quad C = \sum_{k=0}^{d} \lambda_k E_k + \varepsilon(C^n) \quad (n = 0).
\]

**PROOF.** Define $\rho = \max_{\varepsilon > 0} |\lambda_k|$, then $0 < \rho < 1$ and from (2.1.13) we obtain

\[
(2.1.15) \quad C = \sum_{k=0}^{d} \lambda_k E_k + O(\varepsilon^n \rho^n) \quad (n = 0).
\]
Now for any $\epsilon$ with $0 < \epsilon < 1$ relation (2.14) is satisfied. See also
lemma A.1.13 in the appendix.

In the next section we need the following results:

**Lemma 2.1.9.** Let $v$ be any vector such that

\begin{equation}
\lim_{n \to \infty} c^n v = 0.
\end{equation}

Then, with the $v$ of (2.14),

\begin{equation}
|c^n v| \leq c^n \quad \text{for all sufficiently large } n.
\end{equation}

**Proof.** In the appendix, lemma A.1.13, it is proved that (2.16) yields

\begin{equation}
(c^n v)(n) = c^n (v^n) \quad (n \to \infty).
\end{equation}

Relation (2.17) is a trivial consequence of (2.18).

In section 2.3 it will be proved, independently of lemma 2.1.9, that
all eigenvalues of modulus 1 are (integer) roots of unity. Using that result
we can give a simple straightforward proof of lemma 2.1.9, based on (2.14):
since \( c_{n=0}^{d-1} E^n T = 0 \) is a periodic function of $n$, it must be identically zero,
as the remainder term of $c^n v$ tends to 0 as $n \to \infty$.

**Lemma 2.1.10.** If for some subset $T$ of $S$

\begin{equation}
\lim_{n \to \infty} f^{(n)}(T|x) = 0 \quad \text{for all } x \in S,
\end{equation}

then

\begin{equation}
\lim_{n \to \infty} c^n \mathcal{B}(T) = 0.
\end{equation}

**Proof.** Put $a_j^{(n)} := (c^n \mathcal{B}(T))_j$. By lemma 1.2.6 the set $\{a_j^{(n)} \mid j \in \{1, \ldots, r\},$
$n \in \mathbb{N}\}$ is bounded. Let $a_j$ be an arbitrary limit point of the sequence
$a_1^{(1)}, a_2^{(2)}, a_3^{(3)}, \ldots$. Then there are limit points $a_2, \ldots, a_r$ and a subsequence
\( \{n_k^{(n)}\} \) such that \( \lim_{n \to \infty} a_j^{(n_k^{(n)})} = a_j \) for all $j \in \{1, \ldots, r\}$. From (2.1.19) it
follows that

\begin{equation}
\sum_{j=1}^r a_j a_j(x) = 0 \quad \text{for all } x \in S.
\end{equation}
Hence $a_1 = \ldots = a_r = 0$, by the linear independence of the $a_j(x)$. Thus we have found that $\lim_{n \to +\infty} \beta_j^{(n)} = 0$ for $j = 1$, and clearly this also holds for all other $j$. But that is equivalent to (2.1.20).

As a consequence of lemma 2.1.10, if $T$ satisfies (2.1.19), then the vector $v = \beta(T)$ satisfies the condition of lemma 2.1.9, so that $\beta(T)$ satisfies (2.1.17). As a special consequence we get

**Lemma 2.1.11.** If the set $T$ satisfies relation (2.1.19), then

$$\sum_{n=0}^{\infty} |\beta(T)|^n < \infty.$$  

2.2. The eigenvalue $1$; classification of states

As in the case of finite Markov chains the eigenvalues $\lambda$ with $|\lambda| = 1$ are of special interest. They correspond directly to the structure of the chain.

In this section we consider the eigenvalue $\lambda = 1$. We shall see that its multiplicity, which by theorem 2.1.4 is at least 1, determines the number of absorbing subsets. Most of the results about $1$-eigenfunctions that we need here, hold more generally for $\lambda$-eigenfunctions with $|\lambda| = 1$. They are formulated as such for later use, particularly in the next section.

All subsets of the state space $S$ occurring in the sequel of this chapter are supposed to be measurable.

We start by giving some definitions.

**Definition 2.2.1.** If $A$ is a nonempty subset of $S$ such that

$$(2.2.1) \quad P(A|x) = 1 \quad \text{for all } x \in A,$$

then $A$ is said to be closed. If a closed set cannot be split into two disjoint closed sets, it is called absorbing.

If a subset $T$ of $S$ of positive $\mu$-measure (see (1.1.22)) satisfies

$$(2.2.2) \quad \lim_{n \to +\infty} P^{(n)}(T|x) = 0 \quad \text{for all } x \in T,$$

then $T$ is called transient.

A closed set containing no transient subsets is called recurrent.
In the literature there is no uniform terminology for classifying subsets of the state space of a Markov process. The terms introduced above are based on those used by FELLER [6] for finite Markov chains. In DOOB [5] transient sets are defined by (2.2.2) with \( x \in \mathcal{T} \) replaced by \( x \in \mathcal{S} \); see lemmas 2.2.2 and 2.2.3 below.

First entrance probabilities will be very useful in our considerations. For \( n \in \mathbb{N} \) and arbitrary \( A \subset \mathcal{S} \) let

\[
\xi^{(n)}(A|x) := \mathbb{P}(X_0 \notin A, X_1, \ldots, X_n = x | X_0 = x);
\]

\( \xi^{(n)}(A|x) \) is determined recursively by

\[
\xi^{(1)}(A|x) = \mathbb{P}(A|x),
\]

\[
\xi^{(n)}(A|x) = \int_{S \setminus A} \xi^{(n-1)}(A|y) \mathbb{P}(dy|x) \quad \text{for } n \geq 2.
\]

We further define

\[
\xi(A|x) := \sum_{n=1}^{\infty} \xi^{(n)}(A|x),
\]

the probability that the set \( A \) is ever entered after a positive number of transitions.

For formal reasons we introduce

\[
\mathbb{I}(A|x) := \lim_{n \to \infty} \xi^{(n)}(A|x),
\]

the indicator of \( A \).

We first derive some results about transient sets.

**Lemma 2.2.2.** A subset of positive \( \mathbb{P} \)-measure of a transient set is transient. \( \blacksquare \)

**Proof.** Follows immediately from the definition (2.2.2).

In the following lemma we apply the Fatou-Lebesgue lemma: if \( g, f_1, f_2, \ldots \) are measurable functions, such that \( |f_n| \leq g \) for all \( n \) and \( \int g \, d\mu < \infty \), then \( \limsup_{n \to \infty} \int f_n \, d\mu \leq \int \limsup_{n \to \infty} f_n \, d\mu \). A proof can be found in LEVE, [23].
LEMMA 2.2.3. If $T$ is a transient set, then

$$\lim_{n \to \infty} p^{(n)}(T|x) = 0 \quad \text{for all } x \in S .$$

PROOF. For $x \in S \setminus T$ we have, taking $p^{(j)}(T|y) = 0$ if $j < 0$,

$$p^{(n)}(T|x) = \sum_{k=0}^{n} \int_T f^{(k)}(dy|x)p^{(n-k)}(T|y) =$$

$$= \sum_{k=0}^{n} \int_T f^{(k)}(dy|x)p^{(n-k)}(T|y) \leq$$

$$\leq \sum_{k=0}^{n} f^{(k)}(T|x) = f(T|x) \leq 1 .$$

Applying the Fatou-Lebesgue lemma twice (with majorants provided by (2.2.8)) we obtain

$$\limsup_{n \to \infty} p^{(n)}(T|x) = \limsup_{n \to \infty} \sum_{k=0}^{\infty} \int_T f^{(k)}(dy|x)p^{(n-k)}(T|y) \leq$$

$$\leq \sum_{k=0}^{\infty} \limsup_{n \to \infty} \int_T f^{(k)}(dy|x)p^{(n-k)}(T|y) \leq$$

$$\leq \sum_{k=0}^{\infty} \int_T f^{(k)}(dy|x) \limsup_{n \to \infty} p^{(n-k)}(T|y) = 0 .$$

COROLLARY 2.2.4. The union of a finite number of transient sets is transient.

PROOF. Let $T_1, \ldots, T_k$ be transient sets and put $T := \bigcup_{j=1}^{k} T_j$. Then

$$\limsup_{n \to \infty} p^{(n)}(T|x) \leq \sum_{j=1}^{k} \limsup_{n \to \infty} p^{(n)}(T_j|x) = 0 .$$
Lemma 2.2.5. The union of a countable number of transient sets is transient. \[\square\]

Proof. Let \( T_1, T_2, \ldots \) be transient sets. Define

\[
T := \bigcup_{j=1}^{\infty} T_j, \quad T_k := \bigcup_{j=1}^{k} T_j \quad \text{for } k = 1, 2, \ldots
\]

By corollary 2.2.4 the sets \( T_k \) are transient for all \( k \geq 1 \). Choose an arbitrary \( \varepsilon > 0 \). Since \( \lim_{n \to \infty} T_k^n = T \) and since the \( \mathbb{P}_j \) are probability measures, there is a \( k_0 = k_0(\varepsilon) \) such that

\[
\|\mathbb{P}(T) - \mathbb{P}(T_{k_0})\| = \|\mathbb{P}(T \setminus T_{k_0})\| < \varepsilon.
\]

As \( e_n \) is bounded for \( n \in \mathbb{N} \), there exists a constant \( c \) such that \( |c_{jk}| \leq c \) for all \( j, k \in \{1, \ldots, r\} \) and all \( n \in \mathbb{N} \). Applying lemma 2.1.10 we find

\[
\limsup_{n \to \infty} \mathbb{P}(T^n | x) = \limsup_{n \to \infty} \frac{T(x) \cdot e_n \cdot e_x}{\mathbb{P}(T) \cdot e_x} \leq \limsup_{n \to \infty} \left[ \frac{T_a(x) \cdot e_n \cdot e_x}{\mathbb{P}(T) \cdot e_x} + \frac{T_a(x) \cdot e_n \cdot e_x}{\mathbb{P}(T \setminus T_{k_0}) \cdot e_x} \right] \leq 0 + \varepsilon c.
\]

As \( \varepsilon \) was chosen arbitrarily, it follows that \( T \) satisfies the relation (2.2.7). \[\square\]

Remark 2.2.6. The lemmas 2.2.2 and 2.2.3, as well as corollary 2.2.4, are valid for general Markov chains; in their proofs no use is made of the special form of transition kernels of finite rank. Lemma 2.2.5, however, does not hold for general Markov chains. Take, for example, the denumerable Markov chain on \( S = \{0, 1, 2, \ldots\} \) with transition matrix \( P \) given by

\[
P_{jk} = \begin{cases} 
1 & \text{for } k = j + 1 \\
0 & \text{otherwise}.
\end{cases}
\]

Here the sets \( T_j := \{0, 1, \ldots, j\} \) are transient for all \( j > 1 \), but \( \bigcup_{j=1}^{\infty} T_j = \{0, 1, 2, \ldots\} \) is not transient. The chain determined by (2.2.14) is not of finite rank.

Most of the results in the sequel of this section are not valid for general Markov chains. \[\square\]
THEOREM 2.2.7. A set $T$ is transient if and only if
\begin{equation}
\sum_{n=0}^{\infty} p^{(n)}(T|x) < \infty \quad \text{for all } x \in S,
\end{equation}
in which case the sum is even bounded.

**Proof.** The implication $(2.2.15) \Rightarrow (2.2.2)$ is trivial. Now suppose $T$ is transient. Lemma 2.2.3 applies and hence by Lemma 2.1.11 we get
\begin{equation}
\sum_{n=0}^{\infty} p^{(n)}(T|x) = 1_A(x) + \sum_{n=1}^{\infty} e^{-n} E(T) \leq \infty.
\end{equation}
As the $a_j(x)$ are bounded, the above expression is bounded.

For any set $T$ the sum $\sum_{n=0}^{\infty} p^{(n)}(T|x)$ is the expected number of visits to $T$, the process starting at the point $x$. If $T$ is transient this number is finite, even bounded, for all $x \in S$.

Here again a similarity appears between Markov chains of finite rank and finite Markov chains: there are no null-states, i.e. the situation
\begin{equation}
\sum_{n=0}^{\infty} p^{(n)}(T|x) = \infty, \quad \lim_{n \to \infty} p^{(n)}(T|x) = 0
\end{equation}
does not occur.

We remark that $(2.2.2)$ and $(2.2.15)$ are satisfied for any $\mu$-nullset $T$, $p^{(n)}(T|x)$ being equal to 0 for all $n \geq 1$. For this reason in definition 2.2.1 it is required that a transient set has positive $\mu$-measure; in general transient sets are taken modulo $\mu$-nullsets.

We now turn to closed sets. By definition, once the process has entered a closed set it remains there with probability 1. A closed set may contain two or more absorbing sets; the process actually stays within the absorbing set that it has entered. An absorbing set cannot have arbitrarily small $\mu$-measure as is shown in

**Lemma 2.2.8.** If $A$ is a closed set, and if $\sup_{x \in A} a_j(x) < 1$, then $1 > 0$ and
\begin{equation}
\mu(A) \geq \frac{1}{L_0}.
\end{equation}
PROOF. By definition

\[ P(A \mid x) = \prod_{j=1}^{r} B_j(A) \quad \text{for all } x \in A, \]  

hence

\[ \left( \bigcup_{j=1}^{r} B_j(A) \right) = L \tau \pi(A). \]

**COROLLARY 2.2.9.** The number of disjoint absorbing sets is finite. \[ \Box \]

Many properties of closed and transient sets follow from eigenfunction considerations. The next lemma gives an important property of \( \lambda \)-eigenfunctions with \( |\lambda| = 1 \).

**LEMMA 2.2.10.** Let \( \Lambda \) be a nonempty subset of \( S \) and \( v(x) \) a bounded function on \( \Lambda \) such that, for all \( \lambda \) with \( |\lambda| = 1 \),

\[ \lambda v(x) = \int_{x}^{s} v(y) P(dy|x) \quad \text{for all } x \in \Lambda. \]

Then there is at least one \( x_0 \in \Lambda \) such that

\[ \sup_{x \in \Lambda} |v(x)| = |v(x_0)|. \]

Furthermore, if this supremum is positive then

\[ \Lambda_0 := \{ x_0 \in \Lambda \mid |v(x_0)| = \sup_{x \in \Lambda} |v(x)| \} \]

is a closed set. \[ \Box \]

**PROOF.** If \( v(x) = 0 \) on \( \Lambda \), the assertions are trivial. Otherwise, \( v(x) \) can be normalized such that \( \sup_{x \in \Lambda} |v(x)| = 1 \). Assume this has been done and let \( (u_n)_{n=1}^{\infty} \),

with \( u_n \in \Lambda \) for all \( n \), be a sequence with \( \lim_{n \to \infty} |v(u_n)| = 1 \).

As the \( a_j(x) \) are bounded, there is a subsequence \( (x_k)_{k=1}^{\infty} = (u_j)_{j=1}^{\infty} \),

such that \( \lim_{k \to \infty} a_j(x_k) = \alpha_j \) exists for all \( j \). One easily verifies that

\[ F := \bigcup_{j=1}^{r} \bigcup_{k \in \mathbb{N}} a_j B_k \]

is a probability measure (cf. (1.3.8)). Now from
\[(2.2.25) \quad |v(x)| \leq \int_{A} |v(y)| P(dy|x) = \frac{T(x)}{T(x)} \int_{A} |v(y)| B(dy) \quad \text{for all } x \in A\]

one deduces, by substituting \(x = x_{\lambda}\) and letting \(\lambda \to \infty\),

\[(2.2.26) \quad 1 \leq \int_{A} |v(y)| P(dy) .\]

It follows that for \(A_{0} = \{ x \in A \mid |v(x)| = 1 \}\) we must have \(P(A_{0}) = 1\). Hence \(A_{0}\) is nonempty and the first assertion of the lemma is proved. Furthermore we obtain from (2.2.21) for \(x \in A_{0}\)

\[(2.2.27) \quad 1 \leq P(A_{0}|x) + \int_{A \setminus A_{0}} |v(y)| P(dy|x) .\]

If the integral does not vanish, it is less than \(P(A \setminus A_{0} \mid x)\), but that contradicts (2.2.27). Hence \(P(A_{0}|x) = 1\) for all \(x \in A_{0}\), in other words: \(A_{0}\) is a closed set.

The conditions of lemma 2.2.10 are satisfied if \(v(x)\) is a \(\lambda\)-eigenfunction with \(|\lambda| = 1\) and \(A\) is a closed set. The absolute value of every such eigenfunction therefore takes on a maximum on each closed set. Note that, trivially, \(S\) itself is a closed set.

If for a fixed closed set \(A\) the function \(v(x)\) satisfies (2.2.21) with a positive supremum over \(A\) and is normed in such a way that

\[(2.2.28) \quad \max_{x \in A} |v(x)| = 1 = v(x_{0}) \quad \text{for some } x_{0} \in A ,\]

we say that \(v(x)\) is normed on \(A\).

In the special case \(\lambda = 1\) another closed subset of \(A\) can be found, which is contained in the subset \(A_{0}\) given by (2.2.23). This is proved in

**Lemma 2.2.11.** If \(\lambda = 1\), if \(A\) and \(v(x)\) satisfy (2.2.21), and if \(v(x)\) is normed on \(A\), then

\[(2.2.29) \quad A'_{0} := \{ x \in A \mid v(x) = 1 \}\]

is a closed set.
**Proof.** As $v(x)$ is normed on $A$, the set $A_0'$ is nonempty. Rather than (2.2.25) we now have

$$ v(x) = T_x(x) \int_A v(y) \mathbb{E}(dy) \quad \text{for all } x \in A, $$

from which we obtain

$$ 1 = \int_A v(y) \mathbb{P}(dy). $$

It follows that $A_0'$ is closed, similarly to the last part of the proof of the previous lemma.

**Lemma 2.2.12.** If $A$ is an absorbing set then every 1-eigenfunction is constant on $A$. \(\square\)

**Proof.** Suppose $v(x)$ is a 1-eigenfunction, nonconstant and normed on $A$. Then $\tilde{v}(x) := i \cdot v(x)$ is also a 1-eigenfunction, nonconstant on $A$. Applying Lemma 2.2.11 we find two disjoint closed subsets $A_0'$ and $A_0''$ of $A$, both defined by (2.2.29) for $v(x)$ and $\tilde{v}(x)$, respectively. This contradicts the assumption that $A$ is absorbing. \(\blacksquare\)

The following theorem is the main result of this section.

**Theorem 2.2.13.** The number of disjoint absorbing sets equals the multiplicity of the eigenvalue $\lambda_0 = 1$ of $C$. \(\square\)

**Proof.** Let $k_0$ be the multiplicity of $\lambda_0 = 1$. Suppose $A_1, \ldots, A_k$ are disjoint absorbing sets, and put $A_0 := S \setminus \bigcup_{k=1}^k A_k$. With $f(A|x)$ as defined by (2.2.5) we put

$$ v_k(x) := f(A_k|x) \quad \text{for all } x \in S \quad (k = 1, \ldots, k). $$

Clearly,

$$ v_k(x) = \begin{cases} 1 & \text{for all } x \in A_k, \\ 0 & \text{for all } x \in A_j, \quad j \neq k, \end{cases} $$

hence the $v_k(x)$ are linearly independent. Furthermore,

$$ \int v_k(y) \mathbb{P}(dy|x) = v_k(x) \quad \text{for all } x \in S \quad (k = 1, \ldots, k). $$
this is trivial for $x \in \bigcup_{k=1}^{n} A_k$, while for $x \in A_0$ we have

$$
(2.2.35) \quad \int \psi_k(y) \mathbb{P}(dy|x) = \mathbb{P}(A_k | x) + \int_{S \setminus A_k} \sum_{n=1}^{\infty} \mathbb{P}(A_k | y) \mathbb{P}(dy|x) = \\
= \int \psi_k(y) \mathbb{P}(dy|x) + \sum_{n=1}^{\infty} \int \mathbb{P}(A_k | y) \mathbb{P}(dy|x) = f(A_k | x).
$$

Thus we have constructed $k$ linearly independent $1$-eigenfunctions, hence $k_0 \geq k$ by lemma 2.1.3.

Now suppose that $k_0 > 1$. Then by lemma 2.1.5 there exists a 1-eigenfunction $w(x)$, independent of $\psi_1(x), \ldots, \psi_{k_0}(x)$. By lemma 2.2.12 the function $w(x)$ is constant on each set $A_k$, say $w(x) = c_k$ for $x \in A_k$. Consider the function

$$
(2.2.36) \quad \psi(x) := w(x) - \frac{1}{k_0} \sum_{k=1}^{k_0} \psi_k(x) \quad \text{for } x \in S.
$$

This is a 1-eigenfunction, vanishing outside $A_0$. From lemma 2.2.10 it follows that $A_0$ contains a least one closed set, hence at least one absorbing set $A_{k+1}$, which is obviously disjoint from $A_1, \ldots, A_k$. This completes the proof.

When we have determined the disjoint absorbing sets $A_1, \ldots, A_{k_0}$, what can be said about the remainder set? The following theorem answers this question.

**Theorem 2.2.15.** Let $A_1, \ldots, A_{k_0}$ be disjoint absorbing sets, where $k_0$ is the multiplicity of $\lambda_0 = 1$. Then the set $S_0 := S \setminus \bigcup_{k=1}^{k_0} A_k$ is transient or not-null.

**Proof.** Put $A := \bigcup_{k=1}^{k_0} A_k$. The function $v(x) = 1$ is a 1-eigenfunction, which by theorem 2.2.13 is linearly dependent of the $\psi_k(x)$, defined by (2.2.32). From (2.2.33) it follows that we simply have

$$
(2.2.37) \quad 1 = v(x) = \sum_{k=1}^{k_0} \psi_k(x).
$$

Hence

$$
(2.2.38) \quad f(A|x) = 1 \quad \text{for all } x \in S,
$$
so that for all \( N \geq 1 \) and \( x \in S \)

\[
(2.2.39) \quad f(A|x) = \sum_{n=1}^{N} f^{(n)}(A|x) + \sum_{n=N+1}^{\infty} \int_{T_0} f^{(n-N)}(A|y)p^{(N)}(dy|x) = \\
= \sum_{n=1}^{\infty} f^{(n)}(A|x) + p^{(N)}(T_0|x) .
\]

This yields

\[
(2.2.40) \quad \lim_{N \to \infty} p^{(N)}(T_0|x) = 0 \quad \text{for all } x \in S ,
\]

hence by definition \( T_0 \) is transient if \( m(T_0) > 0 \). \[\blacklozenge\]

We have shown that \( S \) consists of \( T_0 \) disjoint absorbing sets \( A_1, \ldots, A_\tau \) and a transient remainder set \( T_0 \). Each of the absorbing sets may, however, still contain a transient part. If one removes a transient set from a closed set, the remaining set is not necessarily closed, but we have

**Lemma 2.2.12.** If \( A \) is a closed set and \( T = A \) is transient, then \( A \setminus T \) contains a closed set \( A_0 \), and \( A_0 \) can be chosen in such a way that \( A \setminus A_0 \) is transient. \( \Box \)

**Proof.** Define

\[
A_0 := \left\{ x \in A \mid \sum_{n=0}^{\infty} p^{(n)}(T|x) = 0 \right\} ,
\]

(2.2.41)

\[
T_0 := \left\{ x \in A \mid \sum_{n=0}^{\infty} p^{(n)}(T|x) > 0 \right\} .
\]

For all \( x \in A \) and fixed \( N \geq 1 \) we have

\[
(2.2.42) \quad \sum_{n=0}^{N-1} p^{(n)}(T|x) = \sum_{n=0}^{N-1} p^{(n)}(T|x) + \sum_{n=N}^{\infty} \int_{T_0} f^{(n-N)}(T|y)p^{(N)}(dy|x) = \\
= \sum_{n=0}^{N-1} p^{(n)}(T|x) + \sum_{n=0}^{\infty} \int_{T_0} f^{(n)}(T|y)p^{(N)}(dy|x) .
\]

If we take \( N = 1 \) (and \( x \in A_0 \)) this yields
\[(2.2.43) \quad P(A_0|x) = 1 - P(T_0'|x) = 1 \quad \text{for all} \quad x \in A_0.\]

By letting \( N = n \) in \((2.2.42)\) and noting that \( \sum_{n=0}^{\infty} p^{(N)}(T|x) = n \) by theorem 2.2.6, it follows that
\[(2.2.44) \quad \lim_{\mathcal{N}} P(A_0|x) = 1 - \lim_{\mathcal{N}} P(T_0'|x) = 1 \quad \text{for all} \quad x \in A.\]

From \((2.2.44)\) we see that \( A_0 \) is nonempty, hence by \((2.2.43)\) \( A_0 \) is closed, whereas \( T_0 = A \setminus A_0 \) is transient (with \( m(T_0) > m(T) > 0 \)). \[\blacksquare\]

\textbf{THEOREM 2.2.16.} Let \( A \) be a closed set. There exist sets \( T \subset A \) and \( R := A \setminus T \), such that \( T \) is transient (or \( \mu \)-null) and \( R \) is recurrent. These sets are uniquely determined modulo \( \mu \)-nullsets.

\textbf{PROOF.} Put
\[(2.2.45) \quad T := \{U \subset A \mid U \text{ is transient or } \mu \text{-null}\},\]
and
\[(2.2.46) \quad t := \sup_{U \in T} \mu(U) .\]

The case \( t = 0 \) is trivial. Suppose \( t > 0 \) and let \( T_1, T_2, \ldots \) be a sequence of sets in \( T \), such that \( \lim_{n \to \infty} \mu(T_n) = t \). By lemma 2.2.5 the set \( T := \bigcup_{n=1}^{\infty} T_n \) is transient, while \( \mu(T) = t \).

Now apply lemma 2.2.15: \( R = A \setminus T \) contains a closed set \( A_0 \) with \( A \setminus A_0 \) transient. Since \( T \) is a transient set of maximal \( \mu \)-measure, we have \( A_0 = R \) (modulo an \( \mu \)-nullset) and \( R \) is recurrent.

As for uniqueness: suppose \( T^* \) and \( R^* \) also satisfy the conditions. Then \( T \cup T^* \) is an element of \( T \), and by the construction of \( T \) it follows that \( T^* \setminus T \) is an \( \mu \)-nullset. On the other hand, also \( T \setminus T^* \) is an \( \mu \)-nullset, otherwise \( R^* \) would contain a transient set. Hence \( T = T^* \) and \( R = R^* \) modulo an \( \mu \)-nullset. \[\blacksquare\]

Applying theorem 2.2.16 to the closed set \( S \), we see that the state space of a Markov chain of finite rank consists of a recurrent part \( R \) and a transient part \( T \), unique modulo a nullset.

In general, the state space of any Markov process can be split into a so-called conservative and a dissipative part. This is known as the Hopf-decomposition, see e.g. FISCHER [9].
By definition a set $A$ is conservative if $m(A) > 0$ and if for every subset $A' \subseteq A$ of positive $m$-measure one has $f(A'|x) = 1$ for $m$-almost all $x \in A'$.

In order to complete the Hopf-decomposition for a Markov chain of finite rank, we prove

**Lemma 2.2.17.** A is recurrent if and only if $A$ is conservative and closed. \( \Box \)

**Proof.** Suppose $A$ is recurrent, and suppose there is a subset $A' \subseteq A$ with $m(A') > 0$ such that $f(A'|x) < 1$ for all $x \in A'$. For a sufficiently small $\delta > 0$ the set

\[
T := \{x \mid f(A'|x) \leq 1 - \delta\} \subset A'
\]

has positive $m$-measure. From

\[
p^{(n)}(T|x) = \sum_{k=1}^{m-1} \int T f^{(k)}(y|x) (dy|x) \quad \text{for all } x \in A
\]

we obtain for $x \in T$, putting $p(x) := \limsup_{n \to \infty} p^{(n)}(T|x)$, cf. (2.2.8) and (2.2.9),

\[
p(x) \leq \int \sup_{y \in T} p(y)f(dy|x) = \sup_{y \in T} p(y)f(T|x) \leq (1 - \delta) \sup_{y \in T} p(y)
\]

It follows that $p(x) = 0$ for $x \in T$, so that $T$ is transient by definition. This contradicts the assumption that $A$ is recurrent.

Conversely suppose $A$ is conservative and closed, and let $A' \subseteq A$ be any subset of positive $m$-measure. For $m$-almost all $x \in A'$ we have

\[
m \sum_{n=0} \frac{p^{(n)}(A'|x)}{n+1} = 1 + \sum_{n=1}^{m} \frac{p^{(n)}(A'|x)}{n+1} = \]

\[
= 1 + \sum_{n=1}^{m} \left( \frac{f^{(n)}(A'|x)}{n+1} + \sum_{k=1}^{n-1} f^{(n-k)}(A'|y) f^{(k)}(dy|x) \right) =
\]

\[
= 1 + \sum_{k=1}^{m} \left( \frac{f^{(n-k)}(A'|y) f^{(k)}(dy|x) \right) =
\]
\[ 1 + 1 + \sum_{k=1}^{\infty} \mathbb{P}^k(A^t|x) = \]
\[ = 1 + \sum_{n=0}^{\infty} p^{(n)}(A^t|x), \]
from which it follows that \( \sum_{n=0}^{\infty} p^{(n)}(A^t|x) = \) and hence, by theorem 2.2.6, that \( A' \) is not transient. We conclude that \( A \) is a closed set, containing no transient subsets, and therefore is recurrent by definition. \( \blacksquare \)

2.3. The eigenvalues of modulus 1; periodicity

If \( A \) is a closed set (not necessarily absorbing) we can construct a new Markov chain of finite rank by restricting the original chain to \( A \). On \( A \) the \( a_{ij}(x) \) may be linearly dependent, in which case the rank of the new chain is less than the rank of the original chain; if so, we re-define the \( a_{ij}(x) \) and the \( b_j(y) \) in order to obtain a standard representation for the new chain.

How about the eigenvalues and eigenfunctions? Suppose \( \psi(x) \) is a \( \lambda \)-eigenfunction of the original chain, and suppose that its restriction \( \psi(x) \) to the closed set \( A \) is not identically 0. Then \( \psi(x) \) is a \( \lambda \)-eigenfunction of the new chain, i.e.

\[(2.3.1) \quad \lambda \psi(x) = \int_A \psi(y) \mathbb{P}(dy|x) \text{ for all } x \in A, \quad \psi(x) \not= 0 \text{ on } A.
\]

In this way eigenvalues are inherited by the new chain. On the other hand, if \( \psi \) satisfies (2.3.1) with \(|\lambda| = 1\), it can be extended to a function on \( S \), by

\[(2.3.2) \quad \varphi(x) := \sum_{n=1}^{\infty} \lambda^{-n} \int_A \psi(y) p^{(n)}(dy|x) \text{ for } x \in S \setminus A.
\]

The extended function \( \varphi \) is a \( \lambda \)-eigenfunction of the original kernel, as one easily verifies. This means that no new eigenvalues of modulus 1 are introduced by restricting the Markov chain to \( A \). Trivially, the multiplicity of \( \lambda \) in the new chain cannot exceed the multiplicity of \( \lambda \) in the original chain. In particular, the extension of \( \psi \) given by (2.3.2) is not necessarily the only one (unless \( S \setminus A \) is transient).
In this section we want to investigate the structure of absorbing sets. For this, the eigenvalues of modulus 1 are essential. On account of the argument given above we may restrict the chain to the absorbing set under consideration. Therefore we shall assume that we are dealing with a chain with only one absorbing set, i.e. a chain with a single eigenvalue \( \lambda_0 = 1 \).

**Definition 2.3.1.** A Markov chain of finite rank with only one absorbing set is called **indecomposable**.

An indecomposable chain with no transient sets is said to be **irreducible**.

The same terms are used for the corresponding kernel matrices.

In an indecomposable chain we may restrict the process to the recurrent part (see theorem 2.2.16). On account of lemma 2.2.10 this does not change the situation regarding the eigenvalues of modulus 1. So we may even assume that we are dealing with an irreducible chain.

Now let \( \lambda \) be an eigenvalue of an irreducible chain, with \( |\lambda| = 1 \) but \( \lambda \neq 1 \), and let \( v(x) \) be a normal \( \lambda \)-eigenfunction (i.e. \( \max_{x \in S} |v(x)| = 1 \) and \( v(x_0) = 1 \) for some \( x_0 \in S \)). Define

\[
S_0 := \{ x \in S \mid v(x) = 1 \},
\]

\[
S_1 := \{ x \in S \mid v(x) = \lambda \}.
\]

For \( x \in S_0 \) we have (cf. the proof of lemma 2.2.10)

\[
(2.3.4) \quad \int_S \left( 1 - \frac{v(x)}{\lambda} \right) P(dy|x) = 1 - v(x) = 0.
\]

Hence, since \( |v(y)/\lambda| = |v(y)| \leq 1 \),

\[
(2.3.5) \quad P(S_1|x) = 1 \quad \text{for all } x \in S_0.
\]

Defining generally

\[
(2.3.6) \quad S_k := \{ x \in S \mid v(x) = \lambda^k \} \quad (k = 0, 1, \ldots)
\]

we find

\[
(2.3.7) \quad P(S_{k+1}|x) = 1 \quad \text{for all } x \in S_k \quad (k = 0, 1, \ldots).
\]
The process moves with probability 1 from $S_0$ to $S_1$, from $S_1$ to $S_2$, and so on. As $S_0$ is nonempty, none of the $S_k$ are empty, and by definition $S_k \cap S_{k+1} = \emptyset$ if and only if $\lambda^k \neq \lambda^{k+1}$.

The set $\tilde{S} := \bigcup_{k=0}^\infty S_k$ is closed: if $x \in \tilde{S}$ then $x \in S_k$ for some $k \geq 0$ and hence $P(\tilde{S}|x) = P(S_{k+1}|x) = 1$. Since the chain is assumed to be irreducible, we have $\tilde{S} = S$ modulo an m-nullset.

Define the functions $\nu_k(x)$ on $\tilde{S}$ by

$$
(2.3.8) \quad \nu_k(x) := \nu_k^d(x) = \lambda^d x_k \quad \text{for} \quad x \in S_k \quad (k = 0,1,\ldots).
$$

Then for $x \in S_k$ ($k = 0,1,\ldots$) we get

$$
(2.3.9) \quad \int_{\tilde{S}} \nu_k(y) \, d\mu(y|x) = \int_{S_k} \nu_k(y) \, d\mu(y|x) = \lambda^{d(k+1)} x_k^{(k+1)} = \lambda^d \nu_k(x),
$$

so $\nu_k(x)$ is a $\lambda^d$-eigenfunction. It follows that $\lambda^d$ can take at most $r$ different values, hence there is an integer $d$, with $1 \leq d \leq r$, such that $\lambda^d = 1$. We have proved

**Lemma 2.3.2.** Every eigenvalue $\lambda$ of modulus 1 is a root of unity of order $d \leq r$, and $\lambda^1, \lambda^2, \ldots, \lambda^{d-1}$ are also eigenvalues of the chain. \(\Box\)

Corresponding to an eigenvalue $\lambda = e^{2\pi i \theta}$ there is a set of so-called "cyclically moving subsets" $S_0 \rightarrow S_1 \rightarrow \cdots \rightarrow S_{d-1} + S_0 = S_0$, for which (2.3.7) holds; the term is due to DOOB [5]. This is a two-way correspondence: if $S_0 \rightarrow S_1 \rightarrow \cdots \rightarrow S_{d-1} = S_0$ are $d$ disjoint cyclically moving subsets, then $\nu(x)$ defined by

$$
(2.3.10) \quad \nu(x) = e^{2\pi i \frac{x}{d}} \quad \text{for} \quad x \in S_k \quad (k = 0,1,\ldots,d-1),
$$

is an $\lambda^d$-eigenfunction, hence $e^{2\pi i \theta}$ is an eigenvalue.

**Definition 2.3.3.** An indecomposable Markov chain of finite rank is called cyclic if there are $d \geq 2$ disjoint cyclically moving subsets $S_0 \rightarrow S_1 \rightarrow \cdots \rightarrow S_{d-1} + S_0$. If $d$ is the largest integer with this property, then $d$ is called the period of the chain. \(\Box\)
Now consider an irreducible cyclic chain with period $d$. As we have seen, all $d$-th order roots of unity

$$
(2.3.11) \quad \lambda_k := e^{\frac{2\pi i k}{d}} \quad (t = 0, 1, \ldots, d-1)
$$

are eigenvalues of the kernel, with eigenfunctions given by (2.3.8). Suppose $\tilde{V}(x)$ is a $\tilde{\lambda}$-eigenfunction for some $\tilde{\lambda}$ with $|\tilde{\lambda}| = 1$, $\tilde{\lambda} \neq 1$. Let

$$
\tilde{S}_0 \rightarrow \tilde{S}_1 \rightarrow \ldots \rightarrow \tilde{S}_{d-1} \rightarrow \tilde{S}_0
$$

be the corresponding cyclically moving subsets with $\tilde{V}(x) = \tilde{V}(\tilde{S}_j)$ on $\tilde{S}_j$; indices of $\tilde{S}$ are taken modulo $\tilde{d}$, indices of $S$ modulo $d$. The two cycles cannot be disjoint (or there would be two disjoint closed sets), hence $\tilde{S}_0 \cap \tilde{S}_{d} \neq \emptyset$ for some $\tilde{j}_0$ and $j_0$. No generality is lost by assuming $\tilde{j}_0 = j_0 = 0$, as this amounts to a renorming of the eigenfunctions. We have

$$
(2.3.12) \quad P(S_1|\tilde{x}) = P(\tilde{S}_1|x) = 1 \quad \text{for all } x \in S_0 \cap \tilde{S}_0,
$$

hence

$$
(2.3.13) \quad P(S_1 \cap \tilde{S}_1|x) = 1 \quad \text{for all } x \in S_0 \cap \tilde{S}_0.
$$

Proceeding in this way we find a cycle $S_0 \cap \tilde{S}_0 + S_1 \cap \tilde{S}_1 + \ldots + S_{d-1} \cap \tilde{S}_{d-1} = S_0 \cap \tilde{S}_0$. The length $n$ of this cycle is at least $d$, but it cannot exceed the period $d$ of the chain. Therefore $n = d$ and $n$ is divisible by $d$. Furthermore

$$
(2.3.14) \quad S_k \cap \tilde{S}_k = S_k \quad \text{(modulo null sets)} \quad \text{for } k = 0, 1, \ldots, d-1.
$$

It follows that $\tilde{\lambda}$ is one of the eigenvalues $\lambda_k$ given in (2.3.11), and that there are no other eigenvalues of modulus 1. Moreover, the $\tilde{\lambda}$-eigenfunction $\tilde{V}(x)$ is given by

$$
(2.3.15) \quad \tilde{V}(x) = \tilde{\lambda}^k \quad \text{for } x \in S_k \quad (k = 0, 1, \ldots, d-1),
$$

which means that there is (up to a norming factor) only one $\tilde{\lambda}$-eigenfunction, i.e., $\tilde{\lambda}$ has multiplicity 1.

We summarize the obtained results in
THEOREM 2.3.4. In an indecomposable chain of finite rank all eigenvalues of modulus 1 are given by (2.3.11) for some integer \( d \geq 1 \), and they all have multiplicity 1.

If \( d \geq 2 \) the chain is cyclic with period \( d \).

The \( \lambda_k \)-eigenfunction \( v_k(x) \), if properly normed, is determined by

\[
(2.3.16) \quad v_k(x) = \lambda_k^x = e^{2\pi i k x/d} \quad \text{for } x \in S_k \quad (k = 0, 1, \ldots, d-1).
\]

where \( S_0 \supseteq S_1 \supseteq \ldots \supseteq S_{d-1} \supseteq S_d = S_0 \) are \( d \)-cyclically moving subsets. \( \Box \)

2.4. The location of eigenvalues

The class of chains of finite rank contains the finite Markov chains as we have seen in Section 1.1. In Section 1.2 some relations between (standard) kernel matrices and transition matrices were derived. Now let \( K_r \) denote the set of all complex numbers that can occur as an eigenvalue of a matrix in \( K(\mathbb{R}) \), and let \( M_{1r} \) denote the set of all complex numbers that can occur as an eigenvalue of a \( r \times r \) transition matrix. The problem of determining \( M_{1r} \) (or, slightly more general, the set of all eigenvalues of \( n \times n \) nonnegative matrices) was posed by Kolmogorov, partly solved by Dmitriev and Dynkin [4], and finally completely solved by Karpelevitch [17].

The set \( M_{1r} \) is a closed, star-shaped subset of the unit disk of the complex plane. The only points of \( M_{1r} \) on the unit circle are the points \( \exp(2\pi ik/r) \) with \( 0 \leq k < r \), and \( M_{1r} \) is symmetrical with respect to the real axis. The boundary of \( M_{1r} \) consists of two straight lines between 1 and \( \exp(2\pi i/r) \) and (for \( r \geq 4 \)) of polynomial arcs elsewhere; see Karpelevitch [17] for the explicit formulas. In Figure 1 on page 50 a sketch of \( M_3 \), \( M_4 \), and \( M_5 \) is given.

The basic observation in Dmitriev and Dynkin [4] and in Karpelevitch [17] is that \( \lambda \in M_{1r} \) if and only if there exists a convex \( k \)-angular polygon \( 0 < k < r \) which is mapped into itself when multiplied by \( \lambda \). Kernel matrices are in general not nonnegative, and it is not clear how this argument can be extended to get information about \( K_r \). The problem of determining \( K_r \) remains open, but some (mainly numerical) results can be given.
For an arbitrary matrix $M$ let $D_M(\lambda) = \det(M - \lambda I)$ be the characteristic polynomial of $M$. From theorem 1.2.3 it follows that $D_C(\lambda) = D_{C^*}(\lambda)$ if $C$ and $C^*$ are two minimal kernel matrices of the same kernel. We have

**Lemma 2.4.1.** If $P$ is an $s \times s$ transition matrix of rank $r$ and $C$ is a corresponding (minimal) kernel matrix, then

$$D_P(\lambda) = (-\lambda)^{s-r} D_C(\lambda) .$$

**Proof.** Let $P = AB$ and $C = BA$, cf. (1.1.20) and (1.2.7). Define the $s \times s$ matrices $A_0$ and $B_0$ by

$$A_0 = \begin{bmatrix} \lambda & \mathbb{R}_1 \end{bmatrix}, \quad B_0 = \begin{bmatrix} \mathbb{R}_0 \end{bmatrix} .$$

where $\mathbb{R}_1$ is any $s \times (s-r)$ matrix that gives $A_0$ rank $s$ and $\mathbb{R}_0$ is the $(s-r) \times s$ matrix with all entries equal to 0. From $P = AB = A_0B_0$ we find

$$D_P(\lambda) = \det(P - \lambda I_s) = \det(A_0B_0 - \lambda I_s) =$$

$$= \det(A_0B_0A_0 - \lambda I_s)A_0^{-1} = \det(B_0A_0 - \lambda I_s) =$$

$$= \det(BA - \lambda I_s) \det(-\lambda I_{s-r}) =$$

$$= \det(C - \lambda I_s)(-\lambda)^{s-r} = (-\lambda)^{s-r} D_C(\lambda) .$$

Lemma 2.4.1 asserts that $C$ inherits all the nonzero eigenvalues of $P$. It does not follow that $C$ has only nonzero eigenvalues; in fact, $C$ can even have an eigenvalue 0 of multiplicity $r - 1$. As an example consider

$$P = \begin{bmatrix} 0 & 1 & 0 & \ldots & 0 \\ 0 & 0 & 1 & \ldots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \ldots & 1 \\ 0 & 0 & 0 & \ldots & 1 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & \ldots & 0 \\ 1 & 0 & 0 & \ldots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \ldots & 1 \\ 0 & 0 & 0 & \ldots & 1 \end{bmatrix} \begin{bmatrix} 0 & 1 & 0 & \ldots & 0 \\ 0 & 0 & 1 & \ldots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \ldots & 1 \\ 0 & 0 & 0 & \ldots & 1 \end{bmatrix} ,$$

where $P$ is an $(r+1) \times (r+1)$ transition matrix of rank $r$. The corresponding kernel matrix is an $r \times r$ matrix of rank $r - 1$, and $D_C(\lambda) = (-\lambda)^{r-1}(1 - 1)$. 
**Lemma 2.6.2.** The set $K_\tau$ is a subset of the unit disk of the complex plane, star-shaped and containing $H_\tau$. \[\square\]

**Proof.** From theorem 2.1.4 it follows that $K_\tau$ is a subset of the unit disk and from lemma 1.2.5 that $M_\tau \subset K_\tau$. Now suppose $\lambda_0 \in K_\tau$. Then there is a standard kernel matrix $C \in K(\tau)$, corresponding to the kernel $P(A|x) = T^r(x)B_1(A)$, such that $\lambda_0$ is an eigenvalue of $C$. For $0 \leq \alpha \leq 1$ define

$$P_\alpha(A|x) := \alpha P(A|x) + (1 - \alpha)B_1(A).$$

This is a transition kernel with standard kernel matrix

$$C_\alpha = \alpha C + (1 - \alpha)I_n \in K(\tau).$$

We have

$$B_{C_\alpha}(\lambda) = \begin{vmatrix}
\alpha c_1^{(1)} - \lambda & \alpha c_{12} & \ldots & \alpha c_{1r} \\
\alpha c_{21}^{(1)} - \lambda & \alpha c_{22} - \lambda & \ldots & \alpha c_{2r} \\
\vdots & \vdots & \ddots & \vdots \\
\alpha c_{r1}^{(1)} - \lambda & \alpha c_{r2} - \lambda & \ldots & \alpha c_{rr} - \lambda \\
\end{vmatrix} = \frac{1}{(1 - \lambda)^r} B_{C_0}(\lambda),$$

from which it follows that $C_\alpha$ has an eigenvalue $\alpha_0$. This means that $K_\tau$ is star-shaped, which completes the proof. \[\square\]

For the case $r = 2$ we have

**Theorem 2.4.3.**

$$K_2 = H_2 = [-1, 1] \subset \mathbb{R}. \[\square\]$$
PROOF. By theorem 1.3.1 each minimal $2 \times 2$ kernel matrix is similar to a $2 \times 2$ transition matrix. From this (2.4.8) follows immediately.

We turn to the case $r = 3$. Every $C \in K(3)$ has an eigenvalue $1$; the other two are either real or complex conjugates. Only the second situation is of interest here. So let $1, x + iy$ and $x - iy$ ($x, y \in \mathbb{R}, y > 0$) be the three eigenvalues of a kernel matrix $C \in K(3)$. The sum of the eigenvalues is nonnegative by lemma 1.2.9, hence $x > -i$. This leads to

**Lemma 2.4.4.** If $\lambda \in K_3$ and $\lambda$ be nonreal, then

\[(2.4.9) \quad \text{Re } \lambda^n = -\lambda^n \quad \text{for all } n = 1, 2, \ldots .\]

**Proof.** If $\lambda$ is an eigenvalue of the matrix $C \in K(3)$, then $\lambda^n$ is an eigenvalue of the matrix $C^n \in K(3)$ (cf. lemma 1.2.7). The other eigenvalues of $C^n$ are $1$ and $\text{Re } \lambda^n$, the complex conjugate of $\lambda^n$. Relation (2.4.9) now follows from lemma 1.2.9.

**Corollary 2.4.5.** Suppose that the eigenvalues of the matrix $C \in K(3)$ are $1, x + iy$ and $x - iy$ ($x, y \in \mathbb{R}$). Then

\[(2.4.10) \quad y^2 < x^2 + \frac{1}{4}.\]

**Proof.** By lemma 2.4.4 we have ($n = 2$)

\[(2.4.11) \quad \text{Re } (x + iy)^2 = x^2 - y^2 > -\lambda,\]

which immediately yields (2.4.10).

Lemma 2.4.4 enables us to exclude a part of the unit disk when searching the area of $K_3$, see figure 2 on the next page. $K_3$ is contained in the shaded area, except the interval $[-1, -i)$ of the real line. On the other hand, it contains the triangle, since $M_3 \in K_3$.

![Figure 1](image-url)
In order to obtain a better inner bound for $K_3$ we consider the following problem: find a $4 \times 4$ transition matrix $P$ of rank 3, such that for a specified $t \in \mathbb{R}$ the eigenvalues of $P$ are $1$, $0$, $x + i tx$ and $x - i tx$, and $x$ is maximal (for $t > 0$) or minimal (for $t < 0$). Since for every $n \times n$ matrix $M = (m_{j k})$ with eigenvalues $\lambda_1, \ldots, \lambda_n$ the relations

$$\sum_{j=1}^{n} \lambda_j = \sum_{j=1}^{n} m_{j j},$$

(2.4.12)

and

$$\sum_{j \neq k} \lambda_j \lambda_k = \sum_{j \neq k} |m_{j k}|$$

(2.4.13)

hold for the first two invariants of the matrix (see e.g. FRANKLIN [10], sec. 4.3), the above problem can be formulated as follows ($t > 0$ fixed):

maximize

$$x = \frac{1}{4} (p_{11} + p_{22} + p_{33} + p_{44} - 1)$$

(2.4.14)

under the conditions
\( p_{jk} \geq 0 \) for all \( j, k \in \{1, 2, 3, 4\} \),

\[ \sum_{k=1}^{4} p_{jk} = 1 \] for all \( j \in \{1, 2, 3, 4\} \),

\[ 2x + x^2 + 2x^2 = \sum_{j=1}^{4} \sum_{k=1}^{4} p_{jj} p_{jk} \]

The maximisation has been carried out by computer for \( t = \tan k\pi/30 \), \( k = 1, \ldots, 14 \). The result is

<table>
<thead>
<tr>
<th>( k )</th>
<th>( t )</th>
<th>( x )</th>
<th>( tx )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.103</td>
<td>0.855</td>
<td>0.090</td>
</tr>
<tr>
<td>2</td>
<td>0.213</td>
<td>0.797</td>
<td>0.161</td>
</tr>
<tr>
<td>3</td>
<td>0.325</td>
<td>0.684</td>
<td>0.222</td>
</tr>
<tr>
<td>4</td>
<td>0.445</td>
<td>0.627</td>
<td>0.279</td>
</tr>
<tr>
<td>5</td>
<td>0.577</td>
<td>0.580</td>
<td>0.335</td>
</tr>
<tr>
<td>6</td>
<td>0.727</td>
<td>0.540</td>
<td>0.392</td>
</tr>
</tbody>
</table>

\[ (2.4.18) \]

<table>
<thead>
<tr>
<th>( 7 )</th>
<th>0.900</th>
<th>0.509</th>
<th>0.458</th>
</tr>
</thead>
<tbody>
<tr>
<td>8</td>
<td>1.111</td>
<td>0.451</td>
<td>0.501</td>
</tr>
<tr>
<td>9</td>
<td>1.376</td>
<td>0.369</td>
<td>0.508</td>
</tr>
<tr>
<td>10</td>
<td>1.723</td>
<td>0.300</td>
<td>0.520</td>
</tr>
<tr>
<td>11</td>
<td>2.246</td>
<td>0.237</td>
<td>0.532</td>
</tr>
<tr>
<td>12</td>
<td>3.078</td>
<td>0.178</td>
<td>0.549</td>
</tr>
<tr>
<td>13</td>
<td>4.705</td>
<td>0.121</td>
<td>0.567</td>
</tr>
<tr>
<td>14</td>
<td>9.514</td>
<td>0.062</td>
<td>0.588</td>
</tr>
</tbody>
</table>

In figure 3 a sketch of the obtained bound is given.
The computer also gave the matrices \( P \) that correspond to the optimal values given in (2.4.18). For \( k = 1, \ldots, 7 \) we could not discover any specific structure in those optimal matrices. However, for \( k = 8, \ldots, 14 \) such a structure was apparent, from which we found

**Lemma 2.4.6.** The arc in Figure 3 between the points \( e^{2\pi i/3} \) and \( 1+i \) is given by

\[
y^2 = \frac{1}{2}(x-1)^2 + 1 \quad \text{for } -\frac{1}{2} \leq x \leq \frac{1}{2}.
\]

\( \square \)

**Proof.** Let \( P \) have the eigenvalues 0, 1, and \( x \pm iy \). For fixed \( x \), i.e., for fixed \( p_{11} + p_{22} + p_{33} + p_{44} \), we obtain from (2.4.12) and (2.4.13)

\[
y^2 \leq \sum_{j \neq k} p_{jj} p_{kk} - x^2 - 2x \leq 6p^2 - x^2 - 2x,
\]

where \( p := \frac{1}{4}(p_{11} + p_{22} + p_{33} + p_{44}) = \frac{1}{2}(2x + 1) \). This yields

\[
y^2 \leq \frac{1}{4}(x-1)^2 + 1 \quad \text{for all } x.
\]

For \( -\frac{1}{2} \leq x \leq \frac{1}{2} \) the upper bound for \( y^2 \) is attained if we take
(2.4.22) \[ p = \begin{pmatrix} \rho & \theta_1 & \theta_2 & 0 \\ 0 & p & 1-p & 0 \\ 0 & 0 & p & 1-p \\ 1-p & 0 & 0 & p \end{pmatrix}, \]

with \( \theta_1 = p(1-p) + p^4(1-p)^{-2} \) and \( \theta_2 = (1-p)^4 - p^4(1-p)^{-2} \).

**Remark 2.4.7.** The construction at the end of the proof of lemma 2.4.6 fails for \( x > \frac{1}{2} \) (then \( p > \frac{1}{2} \) and \( \theta_2 \) is negative). Besides, no transition matrix with all main diagonal entries greater than \( \frac{1}{2} \) can have an eigenvalue 0 (by Gershgorin's theorem, see e.g. Fritz, Huppert and Willems [12]). Moreover, we know that for \( \frac{1}{2} < x < 1 \) we must have

(2.4.23) \[ y \leq 1 - x. \]

At the end of section 1.2 is was shown by an example that the set \( K(3) \) is not closed under matrix multiplication. The same two matrices \( C_1 \) and \( C_2 \) that were introduced in (1.2.21) may serve as an example to show that \( K(3) \) is not convex. Define for \( 0 \leq \alpha \leq 1 \)

(2.4.24) \[ C_\alpha := \alpha C_1 + (1-\alpha) C_2. \]

The matrix \( C_\alpha \) has eigenvalues \( \lambda_1 = 1 \) and

(2.4.25) \[ \lambda_{2,3} = \frac{1}{2} \alpha + \frac{1}{2} \sqrt{\frac{1}{4} \alpha^2 + 1}. \]

By corollary 2.4.5 we must have

(2.4.26) \[ \frac{1}{2} (2\alpha + 1) \leq \left( \frac{1}{4} - \alpha \right)^2 + \frac{1}{4}, \]

which is true for \( 0 \leq \alpha \leq \frac{1}{2} \) and for \( \alpha = 1 \), but not for \( \frac{1}{2} < \alpha < 1 \).

**Remark 2.4.8.** The inequalities (2.4.20) and (2.4.21) can be generalized as follows. If the transition matrix \( P \) is of size \( n \times n \) and of rank 3, with nonzero eigenvalues 1 and \( x \pm iy \), then we obtain from (2.4.12) and (2.4.13)

(2.4.27) \[ y^2 \leq \sum_{j \neq k} p_{jj} - \frac{1}{n} \left( p_{11} + \ldots + p_{nn} \right) + \frac{1}{n} (2x + 1). \]

where now \( p := \frac{1}{n} \left( p_{11} + \ldots + p_{nn} \right) = \frac{1}{n} (2x + 1) \). Hence we find

(2.4.28) \[ y^2 \leq (1 - \frac{1}{n}) x^2 - \frac{2}{n} x + \frac{1}{2} \left( 1 - \frac{1}{n} \right). \]

For \( n = 1 \) this yields (2.4.10) (cf. theorem 1.2.9).
What are the extreme eigenvalues of kernel matrices $G \in K(3)$ that correspond to $5 \times 5$ transition matrices $P$? This question leads to a second optimization problem, analogous to the first one and also solved by computer. The same extreme values, given in (2.4.18), were obtained. The main programming difficulty in this case is formed by the condition that $P$ has rank 3; prescribing linear dependence of the rows of $P$ introduces new variables that are unbounded.

A third optimization problem that has been solved by computer is the finding of extreme eigenvalues of kernel matrices $G \in K(4)$ that correspond to $5 \times 5$ transition matrices $P$. The obtained result is (cf. (2.4.18)):

<table>
<thead>
<tr>
<th>$k$</th>
<th>$t$</th>
<th>$x$</th>
<th>$tx$</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>0.325</td>
<td>0.778</td>
<td>0.253</td>
</tr>
<tr>
<td>4</td>
<td>0.445</td>
<td>0.725</td>
<td>0.323</td>
</tr>
<tr>
<td>5</td>
<td>0.577</td>
<td>0.674</td>
<td>0.389</td>
</tr>
<tr>
<td>6</td>
<td>0.727</td>
<td>0.618</td>
<td>0.449</td>
</tr>
<tr>
<td>7</td>
<td>0.900</td>
<td>0.537</td>
<td>0.486</td>
</tr>
<tr>
<td>(2.4.29) 7.5</td>
<td>1</td>
<td>0.500</td>
<td>0.500</td>
</tr>
<tr>
<td>8</td>
<td>1.111</td>
<td>0.483</td>
<td>0.536</td>
</tr>
<tr>
<td>9</td>
<td>1.376</td>
<td>0.432</td>
<td>0.595</td>
</tr>
<tr>
<td>10</td>
<td>1.732</td>
<td>0.373</td>
<td>0.650</td>
</tr>
<tr>
<td>11</td>
<td>2.246</td>
<td>0.314</td>
<td>0.705</td>
</tr>
<tr>
<td>12</td>
<td>3.078</td>
<td>0.248</td>
<td>0.763</td>
</tr>
<tr>
<td>13</td>
<td>5.705</td>
<td>0.177</td>
<td>0.833</td>
</tr>
<tr>
<td>14</td>
<td>9.514</td>
<td>0.095</td>
<td>0.904</td>
</tr>
</tbody>
</table>

Figure 4 gives a sketch of these values.
Figure 4.

An optimal matrix $P$ for $k = 11$ (also produced by the computer) is

$$
\begin{pmatrix}
0.25 & 0. & 0.75 & 0. & 0. \\
0.44 & 0.25 & 0. & 0.31 & 0. \\
0. & 0. & 0.25 & 0. & 0.75 \\
0.75 & 0. & 0. & 0.25 & 0. \\
0. & 0.45 & 0. & 0.30 & 0.23
\end{pmatrix},
$$

(2.4.30)

and for $k = 8$ each obtained optimal $P$ is of the form (2.4.30), with a constant main diagonal. The optimal matrices $P$ obtained for $k \leq 7$ have a slightly different structure, such as (for $k = 5$)

$$
\begin{pmatrix}
0.63 & 0. & 0.37 & 0. & 0. \\
0. & 0.63 & 0. & 0.37 & 0. \\
0. & 0. & 0.25 & 0. & 0.75 \\
0.37 & 0. & 0. & 0.63 & 0. \\
0. & 0.56 & 0.19 & 0. & 0.25
\end{pmatrix},
$$

(2.4.31)

with a nonconstant main diagonal.
2.5. Examples

We consider the examples of section 1.5.

EXAMPLE 2.5.1 (1.5.1). The eigenvalues of \( C \in \mathbb{K}(2) \) given by (1.5.4) are \( \lambda_0 = 1 \) and \( \lambda_1 = -1/6 \). The chain is indecomposable, noncyclic, and even irreducible. □

EXAMPLE 2.5.2 (1.5.2). The eigenvalues of the kernel matrix \( C \in \mathbb{K}(2) \) given by (1.5.8) are \( \lambda_0 = 1 \) and \( \lambda_1 = \log 2 - 1 \). The chain is irreducible and noncyclic.

EXAMPLE 2.5.3 (1.5.3). The eigenvalues of the kernel matrix \( C \in \mathbb{K}(3) \) given by (1.5.10) are \( \lambda_0 = 1, \lambda_1 = \frac{1}{2} + \frac{1}{2}i \) and \( \lambda_2 = \frac{1}{2} - \frac{1}{2}i \). The chain is irreducible and noncyclic.

Since \( \lambda_{1,2} \notin \mathbb{R} \), no \( 3 \times 3 \) transition matrix is similar to \( C \), hence there is no standard kernel matrix with nonnegative elements for the chain. □

EXAMPLE 2.5.4 (1.5.4). The eigenvalues of the kernel matrix \( C \in \mathbb{K}(3) \) given by (1.5.13) are \( \lambda_0 = 1, \lambda_1 = 1, \lambda_2 = \frac{1}{2}, \lambda_3 = -1/6e^{-1}, \lambda_4 = -1/6 \). The chain has two absorbing sets, \((-\infty, 0] \) and \([1, +\infty) \). The recurrent part \( R \) of the state space \( \mathbb{R} \) is \((-\infty, 0] \cup [1, +\infty) \), the transient part \( T \) is \((0, 1) \). □

EXAMPLE 2.5.5 (1.5.5). The eigenvalues of the kernel matrix \( C \in \mathbb{K}(3) \) given by (1.5.15) are \( \lambda_0 = 1, \lambda_1 = -1 \) and \( \lambda_2 = 0 \). The chain is irreducible and cyclic with period \( d = 2 \). If the \( S = \{0, 1, 2, 3\} \) is the state space for the chain with transition matrix \( T \) given by (1.5.14), then \( S_0 = \{0, 2\} \) and \( S_1 = \{1, 3\} \) are the two cyclically moving subsets. □
CHAPTER 3. LIMIT DISTRIBUTIONS

In this chapter we study the stochastic behaviour for large $n$ of some functions of $(X_1, \ldots, X_n)$, where $X_1, X_2, \ldots$ is a Markov chain of finite rank. In section 3.1 we consider the asymptotic distribution of $X_n$ itself in the indecomposable case. The sections 3.2 and 3.3 are devoted to the central limit theorem, i.e. to the distribution of $X_1 + \cdots + X_n$, and in section 3.4 we study the distribution of $\max(X_1, \ldots, X_n)$. In section 3.5 we briefly consider renewal theory for Markov processes of finite rank.

Many of our results can be found, sometimes with greater generality, elsewhere. The novelty is mainly in the simplicity of the proofs in our special situation. All proofs are based on simple matrix methods, in particular on the spectral decomposition (see lemma 2.1.7 and the appendix). These methods are applied to the matrix $C$ itself, and to matrices closely related to $C$. Although our results imply those for finite Markov chains, our proofs are often simpler than known proofs for that case.

3.1. The invariant distribution

In this section we consider the distribution of $X_n$ for large $n$. The proof of theorem 3.1.2 as given below is briefly indicated in KUNNENBURG AND STEUDEL [30]. We start by giving a definition.

**DEFINITION 3.1.1.** A probability measure $\mu$ satisfying

$$P(A) = \int_A \mu(dx)$$

for all $A \in \mathcal{S}$ is called invariant with respect to $P(\cdot|x)$. □

In the special case of a finite Markov chain with transition matrix $P$ it is well-known that a unique invariant distribution $\pi = \pi_1^T \pi_2 \ldots \pi_n^T$ exists if $P$ is irreducible; $P$ then has a single eigenvalue 1 and $\pi$ is a left 1-eigenvector of $P$ (see e.g. FELLER [62]).
For kernel matrices we have the following analogue.

**Theorem 3.1.2.** If the kernel matrix \( C \) is indecomposable (see definition 2.3.1.), then

\[
(3.1.2) \quad \frac{1}{n} \sum_{k=1}^{n} p^{(k)}(A|x) = C(A) + O\left(\frac{1}{n}\right) \quad (n \to \infty)
\]

for all \( x \in S \) and \( A \in S \). The measure \( G \) is given by

\[
(3.1.3) \quad G = \sum_{j=1}^{K} \gamma_j E_j = \sum_{j=1}^{K} \gamma_j E_j
\]

where \( \gamma \) is the left \( 1 \)-eigenvector of \( G \), normed such that \( \gamma^T 1 = 1 \), and \( G \) is the unique invariant probability measure. \( \square \)

**Proof.** Since the chain is indecomposable, the eigenvalues of modulus 1 are given by (2.3.11) for some fixed \( d \geq 1 \), and they all have multiplicity 1. From the spectral decomposition of \( C \), in particular from lemma 2.1.8, we find

\[
(3.1.4) \quad \frac{1}{n} \sum_{k=1}^{n} c^k = \frac{1}{n} \sum_{k=1}^{n} \left( \frac{d-1}{d} \gamma_k E_k + O\left(\frac{1}{n}\right) \right) =
\]

\[
= \sum_{k=0}^{d-1} \left( \frac{1}{n} \sum_{k=1}^{n} \gamma_k^k \right) E_k + O\left(\frac{1}{n}\right) =
\]

\[
= E_0 + \sum_{k=0}^{d-1} \left( \frac{1}{n} \sum_{k=1}^{n} \gamma_k^k \right) E_k + O\left(\frac{1}{n}\right) \quad (n \to \infty).
\]

For \( t = 1, \ldots, d-1 \) we have, since \( \lambda_k \) is a root of unity,

\[
(3.1.5) \quad \frac{1}{n} \sum_{k=1}^{n} \lambda_k^k = O\left(\frac{1}{n}\right) \quad (n \to \infty),
\]

and hence

\[
(3.1.6) \quad \frac{1}{n} \sum_{k=1}^{n} c^k = E_0 + O\left(\frac{1}{n}\right) \quad (n \to \infty).
\]

It follows that
\[ (3.1.8) \quad \frac{1}{n} \sum_{k=1}^{n} p^{(k)}(A|x) = \frac{1}{n} \left[ \sum_{k=1}^{n} G^{k-1} \right] b(A) = \\
= T^{(1)}_{\hat{\beta}} \left( E_0 + O(\frac{1}{n}) \right) b(A) = \\
= T^{(1)}_{\hat{\beta}} \left( E_0 b(A) + O(\frac{1}{n}) \right) \quad (n \to \infty). \]

Since \( E_0 = \frac{1}{r} \sum_{j} \gamma_j \) with \( \gamma \) the properly normed left 1-eigenvector, and \( T^{(1)}_{\hat{\beta}}(x) \| = 1 \), we have

\[ (3.1.9) \quad T^{(1)}_{\hat{\beta}}(x) E_0 b(A) - T^{(1)}_{\hat{\beta}}(x) \sum_{j} \gamma_j b(A) = T^{(1)}_{\hat{\beta}}(A) - G(A). \]

Obviously, \( G \) is a probability measure, while

\[ (3.1.10) \quad \int_S \hat{\beta}(x) G(dx) = \frac{1}{r} \frac{1}{r} \sum_{j=1}^{r} \sum_{k=1}^{r} a_{j,k} e_{k,j} \gamma_k \gamma_j (dx) = \\
= \frac{1}{r} \frac{1}{r} \sum_{j=1}^{r} \sum_{k=1}^{r} \gamma_k \gamma_j \chi_{e_{j,k}}(A) = \\
= T^{(1)}_{\hat{\beta}} G(A) = \sum_{j} \gamma_j \gamma_j (A) = G(A), \]

so \( G \) is invariant with respect to \( \hat{\beta}(x) \). As for uniqueness: if \( F \) satisfies \( (3.1.1) \), then

\[ (3.1.11) \quad F(A) = \int_S F(x) F(dx) = \int_S T^{(1)}_{\hat{\beta}}(x) b(A) F(dx), \]

i.e. \( F \) is a linear combination of the \( b_j \). Put \( \hat{\beta}_j = \int S a_{j}(x) F(dx), \) then \( \hat{\beta} = 1 \) and by \( (3.1.11) \) we get (cf. \( (3.1.10) \))

\[ (3.1.12) \quad T^{(1)}_{\hat{\beta}}(A) = \int_S T^{(1)}_{\hat{\beta}}(x) b(A) F(dx) = T^{(1)}_{\hat{\beta}} G(A) \quad \text{for all } A \in \mathcal{S}. \]

Hence \( \hat{\beta} \) is a left 1-eigenvector of \( G \), satisfying \( T^{(1)}_{\hat{\beta}} 1 = 1 \). But this means \( \hat{\beta} = \gamma \).

For the special case of a noncyclic chain we have
THEOREM 3.1.3. If the kernel matrix is indecomposable and noncyclic (see definition 2.3.3), then
\[ p^{(n)}(A|x) = G(A) + o(x^n) \quad (n \to \infty) \]
for all \( x \in S \) and \( A \in S \), with \( G \) as in theorem 3.1.2.
\[ \square \]

PROOF. The only eigenvalue of modulus 1 is \( \lambda_0 = 1 \), so (2.1.11) becomes
\[ c^n = E_0 + o(x^n) \quad (n \to \infty) \]
Now (3.1.13) follows in the same way as (3.1.2) followed from (3.1.6).
\[ \square \]

Theorem 2.2.14 asserts that the state space \( S \) consists of a recurrent part \( R \) and a transient part \( T = S \setminus R \), uniquely determined modulo \( m \)-nullsets. The next theorem provides the relation with the invariant probability measure \( G \) in the case of an indecomposable chain.

THEOREM 3.1.4. In an indecomposable chain the support of the invariant probability measure \( G \) is equal to the recurrent part \( R \) of the state space.

PROOF. For the transient part \( T \) we have by lemma 2.2.2
\[ \lim_{n \to \infty} p^{(n)}(T|x) = 0 \quad \text{for all} \; x \in S , \]
and therefore
\[ G(T) = \lim_{n \to \infty} \frac{1}{n} \sum_{k=1}^{n} p^{(k)}(T|x) = 0 . \]
On the other hand, if \( G(A) = 0 \), then it follows from (3.1.2) that
\[ \sum_{k=1}^{n} p^{(k)}(A|x) = O(1) \quad (n \to \infty) , \]
hence \( A \) is transient.
\[ \square \]

We return to the chains given in section 1.5, and determine their invariant distributions.

EXAMPLE 3.1.5 (1.5.1). For the chain with \( p(y|x) = (1-x)y + x(2y-y^2) \) we obtain
(3.1.18) \[ C^n = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{3} & \frac{2}{3} \end{bmatrix} \begin{bmatrix} \frac{1}{2} & \frac{3}{4} \\ \frac{3}{4} & \frac{1}{2} \end{bmatrix} \begin{bmatrix} \frac{3}{4} & \frac{3}{4} \\ \frac{3}{4} & \frac{3}{4} \end{bmatrix} = (-\frac{1}{2})^n \begin{bmatrix} \frac{3}{4} & \frac{3}{4} \\ \frac{3}{4} & \frac{3}{4} \end{bmatrix} , \]

hence for \( x, y \in [0, 1] \) we have

(3.1.19) \[ p^{(n)} (y|x) = \frac{4}{7} y + \frac{3}{7} (2y - y^2) + O\left(\frac{1}{n}\right) \quad (n \to \infty) . \]

EXAMPLE 3.1.6 (1.5.2). The countable Markov chain defined by (1.5.5) has kernel matrix \( C \) given by (1.5.9), which satisfies

(3.1.20) \[ C^n = \begin{bmatrix} p & (1 - \log 2)p \\ p & (1 - \log 2)p \end{bmatrix} + O((1 - \log 2)^n) \quad (n \to \infty) , \]

where \( p := (2 - \log 2)^{-1} \). Hence

(3.1.21) \[ p_{jk}^{(n)} = p b_{1k} + (1 - \log 2)p b_{2k} + O((1 - \log 2)^n) = \begin{cases} \frac{1}{2} p + O((1 - \log 2)^n) & \text{if } k = 1 , \\ \left(\frac{1}{2}\right)^k p + O((1 - \log 2)^n) & \text{if } k \geq 2 . \end{cases} \]

EXAMPLE 3.1.7 (1.5.3). For the chain given by (1.5.9) and (1.5.10) we already found \( \lambda_0 = 1, \lambda_1 = (1+i)/2 \) and \( \lambda_2 = (1-i)/2 \) as the eigenvalues of the kernel matrix, see example 2.5.3. We obtain

(3.1.22) \[ C^n = \begin{bmatrix} 1 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \end{bmatrix} \lambda_1^n \begin{bmatrix} 1 & -1+i \\ 1 & -1-i \\ 1 & -1+i \end{bmatrix} + \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix} \lambda_2^n \begin{bmatrix} -1 & i & -i \\ -1 & i & -i \\ -1 & i & -i \end{bmatrix} . \]

The invariant distribution vector \( \mathbf{v}_x \) \( B \) is the transpose of

(3.1.23) \[ \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} . \]

\( \square \)
EXAMPLE 3.1.8 (1.5.4). The chain given by (1.5.11) and (1.5.12) is reducible, so there is no invariant distribution in this case, independent of the initial state.

EXAMPLE 3.1.9 (1.5.5). The kernel matrix $C$ given by (1.5.15) has eigenvalues $\lambda_0 = 1$, $\lambda_1 = -1$ and $\lambda_2 = 0$. We get

\[
C^n = \begin{pmatrix}
1 & 1 & 1 \\
1 & 1 & 1 \\
1 & 1 & 1
\end{pmatrix} + (-1)^n \begin{pmatrix}
-1 & -1 & -1 \\
-1 & -1 & -1 \\
-1 & -1 & -1
\end{pmatrix},
\]

and the invariant distribution vector is the transpose of

\[
\begin{pmatrix}
0 \\
1 \\
1
\end{pmatrix}.
\]

For all $j,k \in \{1,2,3,4\}$ we now have

\[
\frac{1}{n} \sum_{i=1}^{n} p_{jk}(i) = \frac{1}{n} + \mathcal{O}\left(\frac{1}{n}\right) \quad (n \to \infty).
\]

3.2. The central limit theorem

Let $X_1, X_2, \ldots$ be a sequence of random variables with finite second moments, and let $S_n = X_1 + \ldots + X_n$ $(n \geq 1)$. The sequence is said to obey the central limit theorem if $a_n$ and $b_n > 0$ exist such that

\[
\lim_{n \to \infty} P\left(\frac{S_n - a_n}{b_n} \leq x \right) = \Phi(x) \quad \text{for all } x \in \mathbb{R},
\]

where $\Phi$ is the standard normal distribution function. Several sufficient conditions for (3.2.1) are known, e.g., the Lindeberg-Lévy-Feller condition for independent random variables.

For dependent random variables several results are known under various conditions (see e.g. DÖPP [33]). In this section we give a simple proof of the central limit theorem for Markov processes of finite rank. Our result contains a theorem in KORNOMYSLOV [27] for finite Markov chains. Our proof, however, seems to be simpler and more transparent than the one for the special case treated there. After the completion of our proof we found that
ONICHECQ and MIHOC [25], dealing with the special case of finite Markov chains, have used a similar technique. Not all their arguments, however, are quite clear. As the same is true for ROMANOVSKI [27], we shall give rather detailed proofs here. Throughout this section we work under the following assumption. \( X_0, X_1, X_2, \ldots \) is a Markov chain of rank \( r \) with a standard representation for the kernel and with \( S = \mathbb{R}, S = B(\mathbb{R}) \). It is indecomposable, and the \( B_j \) have finite second moments, i.e.

\[
\int y^2 B_j(dy) < \infty \quad (j = 1, \ldots, r).
\]

The eigenvalues \( \lambda_0, \lambda_1, \ldots, \lambda_r \) of the kernel matrix \( C \) then satisfy (cf. (2.3.11))

\[
\lambda_k = e^{-\frac{2\pi i}{d} k} \quad \text{for } k = 0, \ldots, d-1 \quad (d \geq 1),
\]

\[
|\lambda_k| < c < 1 \quad \text{for } k \geq d,
\]

where \( 1 = \lambda_0, \lambda_1, \ldots, \lambda_{d-1} \) have multiplicity 1.

We introduce the following notations:

\[
S_n = X_1 + \ldots + X_n \quad (n \geq 1), \quad S_0 = 0;
\]

\[
\phi_n(t|x) = \mathbb{E}(e^{itS_n} \mid X_0 = x) \quad (n \geq 0, \ t \in \mathbb{R});
\]

\[
\tilde{C}(t) = \left( \tilde{c}_{jk}(t) \right)_{j,k=1}^r,
\]

\[
\tilde{c}_{jk}(t) = \int a_k(x) e^{itx} B_j(dx) \quad (t \in \mathbb{R}; \ j, k = 1, 2, \ldots, r);
\]

\[
\hat{S}_j(t) = \int e^{itx} B_j(dx) \quad (t \in \mathbb{R}; \ j = 1, 2, \ldots, r).
\]

In the case of independent variables one uses characteristic functions (c.f.'s), the crucial relation being

\[
\varphi_n\left( \frac{t}{\sqrt{n}} \right) = \left[ \varphi\left( \frac{t}{\sqrt{n}} \right) \right]^n,
\]

where \( \varphi_n \) and \( \varphi \) are the c.f. of \( S_n \) and \( X_1 \), respectively.
Of course, we also consider c.f.'s, viz. \( \varphi_n(t|x) \), and we shall show that an asymptotic relation replaces (3.2.8) in our case. We mainly use matrix theory arguments. The essential observation is

**Lemma 3.2.1.** The c.f. \( \varphi_n(t|x) \) is given by

\[
(3.2.9) \quad \varphi_n(t|x) = \sum_{m=0}^{n} \varphi(t)^{m} \Phi(t) \quad (n \geq 1).
\]

**Proof.** By the Markov property we have for \( n \geq 1 \)

\[
(3.2.10) \quad \varphi_n(t|x_0) := E(e^{itS_n} \mid x_0) =
\]

\[
= E(E(e^{it\sum_{j=1}^{n} X_j} \mid X_0, X_1) \mid x_0)
\]

\[
= E(E(e^{it\sum_{j=1}^{n} X_j} \mid X_1) \mid x_0)
\]

\[
= E(e^{it\sum_{j=1}^{n-1} \varphi_{n-1}(t|x_j) \mid x_0}),
\]

and hence for \( n \geq 2 \)

\[
(3.2.11) \quad \varphi_n(t|x) = \int \int_{\mathbb{R}^n} e^{itY} \varphi_{n-1}(t|y) \mathcal{P}(dy|x) \]

\[
= \int \int_{\mathbb{R}^n} \frac{1}{a_j(x)} e^{itY} \varphi_{n-1}(t|y) b_j(dy) \]

\[
= \frac{1}{a_j(x)} \int \int_{\mathbb{R}^n} e^{itY} \varphi_{n-1}(t|z) \mathcal{B}_j(dy)
\]

\[
= \frac{1}{a_j(x)} \int \int_{\mathbb{R}^n} e^{itY} \varphi_{n-1}(t|z) \mathcal{B}_j(dy)
\]

\[
= \frac{1}{a_j(x)} \int \int_{\mathbb{R}^n} e^{itY} \varphi_{n-1}(t|z) \mathcal{B}_j(dy)
\]

In matrix notation (3.2.11) reads

\[
(3.2.12) \quad \varphi_n(t|x) = \frac{1}{a_j(x)} \int \int_{\mathbb{R}^n} e^{itz} \varphi_{n-2}(t|z) \mathcal{B}_j(dz)
\]

Now (3.2.9) follows by induction from (3.2.12) for \( n \geq 2 \), whereas by definition \( \varphi_1(t|x) = \varphi(t) \). ☑
**Lemma 3.2.2.** The elements $\overline{\gamma}_{jk}(t)$ of $\overline{\gamma}(t)$ have uniformly continuous first and second derivatives on $\mathbb{R}$ for all $j$ and $k$.

**Proof.** By Lemma 1.1.5 the functions $a_j(x)$ are bounded, say $|a_j(x)| \leq L$. We therefore have

$$
(3.2.13) \quad \left| \frac{d^k}{dt^k} \left( a_k(x)e^{itx} \right) \right| \leq L |x|^k \quad (k = 1, \ldots, r; \; \xi = 0,1,2; \; t \in \mathbb{R}).
$$

From (3.2.2) it follows that $\int |x|^\xi \beta_j(dx) < \infty$ for $j = 1, \ldots, r$ and $\xi = 0,1,2$, hence the derivatives

$$
(3.2.14) \quad \left( \frac{d}{dt} \right)^\xi c_{jk}(t) = (i)^\xi \int x^\xi a_k(x)e^{itx} \beta_j(dx) \quad (\xi = 0,1,2)
$$

exist and are uniformly continuous on $\mathbb{R}$.

Before passing to the next lemmas, stating properties of the eigenvalues of $\overline{\gamma}(t)$, we note that $\overline{\gamma}(0) = \gamma$, the kernel matrix. The eigenvalues of $\overline{\gamma}(t)$ will be denoted by $\overline{\lambda}_k(t)$. Since the eigenvalues of a matrix, when properly identified, are continuous functions of the matrix elements, and since by Lemma 3.2.2 the $c_{jk}(t)$ are continuous, the eigenvalues $\overline{\lambda}_k(t)$ are continuous functions on $\mathbb{R}$. We number them in such a way that $\overline{\lambda}_k(0) = \lambda_k$ for $k = 0,1,\ldots,r-1$.

**Lemma 3.2.3.** Every eigenvalue $\overline{\lambda}(t)$ of $\overline{\gamma}(t)$ satisfies

$$
(3.2.15) \quad |\overline{\lambda}(t)| \leq 1.
$$

**Proof.** Let $\varphi(t)$ be a right $\overline{\lambda}(t)$-eigenvector, i.e.

$$
(3.2.16) \quad \sum_{j=1}^r \nu_j(t) \int a_j(y)e^{ity} \beta_j(dy) = \overline{\lambda}(t)\nu_k(t) \quad (k = 1, \ldots, r).
$$

If we multiply both sides of (3.2.16) by $a_k(x)$ and sum over $k$, we obtain, putting $\varphi_k(x) := \sum_{k=1}^r \nu_k(t)a_k(x)$,

$$
(3.2.17) \quad \int \varphi_k(y)e^{ity} \beta(dy|x) = \overline{\lambda}(t)\varphi_k(x).
$$

It follows that
(3.2.18) \[ | \tilde{x}(t)| \leq \int \frac{d}{d\tau} \int \theta_{\tau}(y) | y \leq \sup_{\tau} | \theta_{\tau}(y) | \] for all \( \tau \).

Because of the linear independence of the \( a_k(x) \) the function \( x \mapsto \theta_{\tau}(x) \) cannot vanish identically, so from (3.2.18) we conclude that \( | \tilde{x}(t)| \leq 1. \)

**Lemma 3.2.4.** There exists a positive number \( t_0 \) such that for all \( t \) with \( |t| < t_0 \) the eigenvalues \( \lambda_0(t), \ldots, \lambda_{d-1}(t) \) have multiplicity 1 and the other eigenvalues satisfy \( | \tilde{\lambda}_k(t) | < \varepsilon \).

**Proof.** As \( \lim_{\tau \to 0} \tilde{\lambda}_k(t) = \lambda_k \) and (by assumption) \( \lambda_0, \ldots, \lambda_{d-1} \) are distinct eigenvalues of \( C \) with multiplicity 1, also \( \tilde{\lambda}_0(t), \ldots, \tilde{\lambda}_{d-1}(t) \) are distinct and hence have multiplicity 1 for \( t \) sufficiently small. Also we have, cf. (3.2.3), \( | \tilde{\lambda}_k(t) | < \varepsilon \) for \( k \neq d \) and \( t \) sufficiently small.

Let \( J \) be the interval \((-\varepsilon_0, \varepsilon_0)\). From here on we consider only \( t \in J \).

**Lemma 3.2.5.** The eigenvalue \( \tilde{\lambda}_k(t) \) has continuous first and second derivatives on \( J \) \((k = 0, 1, \ldots, d-1)\).

**Proof.** \( \tilde{\lambda}_k(t) \) is a zero of the polynomial

\[ \det(\tilde{C}(t) - \lambda I) = \prod_{k=0}^{d} \theta_{\tau}(t) \lambda^k, \]

where by lemma 3.2.2 the \( s_k(t) \) have continuous second derivatives. Obviously,

\[ \prod_{k=0}^{d} s_k(t) \tilde{\lambda}_k(t)^k = 0, \]

whereas

\[ \prod_{k=0}^{d} s_k(t) \tilde{\lambda}_k(t)^{k-1} \neq 0 \quad \text{for} \quad t \in J, \]

because \( \lambda_k(t) \) is by lemma 3.2.4 a simple zero of \( \theta_{\tau}(t) \). Formally differentiating (3.2.20) we find that \( \tilde{\lambda}_k(t) \) exists on \( J \) and that
\[ (3.2.22) \quad \bar{\lambda}(t) = \frac{\sum_{k=0}^{k} s_k(t) \lambda_k(t)^k}{\sum_{k=0}^{k} s_k(t) \lambda_k(t)^{k-1}}. \]

In view of (3.2.21) it is now clear that also \( \lambda(t) \) exists and is continuous on \( J \).

**Lemma 3.2.6.** If \( y_k(t) \) is a suitably normed right or left \( \bar{\lambda}(t) \)-eigenvector, then \( y_k(t) \) has a continuous second derivative in a neighborhood \( J_0 \) of \( 0 \).

**Proof.** We give the proof only for a right eigenvector, as the other case is completely analogous, and we omit the index \( i \). The equation

\[ (3.2.23) \quad [\mathcal{C}(t) - \bar{\lambda}(t)I]y(t) \equiv 0 \]

determines \( y(t) \) up to a norming factor. For each \( t \in J \) the matrix

\[ \mathcal{C}(t) = \bar{\lambda}(t)I \]

has rank \( r-1 \). Without loss of generality we may assume that the \( r \)-th row of \( \mathcal{C}(0) - \bar{\lambda}(0)I = \mathcal{C} - \lambda I \) is dependent of the other \( r-1 \) rows. By continuity arguments we then have the same for each matrix \( \mathcal{C}(t) - \bar{\lambda}(t)I \) with \( t \) in some open set \( J_0 \subset J \), containing \( 0 \). It follows that we can find a constant vector \( \bar{y} \), such that the row vector \( \bar{y} \) is independent of the first \( r-1 \) rows of \( \mathcal{C}(t) - \bar{\lambda}(t)I \), for all \( t \in J_0 \).

Let \( D(t) \) be the matrix obtained by replacing the \( r \)-th row of \( \mathcal{C}(t) - \bar{\lambda}(t)I \) by \( \bar{y} \). Then \( D(t) \) has full rank and \( y(t) \) is completely determined for \( t \in J_0 \) by

\[ (3.2.24) \quad D(t)y(t) = \bar{y}(0,0,\ldots,0,1). \]

Using the lemmas 3.2.2 and 3.2.5 we see that \( D'(t) \) exists and is continuous. By formal differentiation of (3.2.24) we get

\[ (3.2.25) \quad y'(t) = - D^{-1}(t)D'(t)y(t) \]

and

\[ (3.2.26) \quad y''(t) = - D^{-1}(t)[DD'(t)y(t) - 2D'(t)y'(t)]. \]

Clearly, \( y'(t) \) and \( y''(t) \) exist and are continuous on \( J_0 \).
From here on we take for \( J \) a neighbourhood of \( t = 0 \), such that both \( v_k(t) \) and \( u_k(t) \) have continuous first and second derivatives on \( J \) for \( k = 0, 1, \ldots, d-1 \).

The spectral decomposition of \( \hat{C}(t) \) is given by

\[
(3.2.27) \quad \hat{C}(t) = \sum_{k=0}^{d-1} \lambda_k(t) v_k(t) \bar{v}_k(t) + \sum_{k=0}^{d-1} \left[ \lambda_k(t) E_k(t) \right] E_k(t) \quad (t \in J). 
\]

The number of terms of the second sum does not exceed \( r - d \), but it may depend on \( t \). This is of no importance; only the first sum matters, as we shall see.

**Lemma 3.2.7.** The matrix \( E_k(t) \) has continuous first and second derivatives on \( J \) for \( k = 0, 1, \ldots, d-1 \). \( \square \)

**Proof.** For \( k = 0, 1, \ldots, d-1 \) we have \( E_k(t) = v_k(t)\bar{v}_k(t) \). \( \square \)

**Lemma 3.2.8.** For all \( x \)

\[
\lim_{t \to 0} \frac{\hat{a}(x) E_k(t) \bar{v}_k(t)}{t} = \begin{cases} 1 & \text{for } k = 0, \\ \frac{1}{k} & \text{for } k = 1, \ldots, d-1. \\ \end{cases} \quad \square 
\]

**Proof.** Since both \( E_k(t) \) and \( \bar{v}_k(t) \) are continuous on \( J \), we need only consider \( \hat{a}(x) = \frac{1}{\hat{a}(x) \bar{v}_k(0) \bar{v}_k(0)} \). Now \( \bar{v}_k(0) = 1 \) is a right \( k \)-eigenvector of \( \hat{C}(0) \). From the orthogonality properties of eigenvectors corresponding to different eigenvalues it follows that \( \bar{v}_k(0) \bar{v}_k(0) = 0 \) for \( k = 1, \ldots, d-1 \). For \( k = 0 \) we have \( E_0(0) = E_0 \) and hence \( \hat{a}(x) E_0(0) \bar{v}_k(0) = 1 \). \( \square \)

For brevity we shall write

\[
(3.2.29) \quad u_k(t) := \frac{\hat{a}(x) E_k(t) \bar{v}_k(t)}{t}. 
\]

We are now ready to prove the main result of this section, a central limit theorem for Markov chains of finite rank.
THEOREM 3.2.9. Let \( X_0, X_1, \ldots \) be an indecomposable Markov chain of finite rank \( r \) satisfying (3.2.2). Put \( \nu := -t \gamma^\prime_0(0) \) and \( \nu := (\gamma^\prime_0(0))^2 - \gamma_0(0) \). Then \( \nu \) and \( \nu \) are real constants, \( \nu \) is nonnegative, and \( \phi_n \), the c.f. defined by (3.2.5), satisfies the following relation:

\[
\lim_{n \to \infty} e^{-iut\nu/n} \phi_n \left( \frac{x}{\sqrt{n}} \right) = e^{-iut^2} \quad (t \in \mathbb{R}, \ x \in \mathbb{R}).
\]

PROOF. We substitute the spectral decomposition (3.2.27) of \( \tilde{C}(t) \) in (3.2.9) and obtain for \( t \in \mathbb{T} \)

\[
\phi_n(t|x) = \psi_n(t|x) + \zeta_n(t|x),
\]

where, see (3.2.3) and lemma 2.1.8,

\[
\phi_n(t|x) := \frac{d_{-1}}{\sqrt{n}} \gamma_n^{-1}(t) \nu(t|x)
\]

and

\[
\zeta_n(t|x) := \sum_{k=0}^{\infty} \gamma_k(t) E_k(t) (k+1) \delta(t) (n=0).
\]

Using lemma 3.2.8 we find

\[
\zeta_n \left( \frac{r_0}{\sqrt{n}} \right) = \gamma_n^{-1}(t) \left[ 1 + o(1) \right] \quad (n \to \infty).
\]

By lemma 3.2.5 the function \( \gamma_0(t) \) has a continuous second derivative on \( J \), hence

\[
\zeta_n \left( \frac{r_0}{\sqrt{n}} \right) = \exp \left( (n-1) \log \left[ 1 + \frac{t^2}{2n} \gamma_0(0) + o \left( \frac{1}{n} \right) \right] \right) =
\]

\[
= \exp \left( (n-1) \left[ \frac{1}{\sqrt{n}} \gamma_0(0) + \frac{t^2}{2n} \gamma_0(0) - \frac{1}{2n} \gamma_0(0) + o \left( \frac{1}{n} \right) \right] \right) =
\]

\[
= \exp \left( (n-1) \left[ \frac{1}{\sqrt{n}} \gamma_0(0) - \frac{1}{2n} \gamma_0(0) + o \left( \frac{1}{n} \right) \right] \right) =
\]

\[
= \exp \left( r_0 t \gamma_0(0) - \frac{1}{2n} \gamma_0(0) + o(1) \right) \quad (n \to \infty).
\]

Substitution of (3.2.35) into (3.2.34) proves the relation (3.2.30).
It remains to show that $\nu = -i \lambda_0'(0)$ is real and that $\nu = (\lambda_0'(0))^2 + - \lambda_0''(0)$ is real and nonnegative. Now first suppose that $\nu$ is real. Then the function $e^{-i\nu t} / n \langle \mathcal{F}_{n} \frac{1}{n} | x \rangle$ is a.c.f. for every $n$, and we just have proved that this sequence of c.f.'s has a limit function $e^{-i\nu t^2}$ for $n \to \infty$. This limit function is continuous at $t = 0$ and is therefore itself a c.f. From this it follows that $\nu$ is real and nonnegative.

To complete the proof we must show that $\nu$ is real. In order to do this we differentiate (3.2.31) and substitute $t = 0$. First we note that $\varphi_n(t|x)$ has a continuous derivative, in view of lemmas 3.2.1 and 3.2.2 and of the fact that $\xi(t)$ has a continuous derivative. We have

(3.2.36) \quad \varphi_n'(0|x) = iE(S_n | X_0 = x).

As $\varphi_n(t|x)$ (cf. (3.2.32)) has a continuous derivative on account of the lemmas 3.2.5 and 3.2.7, with

(3.2.37) \quad \varphi_n'(0|x) = \left[ (n-1)\tilde{\Sigma}_0'(0) + \varphi_n'(0|x) \right] + \frac{d-1}{k=1} \tilde{\Sigma}_k^{-1}(0)\varphi_k'(0|x),

it follows that $\xi_n(t|x)$ also has a continuous derivative. From (3.2.33) it follows, by a result in matrix theory proved in the appendix as lemma A.2.1, that

(3.2.38) \quad \xi_n'(0|x) = o(n^{-1}) \quad (n \to \infty).

Combining these results we obtain by differentiation of (3.2.31)

(3.2.39) \quad iE(S_n | X_0 = x) = (n-1)\tilde{\Sigma}_0'(0) + \varphi_n'(0|x) + \frac{d-1}{k=1} \tilde{\Sigma}_k^{-1}(0)\varphi_k'(0|x) + o(1)

\quad \quad (n \to \infty),

which yields

(3.2.40) \quad \nu = -i \lambda_0'(0) = \frac{1}{n} E(S_n | X_0 = x) + o(n^{-1}) \quad (n \to \infty).

But this means that $\nu$ is real, which completes the proof.
3.3. Expectation and variance of $S_n$

We now explore relation (3.2.30) further to obtain additional information about the constants $\nu$ and $\nu$. From (3.2.40) we see that we may replace $\nu$ by $\frac{1}{n} E(S_n \mid X_0 = x)$ in (3.2.30). The relation with the invariant distribution function $G(y)$ is given in the following theorem.

**Theorem 3.3.1.**

\begin{align*}
\frac{1}{n} E(S_n \mid X_0 = x) \rightarrow \nu = \int y G(dy) & \quad \text{as } n \rightarrow \infty. \quad \square
\end{align*}

**Proof.** We have, using (3.1.6),

\begin{align*}
\frac{1}{n} E(S_n \mid X_0 = x) &= \frac{1}{n} \sum_{k=1}^{n} E(X_k \mid X_0 = x) = \\
&= \frac{1}{n} \sum_{k=1}^{n} \sum_{i=1}^{\nu} \Gamma_{\frac{x}{\nu}}^i \theta_{x}^{i-1} \int \frac{G(dy)}{\nu} = \\
&= \frac{\nu}{n} \sum_{i=1}^{\nu} \Gamma_{\frac{x}{\nu}}^i \theta_{x}^{i-1} \int \frac{G(dy)}{\nu} = \\
&= \frac{\nu}{n} \left( \frac{1}{\nu} \sum_{i=1}^{\nu} \theta_{x}^{i-1} \right) \int \frac{G(dy)}{\nu} = \\
&= \frac{\nu}{n} \left( \frac{1}{\nu} \sum_{i=1}^{\nu} \theta_{x}^{i-1} \right) \int \frac{G(dy)}{\nu} = \\
&= \frac{1}{n} \int y G(dy) + o\left(\frac{1}{n}\right) \quad (n \rightarrow \infty). \quad \square
\end{align*}

We shall write

\begin{align*}
E_x \ Y &:= E(Y \mid X_0 = x), \\
\text{Var}_x \ Y &:= E_x \ Y^2 - (E_x \ Y)^2, \\
\text{Cov}_x (Y, Z) &:= E_x \ YZ - E_x \ E_x \ Y \ E_x \ Z.
\end{align*}

The constant $\nu$ is related to $\text{Var}_x S_n$ in the same way as $\nu$ is related to $E_x S_n$, just as one would expect. The calculations are a bit more cumbersome, but they follow essentially the same pattern. As we show by an
example at the end of this section, there is no direct relation with the variance of the invariant distribution \( \gamma \), similar to (3.3.1).

**THEOREM 3.3.2.**

\[(3.3.4) \quad \nu = \frac{1}{n} \operatorname{var}_n S_n + o(1) \quad (n \to \infty) \]

**PROOF.** We have to determine second derivatives, the existence of which is guaranteed by the lemmas preceding theorem 3.2.9. The existence of \( \zeta_n^*(0|x) \) again follows from the existence of the second derivatives of the other terms in (3.2.31). We have (cf. (3.2.36), (3.2.37) and (3.2.38))

\[(3.3.5) \quad \zeta_n^*(0|x) = -E_x(S_n - E_x S_n)^2 = \operatorname{var}_n S_n ;
\]

\[(3.3.6) \quad \zeta_n^*(0|x) = [(n-1)\tilde{\lambda}_0^*(0) + (n-2)(n-1)\tilde{\lambda}_1^*(0))^2 + 2(n-1)\tilde{\lambda}_0^*(0)\tilde{\lambda}_1^*(0)|x| + O(1)] +
\]

\[+ \frac{d-1}{\Delta} \{2(n-1)\tilde{\lambda}_0^*(0)\tilde{\lambda}_1^*(0) + O(1) \quad (n \to \infty) ;
\]

\[(3.3.7) \quad \zeta_n^*(0|x) = o(e^{b^2}) \quad (n \to \infty) ,
\]

where for (3.3.7) we applied lemma A.2.2 of the appendix. This leads to

\[(3.3.8) \quad \frac{1}{n} \operatorname{var}_n S_n = (n-1)\tilde{\lambda}_0^*(0) + (n-2)(n-1)\tilde{\lambda}_1^*(0) +
\]

\[+ \frac{d-1}{\Delta} \{2(n-1)\tilde{\lambda}_0^*(0)\tilde{\lambda}_1^*(0) + O(1) \quad (n \to \infty) ,
\]

which by (3.2.39) yields (recall that \( \nu = (\tilde{\lambda}_1^*(0))^2 - \tilde{\lambda}_0^*(0) \))

\[(3.3.9) \quad \nu = \frac{1}{n-1} \operatorname{var}_n S_n + 2 \frac{d-1}{\Delta} \{ \tilde{\lambda}_0^*(0)\tilde{\lambda}_1^*(0) - \tilde{\lambda}_0^*(0)\tilde{\lambda}_1^*(0)\tilde{\lambda}_0^*(0)\tilde{\lambda}_1^*(0) + O(1) \quad (n \to \infty) .
\]

The sum in the right-hand side of (3.3.9) is bounded; it takes only a finite number of values since \( \tilde{\lambda}_1^*(0) = \tilde{\lambda}_1 = \hat{\alpha} n^{1/(2d)} \), and hence \( \frac{1}{n-1} \operatorname{var}_n S_n \) is bounded.
We proceed by proving that the sum only takes the value 0. We have

\[(3.3.10) \quad \var_{X_n} = \var_{X_n} (X_{n-1} + X_n) =\]
\[= \var_{X} X_{n-1} + \var_{X} X_n + 2 \cov_{X} (X_{n-1}, X_n) =\]
\[= \var_{X} X_{n-1} + \var_{X} X_n + 2 \sum_{j=1}^{n-1} \cov_{X_n} (X_{j}, X_{n-1}).\]

From

\[(3.3.11) \quad E_{X_n} X_n^2 = \int y^2 f_{X_n}(dy|x) = \int y^2 E_{X_n}(y)\]

we see that \(\var_{X_n} X_n\) is bounded.

Further for \(j < n\) we obtain \((n - j, j = 0), \) using theorem 3.3.1 and lemma 2.1.7, and writing \(U(x) := E_{X_n} (\lambda_k X_n + X_k)^K,\)

\[(3.3.12) \quad E_{X_n} X_j X_n = \int \int \int y z f_{X_n}(dy|x) z^{(n-j)} (dz|y) =\]
\[= T_{\var}(x) c^{j-1} \int y \var(dx) \var(y) c^{n-j-1} \int z \var (dz) =\]
\[= T_{\var}(x) [s_0 + \sum_{k=1}^{d-1} \lambda_k^{j-1} \var_k + U(j-1)] \int y \var (dy) -\]
\[= T_{\var}(y) [s_0 + \sum_{k=1}^{d-1} \lambda_k^{n-j-1} \var_k + U(n-j-1)] \int z \var (dz) =\]
\[= \left[\var + \sum_{k=1}^{d-1} \lambda_k^{n-j-1} \var_k + T_{\var}(x) U(j-1)\right] \int y \var (dy) -\]
\[= \left[\var + \sum_{k=1}^{d-1} \lambda_k^{n-j-1} \var_k + T_{\var}(y) U(n-j-1)\right] \int z \var (dz),\]

where \(\var_k(x)\) is given by \((2.3.8)\)

From theorem 3.3.4 we know that there are cyclically moving sets \(S_0 + S_1 + \ldots + S_d = S_0.\) If \(x \in S_k,\) then the first integral is in fact over \(S_{k+j}\) and the second over \(S_{k+n}\) (indices taken modulo 4). We obtain
\[ E_{x_j} X_k = \left[ \sum_{i=1}^{d-1} \lambda_i^k \right] y \mathbb{B}(dy) + \sum_{S_k \neq j} \left[ \sum_{i=1}^{d-1} \lambda_i^k \right] z \mathbb{B}(dz) \]

In exactly the same way we find an expression for \( E_{x_j} X_k \), equal to (3.3.13), except that the term \( \frac{1}{\alpha}(\gamma) \mathbb{W}(n-j-1) \) is replaced by \( \frac{1}{\alpha}(\gamma) \mathbb{W}(n-1) \).

If \( x \) is not an element of one of the transient parts \( S_k \), but an element of the transient part \( T \) of \( S \), then we have to substitute (cf. (2.3.2))

\[ v_k(x) = \sum_{n=1}^{\infty} \lambda_k^n \sum_{k=0}^{d-1} \left[ v_{k}(y)f^{(n)}(dy|x) = \sum_{n=1}^{\infty} \sum_{k=0}^{d-1} \lambda_k^n f^{(n)}(S_k|x) \right] \]

into (3.3.12). In any event we obtain, using \( W(n) = o(n) \) \( n \to \infty \),

\[ \sum_{j=1}^{n-1} \text{cov}(X_j, X_n) = \sum_{j=1}^{n-1} \left( \sum_{j=1}^{m-1} \sum_{k=0}^{d-1} \frac{1}{\alpha}(\gamma) \mathbb{W}(n-j-1) \right) = O(1) \quad (n \to \infty). \]

It is now clear from (3.3.10) and (3.3.15) that

\[ \frac{\text{var}_x S_n}{n-1} = \frac{\text{var}_x S_n \to \mathbb{E}(S_n^2)}{n-2} + O(1) \quad (n \to \infty). \]

Returning to relation (3.3.9), we conclude from (3.3.16) that

\[ \sum_{i=1}^{d-1} \lambda_i^{n-2} \gamma_i(0) = \gamma_x(0) \mathbb{B}(0|x) = \text{constant}. \]

It is a simple algebraic result, given in the appendix as lemma A.2.5, that the constant in (3.3.17) must be \( 0 \). Now (3.3.4) follows from (3.3.9).
Theorem 3.2.9 asserts that

\[(3.3.18) \quad \lim_{n \to \infty} P \left( \frac{S_n - nu}{(nv)^{\frac{1}{2}}} \leq y \mid X_0 = x \right) = \Phi(y), \]

where \(\Phi\) is the standard normal distribution function. The theorems in this section yield

**Corollary 3.3.3.** In (3.3.18) one may replace the normalizing constants \(nu\) and \(nv\) by \(E_S n\) and \(\text{var}_n S\), respectively.

**Proof.** On account of the theorems 3.3.1 and 3.3.2 we have

\[(3.3.19) \quad \frac{S_n - E_S n}{(\text{var}_n S)^{\frac{1}{2}}} = \frac{S_n - nu + O(1)}{(nv)^{\frac{1}{2}}} (1 + o(1)) = \frac{S_n - nu}{(nv)^{\frac{1}{2}}} (1 + o(1)) + o(1) (n \to \infty), \]

hence for each \(y \in \mathbb{R}\)

\[(3.3.20) \quad P \left( \frac{S_n - E_S n}{(\text{var}_n S)^{\frac{1}{2}}} \leq y \mid X_0 = x \right) = P \left( \frac{S_n - nu}{(nv)^{\frac{1}{2}}} \leq y + o(1) \mid X_0 = x \right), \]

so that by (3.3.18) and by the continuity of \(\Phi\) we obtain

\[(3.3.21) \quad \lim_{n \to \infty} P \left( \frac{S_n - E_S n}{(\text{var}_n S)^{\frac{1}{2}}} \leq y \mid X_0 = x \right) = \Phi(y). \]

From (3.3.11) we can deduce, cf. (3.3.2),

\[(3.3.22) \quad \frac{1}{n} \sum_{i=1}^{n} E_x X_i^2 = \int y^2 G(dy) + o(1/n) \quad (n \to \infty), \]

and hence

\[(3.3.23) \quad \frac{1}{n} \sum_{i=1}^{n} \text{var}_x X_i = \int y^2 G(dy) - y^2 + o(1/n) \quad (n \to \infty). \]
Due to the covariances there is no direct relation between \( \text{var} X_n \) and the variance of \( C \), similar to (3.3.1) and (3.3.23). This shows in the following example.

Consider the Markov chain of example 1.5.5 with \( S = \{1,2,3,4\} \) and transition matrix

\[
P = \begin{pmatrix}
0 & 1 & 0 & 0 \\
0 & 0 & 1 & 0 \\
0 & 1 & 0 & 0 \\
1 & 0 & 0 & 0 \\
\end{pmatrix}
\]

One easily verifies that for all \( n \geq 1 \)

\[
P(S_{2n} = 5n \mid X_0 = 1) = 1,
\]
(3.3.24)

\[
P(S_{2n+1} = 5n+2 \mid X_0 = 1) = P(S_{2n+1} = 5n+4 \mid X_0 = 1) = \frac{1}{2},
\]

and hence that

(3.3.25) \( \text{var}_1 S_{2n} = 0 \), \( \text{var}_1 S_{2n+1} = 1 \).

Similar results hold for the conditions \( X_0 = 2 \), \( X_0 = 3 \) and \( X_0 = 4 \). On the other hand we have found in (3.1.25) that the invariant distribution vector in this case is \( \pi = (\frac{1}{4} \frac{1}{4} \frac{1}{4} \frac{1}{4}) \), so that

(3.3.26) \[ \int y^2 g(dy) - y^2 = \frac{5}{8}, \]

which is larger than \( \text{var}_1 S_n \) due to the negative correlation apparent from the matrix \( P \).

### 3.4. Extreme values

In classical extreme-value theory one considers a sequence \( X_1, X_2, \ldots \) of i.i.d. random variables and studies the distribution of \( M_n := \max(X_1, \ldots, X_n) \) as \( n \to \infty \). In particular, one asks whether sequences of normalizing constants \( a_n \) and \( b_n \) exist, such that \( a_n(M_n - b_n) \) has a nondegenerate limiting distribution. The following classical theorem states that there are only three possible types of such limiting distributions.
THEOREM 3.4.1 (Goedenko). If $X_1, X_2, \ldots$ are i.i.d. random variables, $N_n := \max(X_1, \ldots, X_n)$, and for some sequence of real numbers $a_n > 0$ and $b_n$, the distribution function of the normalized maximum $a_n(N_n - b_n)$ converges to a nondegenerate distribution function $F$, then $F$ is of one of the following types:

$$F_1(x) = \begin{cases} \exp(-e^{-x}) & \text{for } -\infty < x < \infty; \\ 0 & \text{for } x \leq 0; \\ \exp(-e^{-x}) & \text{for } x > 0; \end{cases}$$

$$F_{II}(x) = \begin{cases} \exp(-x^2) & \text{for } x \leq 0; \\ 0 & \text{for } x > 0; \end{cases}$$

$$F_{III}(x) = \begin{cases} \exp(-(-x)^2) & \text{for } x \leq 0; \\ 1 & \text{for } x > 0; \end{cases}$$

Here $a$ is a positive constant.

A detailed proof can be found, for instance, in DE HAAN [14], where also the conditions for convergence to the various types are studied.

If $F$ is the common distribution function of the $X_i$, we can summarize theorem 3.4.1 by the following implication, denoting weak convergence to a nondegenerate distribution function by $\overset{w}{\Rightarrow}$,

$$F_n^A(a_n y + b_n) \overset{w}{\Rightarrow} F(y) \text{ as } n \to \infty \Rightarrow F \in \{F_1, F_{II}, F_{III}\}.$$  (3.4.2)

Theorem 3.4.1 can be generalized to cases where the variables $X_1, X_2, \ldots$ are dependent; of course, restrictions have to be imposed on the nature of the dependence. LEADBETTER [122] considers stationary sequences satisfying a weak mixing condition. Here we consider Markov chains of finite rank, and prove an analogue of theorem 3.4.1 without essential further restrictions.

We define the distribution function $F_n$ by

$$F_n(y|x) := P(N_n \leq y | X_0 = x),$$  (3.4.3)

and the $\tau \times \tau$ matrix $C(y) = (c_{jk}(y))$ by

$$c_{jk}(y) := \int_{-\infty}^{y} a_k(x) b_j(dx).$$  (3.4.4)
We proceed along the following lines. First we prove that the behavior of \( P_n(y|x) \) as a function of \( n \) is governed by \( C_n^{n-1}(y) \), and that \( C(y) \) in many respects behaves like the matrix \( \tilde{C}(x) \) of section 3.2. We shall show that \( C(y) \) has a dominant eigenvalue \( \lambda_0(y) \), and that 
\[
P_n(y_n|x) = \lambda_0^n(y_n) \quad \text{as} \quad n \to \infty. \]
Then, we prove the analogue of (3.4.2), with \( P_0 \) replaced by \( \lambda_0 \). In the last part of this section we study the relation between \( P_n(y|x) \) and \( C(y) \), where \( C \) is the invariant distribution of the chain.

Again we restrict ourselves to indecomposable chains, with \( S = \mathbb{N} \). But even in an indecomposable chain we cannot, in general, expect \( P_n(y_n|x) \) to be asymptotically independent of the initial state \( X_0 = x \). The influence of transient states that are greater than the recurrent states does not disappear. Since transient sets are not particularly interesting here, we restrict to irreducible chains (cf. definition 2.3.1). So we assume (theorem 3.1.4) that the support of the invariant distribution \( C \) has \( m \)-measure 1. Part of the results, however, also hold for indecomposable reducible chains.

The following lemma is the analogue of lemma 3.2.1.

**Lemma 3.4.2.** In an irreducible Markov chain of finite rank, the distribution function \( P_n \) defined in \((3.4.3)\) satisfies

\[
F_n(y|x) = T_n(x)C^{n-1}(y)\mathcal{B}(y) \quad (n = 1, 2, \ldots),
\]

with \( C(y) \) given by \((3.4.4)\).

**Proof.** Obviously \((3.4.5)\) holds for \( n = 1 \). Now suppose \((3.4.5)\) holds for a certain \( n \). Then

\[
F_{n+1}(y|x) = \int F_n(y|x)F(dy|x) = \int \int \mathcal{B}(y)\mathcal{B}(y) = \int a_n(x)B_j(dy) - \int a_n(x)B_j(dy) = \int a_n(x)C^n(y)\mathcal{B}(y).
\]

By induction \((3.4.5)\) holds for all \( n \geq 1 \).
As an analogue to lemma 3.2.3 we have

**Lemma 3.5.3.** Every eigenvalue \( \lambda(y) \) of \( C(y) \) satisfies

\[
(3.4.7) \quad |\lambda(y)| \leq 1.
\]

**Proof.** Let \( y(y) \) be a right \( \lambda(y) \)-eigenvector; then, writing \( \int_0^y \) for \( \int_{-\infty}^y \),

\[
(3.4.8) \quad \sum_{j=1}^r v_j(y) \int_0^y a_j(z) e_k(\lambda) = \lambda(y) v_k(y) \quad (k = 1, \ldots , r).
\]

If we multiply both sides of (3.4.8) by \( s_k(x) \) and then sum over \( k \), we obtain, putting \( \theta_y(x) := \sum_{k=1}^r v_k(y) s_k(x) \),

\[
(3.4.9) \quad \int_0^y \theta_y(z) P(dz|x) = \lambda(y) \theta_y(x).
\]

It follows that

\[
(3.4.10) \quad |\lambda(y)| |\theta_y(x)| \leq \int_0^y |\theta_y(z)| P(dz|x) \leq \sup_{x} |\theta_y(x)| \quad \text{for all } x.
\]

Taking the supremum over \( x \) in (3.4.10) yields (3.4.7). \( \square \)

**Lemma 3.5.4.** If the sequence \( y_1, y_2, \ldots \) is such that \( \lim_{n \to \infty} C(y_n) = 0 \), then \( \lim_{n \to \infty} |\lambda(y_n)| < 1 \) for every eigenvalue of \( C(y) \).

**Proof.** From (3.4.10) it follows that for all \( x, y \in \mathbb{R} \) we have

\[
(3.4.11) \quad |\lambda(y)| |\theta_y(x)| \leq \int_0^y |\lambda(y)| |\theta_y(z)| P(dz|x) \leq
\]

\[
\leq \int_0^y \{ \int_0^y |\theta_y(z)| P(dz|u) \} P(du|x) \leq
\]

\[
\leq \int_0^y |\theta_y(z)| \gamma(2)(dz|u),
\]

and in general we find
(3.4.12) \[ |\lambda(y)|^k |\mathcal{g}_y(x)| \leq \sum_{k=1}^Y |\mathcal{g}_y(x)|^k \mathcal{P}^k(d\xi|x) \quad (k = 1, 2, \ldots). \]

Applying theorem 3.1.2, in particular relation (3.1.6), we obtain

\[
\sum_{k=1}^N \left( \int |\mathcal{g}_y(x)|^k \mathcal{P}^k(d\xi|x) \right) - \int |\mathcal{g}_y(x)| \mathcal{T}_\lambda(d\xi) \mathcal{E}_0 + \mathcal{O}(\frac{1}{N}) \mathcal{P}(d\xi) - \int |\mathcal{g}_y(x)| \mathcal{G}(d\xi) + \mathcal{O}(\frac{1}{N}) \leq \sup_x |\mathcal{g}_y(x)| \mathcal{G}(y) + \mathcal{O}(\frac{1}{N}) \quad (N = \infty).
\]

Now (3.4.12) and (3.4.13) yield

\[
\frac{1}{N} \sum_{k=1}^N |\lambda(y)|^k \leq \mathcal{G}(y) + \mathcal{O}(\frac{1}{N}) \quad (N = \infty).
\]

Hence for a sufficiently large, fixed \(N\) we have

\[
\limsup_{n \to \infty} \frac{1}{N} \sum_{k=1}^N |\lambda(y_n)|^k < 1.
\]

It follows that

\[
\frac{1}{N} \sum_{k=1}^N (\limsup_{n \to \infty} |\lambda(y_n)|)^k < 1
\]

and hence that \(\limsup_{n \to \infty} |\lambda(y_n)| < 1.\)

**Lemma 3.4.5.** If the sequence \(y_1, y_2, \ldots\) is such that \(\lim_{n \to \infty} \mathcal{G}(y_n) = 1\), then

\[
\lim_{n \to \infty} \mathcal{B}(y_n) = 1
\]

and

\[
\lim_{n \to \infty} \mathcal{C}(y_n) = \mathcal{C}.
\]

**Proof.** From the assumption that the chain is irreducible it follows that the interval \(\lim_{n \to \infty} y_n = 0\) is an \(\mathcal{N}\)-nullset and hence a \(\mathcal{B}\)-nullset. This proves (3.4.17). Relation (3.4.18) follows from
\[ (3.4.19) \quad c_{jk}(y_n) = \int_{-\infty}^{y_n} a_k(x) E_j(x) \, dx = \int_{E_k} a_k(x) 1_{(-\infty, y_n]}(x) \, dx. \]

\[ = \int_{E_k} a_k(x) E_j(x) = c_{jk} \quad \text{as} \quad n \to \infty. \]

The eigenvalues \( \lambda_i(y) \) of \( C(y) \) can be identified in the same way as the eigenvalues \( \lambda(t) \) of \( \overline{C}(t) \) in section 3.2 (cf. the remarks preceding lemma 3.2.3), since \( \lim_{n \to \infty} G(y_n) = 1 \) implies, on account of (3.4.18),

\[ (3.4.20) \quad \lim_{n \to \infty} \lambda_i(y_n) = \lambda_i \quad (i = 0, 1, \ldots, r-1), \]

i.e. \( \lambda_i \) is continuous at "infinity". It follows that for large \( y \) (i.e. for \( G(y) \) close to 1) the spectral decomposition of \( C(y) \) can be written as

\[ (3.4.21) \quad C(y) = \sum_{k=0}^{d-1} \lambda_k(y) E_k(y) + \sum_{k=0}^{d} \left[ \lambda_k(y) E_k(y) + \nu_k(y) \right]. \]

Here we have, if \( \lim_{n \to \infty} G(y_n) = 1 \), that (cf. lemma 3.2.7)

\[ (3.4.22) \quad \lim_{n \to \infty} E_k(y_n) = E_k \]

and (cf. lemma 3.2.8)

\[ (3.4.23) \quad \lim_{n \to \infty} \left[ a_k(x) E_k(y_n) \right] = \begin{cases} 1 & \text{for } k = 0, \\ 0 & \text{for } k = 1, \ldots, d-1. \end{cases} \]

Finally, we need the following lemma.

**Lemma 3.4.6.** Let \( F_1, F_2, \ldots \) be a sequence of distribution functions, and suppose there exist real constants \( a_n > 0, b_n, \gamma_n > 0 \) and \( \alpha_n \) such that

\[ (3.4.24) \quad \frac{a_n}{\alpha_n} + b_n \gamma_n \leq V(y), \]

where \( U \) and \( V \) are non-degenerate distribution functions. Then

\[ (3.4.25) \quad \lim_{n \to \infty} \frac{a_n}{\alpha_n} = \rho > 0, \quad \lim_{n \to \infty} \frac{b_n}{\alpha_n} = \beta, \quad \text{and} \quad V(y) = U(\rho y + \beta). \]
PROOF. We refer to FELLER [7] for a proof of this lemma.

We now state the main theorem of this section:

**Theorem 3.4.2.** Let $X_0, X_1, \ldots$ be an irreducible Markov chain of finite rank. Suppose for some fixed $x \in \mathcal{S}$ there exist real constants $a_n > 0$ and $b_n$, and a nondegenerate distribution function $F$, such that

$$
(3.4.26) \quad F_n(a_n y + b_n | x) \overset{p}{\rightarrow} F(y).
$$

Then (3.4.26) holds for all $x \in \mathcal{S}$ (the limit being independent of $x$) and we have

$$
(3.4.27) \quad \lambda_0^n(a_n y + b_n) \overset{p}{\rightarrow} F(y)
$$

and $P \in (\mathcal{F}_1, \mathcal{F}_{II}, \mathcal{F}_{III})$ (cf. (3.4.1)).

**Proof.** Let $y$ be a continuity point of $F_n$ and put $y_n := a_n y + b_n$. Substituting (3.4.21) into (3.4.5) we obtain

$$
(3.4.28) \quad F_n(y_n | x) =
$$

$$
T_{\mathcal{S}}(x) \left( \sum_{k=0}^{d-1} \lambda_0 \rho_{n} \left( y_n \right) \rho_k \left( y_n \right) \right) + \sum_{k \in \mathcal{S}} \left( \lambda_k \left( y_n \right) \rho_k \left( y_n \right) + \lambda_k \left( y_n \right) \rho_k \left( y_n \right) \right) = F(y_n).
$$

Now either we have $\lim_{n \to \infty} G(y_n) = 1$ or there is a subsequence $y_{n_1}, y_{n_2}, \ldots$ with $\limsup_{k \to \infty} G(y_{n_k}) < 1$. In the first case lemma 3.4.5 applies and from (3.4.28) it follows that

$$
(3.4.29) \quad F(y) = F(y_n | x)(1 + o(1)) = \lambda_0 \rho_0 \left( y_n \right)(1 + o(1)) =
$$

$$
= \lambda_0 \rho_0 \left( y_n \right)(1 + o(1)) \quad (n \to \infty),
$$

just as (3.2.34) followed from (3.2.31). In the second case lemma 3.4.4 applies to the subsequence $y_{n_1}, y_{n_2}, \ldots$ and we have $\limsup_{k \to \infty} |\lambda_k \rho_k| < 1$ for all $k$, so that

$$
(3.4.30) \quad F(y) = \lim_{k \to \infty} F_k(y_n | x) = 0.
$$

We see that the second case can only occur if $F(y) = 0$, but even then relation (3.4.29) holds. This completes the proof of (3.4.27). It follows
from (3.4.28) that \( \lim_{n \to \infty} F_n(y_n | x) \) exists for all \( x \in S \) as soon as it exists for one value of \( x \). The limit depends on \( x \) only through the terms \( \tau \alpha(x) F_n(y_n) \beta(y_n) \), which by (3.4.23) have limit 1 (for \( \ell = 0 \)) or 0 (for \( \ell = 1, \ldots, d-1 \)), independently of \( x \).

It remains to prove the last assertion of the theorem. Let \( k \) be an arbitrary positive integer. By (3.4.27) we obtain

\[
\lambda_n^{(y)}(a_n x + b_n) \leq F(y)
\]

and hence

\[
\lambda_n^{(y)}(a_n x + b_n) \leq F^{1/k}(y).
\]

Obviously, \( \lim_{n \to \infty} G(a_n x + b_n) = 1 \). Replacing \( y_n \) in (3.4.28) by \( a_n y + b_n \) we find for all \( x \in S \)

\[
F_n(a_n y + b_n | x) \leq F^{1/k}(y) \quad \text{for all } k \geq 1.
\]

It follows by lemma 3.4.6 that there are constants \( p_k > 0 \) and \( q_k \) such that

\[
F^{1/k}(y) = F(p_k y + q_k) \quad \text{for all } k \geq 1.
\]

We can apply (3.4.2) now.

The following theorem establishes the role of the invariant distribution.

**Theorem 3.4.8**. Let \( X_0, X_1, \ldots \) be an irreducible chain of finite rank, with invariant distribution function \( F \). Suppose there are real constants \( a_n > 0 \) and \( b_n \) such that

\[
\lambda_n^{(y)}(a_n y + b_n) \leq F(y) \quad \text{as } n \to \infty,
\]

where \( F \) is a nondegenerate distribution function. Then also

\[
F_n(a_n y + b_n | x) \leq F(y) \quad \text{as } n \to \infty.
\]
Let $y_1, y_2, \ldots$ be a sequence such that $\lim_{n \to \infty} G(y_n) = 1$. For every fixed positive integer $k$ we then have for all $x \in S$

\begin{equation}
F_{n_k}(y_{n_k} | x) = F_{n_k}(y_{n_k} | x) (1 + o(1)) \quad (n \to \infty) .
\end{equation}

\[\square\]

PROOF. In the proof of theorem 3.6.7 we found, under the condition that

\begin{equation}
F_n(y_n | x) = \frac{\tau^2}{n} G(y_n)(1 + o(1)) \quad (n \to \infty) ;
\end{equation}

this is actually (3.4.29), only this time we do not know whether

\[\lim_{n \to \infty} F_n(y_n | x) \]

exists for at least one $x$. Now (3.4.37) follows immediately from (3.4.38), if we consider only the subsequence $y_{n_k}, y_{2n_k}, \ldots$. \[\square\]

**Lemma 3.4.10.** Suppose $y_1, y_2, \ldots$ is a sequence, such that for a certain $\tau > 0$ we have

\begin{equation}
1 - G(y_n) = \frac{\tau}{n} + o\left(\frac{1}{n}\right) \quad (n \to \infty) .
\end{equation}

Then for all $x \in S$

\begin{equation}
\lim_{n \to \infty} F_n(y_n | x) = e^{-\tau} .
\end{equation}

\[\square\]

PROOF. From (3.4.35) it follows that $\lim_{n \to \infty} G(y_n) = 1$ and hence

$\lim_{n \to \infty} \mathbb{E}(y_n) = 1$ by lemma 3.4.4. In the proof of theorem 3.1.2 we derived relation (3.1.18), which now yields

\begin{equation}
1 - \frac{1}{n} \sum_{k=1}^{n} \mathbb{P}(x) = \mathbb{E}(x) \mathbb{E}(x) + o\left(\frac{1}{n}\right) \quad (n \to \infty) .
\end{equation}

Take a fixed integer $k$. We obtain (first Bonferroni inequality)

\begin{equation}
F_n(y_{n_k} | x) = \mathbb{P}(X_1 \leq y_{n_k}, \ldots, X_n \leq y_{n_k} | X_n = x) \geq
\end{equation}

\begin{equation}
1 - \frac{1}{n} \sum_{k=1}^{n} \mathbb{P}(X_1 > y_{n_k} | X_n = x) =
\end{equation}
- 1 - n\left[ - 1 - \frac{1}{n} \sum_{i=1}^{n} p^{(i)}(\gamma_{nk} | x) \right] = \\
\approx 1 - n \{ 1 - C(\gamma_{nk}) + o(\frac{1}{n}) \} = 1 - \frac{1}{k} + o(1) \quad (n \to \infty).

On the other hand, we have (second Bonferroni inequality)

\begin{equation}
(3.4.43) \quad F_n(\gamma_{nk} | x) \leq 1 - \frac{n}{l} \sum_{i=1}^{n} P(X_i > \gamma_{nk} | X_0 = x) + \\
+ \sum_{j \neq k} P(X_j > \gamma_{nk}, X_i > \gamma_{nk} | X_0 = x),
\end{equation}

with

\begin{equation}
(3.4.44) \quad \sum_{j \neq k} P(X_j > \gamma_{nk}, X_i > \gamma_{nk} | X_0 = x) = \\
= \frac{2}{n} \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} P(X_j > \gamma_{nk}, X_i > \gamma_{nk} | X_0 = x) \leq \\
\leq \frac{2}{n} \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \left[ \frac{\Gamma_{\gamma}(x) \zeta_{n-1}^{(n)}}{d(y)} \right] \left[ \Gamma_{\gamma}(y) \frac{\zeta_{n-1}^{(n)}}{d(y)} \right] = \\
\leq 2n^2 \sum_{i=1}^{n-1} \left[ \frac{\Gamma_{\gamma}(x) \zeta_{n-1}^{(n)}}{d(y)} \right] \left[ \Gamma_{\gamma}(x) \frac{\zeta_{n-1}^{(n)}}{d(y)} \right] = \\
= 2n^2 \left[ 1 - C(\gamma_{nk}) + o(\frac{1}{n}) \right] = 2n^2 \left[ \frac{\gamma_{nk}}{k} + o(1) \right] = \\
= 2\left( \frac{\gamma_{nk}}{k} + o(1) \right)^2 = \frac{2\gamma_{nk}^2}{k^2} + o(1) \quad (n \to \infty).
\end{equation}

Substituting (3.4.42) and (3.4.44) into (3.4.43) we obtain

\begin{equation}
(3.4.45) \quad F_n(\gamma_{nk} | x) \leq 1 - \frac{1}{k} + \frac{2\gamma_{nk}^2}{k^2} + o(1) \quad (n \to \infty).
\end{equation}
We thus have found for all fixed $k$

$$(3.4.46) \quad 1 - \frac{\tau}{k} \leq \operatorname{linsup}_{n \to \infty} P_n(y_{nk}|x) \leq \operatorname{limsup}_{n \to \infty} P_n(y_{nk}|x) \leq 1 - \frac{\tau}{k} + \frac{2\tau^2}{k^2},$$

which is equivalent to

$$(3.4.47) \quad \left(1 - \frac{\tau}{k}\right)^k \leq \operatorname{linsup}_{n \to \infty} P_n^k(y_{nk}|x) \leq \operatorname{limsup}_{n \to \infty} P_n^k(y_{nk}|x) \leq \left(1 - \frac{\tau}{k} + \frac{2\tau^2}{k^2}\right)^k.$$

From this and Lemma 3.4.9 it follows that for all fixed $k$

$$(3.4.48) \quad \left(1 - \frac{\tau}{k}\right)^k \leq \operatorname{linsup}_{n \to \infty} P_n^k(y_{nk}|x) \leq \operatorname{limsup}_{n \to \infty} P_n^k(y_{nk}|x) \leq \left(1 - \frac{\tau}{k} + \frac{2\tau^2}{k^2}\right)^k.$$

We now first note, that in (3.4.42)-(3.4.48) we may replace $y_{nk}$ by $y_{nk+\ell_n}$, where $\ell_n$ is any sequence of integers with $\ell_n = o(n) (n \to \infty)$, because in that case

$$(3.4.49) \quad 1 - o(y_{nk+\ell_n}) = \frac{-\tau}{nk+\ell_n} + o\left(\frac{1}{nk+\ell_n}\right) =$$

$$= \frac{-\tau}{nk} \left(1 - \frac{\ell_n}{nk+\ell_n}\right) + o\left(\frac{1}{n(k+\ell_n)}\right) =$$

$$= \frac{-\tau}{nk} + o\left(\frac{1}{n}\right) \quad (n \to \infty),$$

so that all calculations remain the same. For each $n$ and fixed $k$ let $r_n$ and $\ell_n$ be determined by

$$(3.4.50) \quad n = r_nk + \ell_n, \quad \ell_n \in \{0, 1, \ldots, k-1\}.$$

Clearly $n \to \infty$ if $r_n \to \infty$ and $\ell_n = o(r_n)$, since even $\ell_n < k$ for all $n$. Therefore we obtain from (3.4.48)

$$(3.4.51) \quad \operatorname{linsup}_{n \to \infty} P_n(y_{nk}|x) \leq \operatorname{linsup}_{n \to \infty} P_{r_nk}^k(y_{nk+\ell_n}|x) =$$

$$= \operatorname{linsup}_{n \to \infty} P_{r_nk}^k(y_{r_nk+\ell_n}|x) \leq \left(1 - \frac{\tau}{k} + \frac{2\tau^2}{k^2}\right)^k \quad \text{for all } k,$$
and

\[
\liminf_{n \to \infty} F(n | x) \geq \liminf_{n \to \infty} F(n | x) = \frac{1}{k}
\]

\[
\geq \left( 1 - \frac{1}{k} \right) \quad \text{for all } k.
\]

Letting \( k \to \infty \) in (3.4.51) and (3.4.52) completes the proof of (3.4.50).

\[\text{PROOF of theorem 3.4.8. Let } y \text{ be a continuity point of } F \text{ with } F(y) > 0. \]

Putting again \( y_n := a_n y + b_n \), we obtain from (3.4.35) that

\[
(3.4.53) \quad 1 - G(y_n) = 1 - \frac{1}{\pi} \log F(y) (1 + O(\frac{1}{n})) =
\]

\[
= 1 - \exp \left( - \frac{1}{n} \log F(y) \right) + O(\frac{1}{n}) =
\]

\[
= - \frac{1}{n} \log F(y) + O(\frac{1}{n}) \quad (n \to \infty).
\]

With \( r = - \log F(y) \) the condition of lemma 3.4.10 is satisfied, hence

\[
(3.4.54) \quad \lim_{n \to \infty} F(n | x) = e^{r \log F(y)} = F(y) \quad \text{for all } x \in S.
\]

On account of the classical theory we know that \( F \in \{ F_1, F_2, F_3, F_4 \} \), hence \( F \)

is a continuous distribution function. By a monotonicity argument (3.4.54) can therefore be extended to \( y \) with \( F(y) = 0 \).

To finish this section we remark that the reverse of theorem 3.4.8,

which seems likely to hold, is hard to prove.

3.5. Renewal theory

If \( X_1, X_2, \ldots \) are nonnegative i.i.d. random variables, \( S_n := \sum_{k=1}^{n} X_k \)

and \( S_0 := 0 \), then the renewal function \( U \), defined by

\[
(3.5.1) \quad U(t) := \sum_{n=0}^{\infty} P(S_n \leq t) \quad \text{for } t \geq 0\,
\]

is the expected number of renewals in the interval \([0, t]\), i.e. with
(3.5.2) \( U(t) := \max \{ n \in \mathbb{N} \mid S_n \leq t \} \) for \( t > 0 \)

we have

(3.5.3) \( U(t) = E \mathbb{R}(t) \) for \( t \geq 0 \).

Blackwell's well-known renewal theorem asserts that, with \( \mu := E \mathbb{R} \leq \infty \) and \( 1/\mu := 0 \) if \( \mu = \infty \),

(3.5.4) \( \lim \limits_{t \to \infty} \left[ U(t) - U(t-h) \right] = \frac{h}{\mu} \) for every \( h \geq 0 \)

(if the common distribution of the \( X_k \) is nonlattice; otherwise only when \( h \) is a multiple of the span of the distribution).

This theorem is not easy to prove and things get even more complicated if one allows the \( X_k \) to be dependent. In his thesis, RUNNENBURG [28] proves the renewal theorem for a large class of Markov chains containing the chains of finite rank. Here we shall only give the elementary renewal theorem, which has an easy proof, similar to that of the central limit theorem in section 3.2. In this case the renewal function is

(3.5.5) \( U(t|x) := \sum \limits_{n=0}^{\infty} P(S_n \leq t \mid X_0 = x) = E \mathbb{R}(t \mid X_0 = x) \).

**Lemma 3.5.1.** If \( U(t|x) \) is the renewal function corresponding to an irreducible Markov chain of finite rank, then

(3.5.6) \( U(t|x) < \infty \) for all \( t \) with \( G(t) < 1 \).

**Proof.** With the notations of section 3.4 we have (for all \( x, t \geq 0 \))

(3.5.7) \( U(t|x) = \sum \limits_{n=0}^{\infty} P(S_n \leq t \mid X_0 = x) = \sum \limits_{n=0}^{\infty} P(X_n \leq t \mid X_0 = x) = \sum \limits_{n=0}^{\infty} F_n(t|x) \).

Now let \( G(t) < 1 \). Then \( |\lambda(t)| < 1 \) for every eigenvalue \( \lambda(t) \) of \( C(t) \), as a consequence of lemma 3.4.6 (take \( y_n = t \) for all \( n \)). Now from (3.4.28) it follows that there are constants \( c \) and \( c \) with \( c < 1 \), such that

(3.5.8) \( F_n(t|x) \leq ce^n \) for all \( n \in \mathbb{N} \).

Hence \( \sum \limits_{n=0}^{\infty} F_n(t|x) \leq \infty \), and by (3.5.7) also \( U(t|x) < \infty \).
**Theorem 3.5.2.** Let \( X_0, X_1, X_2, \ldots \) be an indecomposable Markov chain of finite rank with state space \( S = \{0, 1, \ldots, n\} \) and invariant distribution \( \pi \) which is not the degenerate distribution with mass 1 at 0. Then \( \nu = \int_0^\infty \pi(dx) \) satisfies \( 0 < \nu < \infty \) and the renewal function \( V(t|x) \) given by (3.5.9), satisfies

\[
\lim_{t \to \infty} \frac{1}{t} V(t|x) = \frac{1}{\nu} \text{ for all } x \in S,
\]

where \( 1/\nu := 0 < 1 \).

**Proof.** For the Laplace transform of \( S_n \) we have, similar to (3.2.9),

\[
\chi_n(t|x) := E(e^{-tS_n} | X_0 = x) =
\begin{cases}
1 & \text{for } n = 0, \\
\frac{\lambda(x)e^{-\lambda(x)\tau}}{\lambda(x) - \lambda'(x)} & \text{for } n \geq 1,
\end{cases}
\]

with \( \zeta(t) = (\zeta_{jk}(t)) \) given by

\[
\zeta_{jk}(t) := \int_0^\infty e^{-tN} a_k(x) \mathbf{b}_j(dx),
\]

and

\[
\tilde{\mathbf{b}}(t) := \int_0^\infty e^{-tN} \mathbf{b}(dx).
\]

The relation with \( \tilde{\mathbf{a}}(t) \) and \( \tilde{\mathbf{b}}(t) \) of section 3.2 is

\[
\tilde{\mathbf{a}}(t) = \tilde{\mathbf{b}}(t), \quad \tilde{\mathbf{b}}(t) = \tilde{\mathbf{b}}(it).
\]

Suppose \( \nu < \infty \). It is easily verified that the results about \( \tilde{\mathbf{a}}(t) \), its eigenvalues and its spectral decomposition, remain valid if \( t \) is replaced by \( it \), except possibly those concerning second derivatives. Since

\[
\chi_n(t|x) = \chi_n(it|x) = 1 + it\chi_n(0) + o(t) = 1 - \mu t + o(t) \quad (t \to 0)
\]

we have \(|\tilde{\mathbf{a}}(t)| < 1\) for \( t > 0 \) sufficiently small. Hence we obtain
\( (3.5.15) \quad w(t|x) := \int_0^t e^{-\tau x} \mathcal{U}(dt|x) = \)

\[ = \frac{1}{n} \sum_{n=0}^\infty \int_0^\infty \frac{1}{\sqrt{\pi}} e^{-\tau^2/4} \left( e^{-\lambda_n \tau} + \lambda_n \int_0^\tau e^{-\lambda_n \tau} \right) d\tau = \]

\[ = \frac{1}{t} \int_0^\infty \frac{1}{\sqrt{\pi \tau}} e^{-\lambda_n \tau} \left( 1 + \frac{1}{\tau} \right) + O(1) \frac{d-1}{\sum_{k=1}^{d-1}} \int_0^\tau \frac{1}{\sqrt{\tau}} e^{-\lambda_n \tau} \left( 1 + \frac{1}{\tau} \right) d\tau + O(1) \]

By a simple Tauberian theorem, see e.g. Fellner [8], theorem XIII 5.1, it follows that

\[ (3.5.16) \quad U(t|x) = \frac{1}{t} + o(t) \quad \text{for all } x \in S, \]

which is the same as (3.5.9).

Now suppose \( u = \infty \). Take an \( \epsilon > 0 \) and define \( X_n^\epsilon = X_n(1 + \epsilon X_n)^{-1} \) for all \( n \). It is easily verified that \( X_1^\epsilon, X_2^\epsilon, \ldots \) is a Markov chain of finite rank on \( [0, \infty) \), with an invariant distribution \( G^\epsilon \), and \( p^\epsilon := \int_0^\infty x G^\epsilon (dx) < \infty \) (note that the \( X_n^\epsilon \) are bounded by \( 1/\epsilon \)). Let \( U^\epsilon(t|x) \) be the corresponding renewal function. We have, since \( X_n^\epsilon \leq X_n \) for all \( n \),

\[ (3.5.17) \quad U(t|x) \leq U^\epsilon(t|x) = \frac{1}{p^\epsilon} + o(t) \quad \text{as } t \to \infty. \]

For \( \epsilon > 0 \) we find \( u^\epsilon = \infty \) and \( U(t|x) = o(t) \) as \( t \to \infty \), which completes the proof of (3.5.9).

As another aspect of renewal theory we briefly consider the following open question of Kingman [19]. Let \( (u_n) \) be a renewal sequence corresponding to a denumerable Markov chain \( X_0, X_1, X_2, \ldots \), i.e. for some \( x \in S \) one has

\[ (3.5.18) \quad u_n = P(X_n = x \mid X_0 = x) ; \]

for which sequences \( (u_n) \) can the chain be taken of finite rank?
The question seems to be quite hard to answer in general, but something can be said. If we are dealing with a chain of finite rank, then

\[ u_n = T_n(x) e^{\lambda n} B(x) , \]

and substituting the spectral decomposition of \( C \) we obtain

\[ u_n = T_n(x) E_0 B(x) + \sum_{t=1}^{d-1} T_n(x) \lambda B(x) + o(e^n) \quad (n \to \infty). \]

It follows that, if \( u_n \) converges to a limit \( u \) as \( n \to \infty \), then it does so exponentially fast. Hence, for example, the sequence \( u_n = 1/(n+1) \), which is completely monotone and thereby a renewal sequence, does not correspond to a Markov chain of finite rank.
CHAPTER 4. GENERALIZATIONS AND APPLICABILITY

KINGMAN [19], at the end of his paper, poses some open problems about chains of finite rank, one of which concerns the possibility of a generalization to a continuous time parameter. This question is considered in the first section, after we have generalized some of the results of chapter 3.

In the second section we investigate ways of approximating general Markov chains by chains of finite rank. Special attention is paid to correlation coefficients as a way of measuring the generality of chains of finite rank. The possibility of applying discrete dynamic programming techniques is discussed, briefly, in the final section.

4.1. Generalizations

4.1.1. Multidimensional central limit theorem

Suppose that for \( n = 0,1, \ldots \) we have a random vector \( \mathbf{X}(n) = (X_j(n), \ldots, X_\ell(n)) \), taking values in the space \( (S, \Sigma) \) with \( S = \mathbb{R}^k \) and \( \Sigma = \mathcal{B}(\mathbb{R}^k) \).

Further suppose that \( X(1), X(2), \ldots \) is a Markov chain of rank \( r \). We introduce

\[
S_j(n) := \sum_{k=0}^{n} X_j(k) \quad \text{for } j = 1, \ldots, \ell \text{ and } n = 0,1, \ldots,
\]

and for all \( \mathbf{x} \in S \) the characteristic function

\[
\varphi_n(\mathbf{t}| \mathbf{y}) := \mathbb{E}(e^{i \mathbf{t}^T S(n)} | S_0 = \mathbf{x}) \quad \text{for } \mathbf{t} \in \mathbb{R}^k
\]

of \( S(n) = (S_1(n), \ldots, S_\ell(n)) \), cf. (3.2.5).

The \( k \)-dimensional generalization of lemma 3.2.1 is

**LEMMA 4.1.1.** The characteristic function \( \varphi_n(\mathbf{t}| \mathbf{y}) \) is given by

\[
\varphi_n(\mathbf{t}| \mathbf{y}) = e^{i \mathbf{t}^T \mathbf{C}(\mathbf{y}) \mathbf{a}^T} \mathbf{a}(\mathbf{y})
\]

with

\[
\mathbf{a}(\mathbf{y}) = \mathbf{a}(\mathbf{y}) \mathbf{a}(\mathbf{y})^T
\]
\[ (4.1.6) \quad \xi_{t_k}(x) = \int_{\mathbb{R}^d} a_k(x)e^{i\sum \xi_j \partial_j} \beta_j(dx) = \int_{\mathbb{R}^d} \cdots \int_{\mathbb{R}^d} a_k(x_1, \ldots, x_d)e^{i(t_1x_1 + \cdots + t_dx_d)} \beta_j(dx_1, \ldots, dx_d) \]

and

\[ (4.1.5) \quad \xi_t(x) := \int_{\mathbb{R}^d} e^{i\sum \xi_j \partial_j} \beta_j(dx) = \int_{\mathbb{R}^d} \cdots \int_{\mathbb{R}^d} e^{i(t_1x_1 + \cdots + t_dx_d)} \beta_j(dx_1, \ldots, dx_d) \]

of (3.2.5) and (3.2.7).

\[ \Box \]

**Proof.** One easily verifies that the proof of lemma 3.2.1 can be extended to the \( \lambda \)-dimensional case without any difficulty.

Not only lemma 3.2.1, but all results of section 3.2 preceding theorem 3.2.9 can easily be extended to the \( \lambda \)-dimensional situation. We note that we now have to consider the derivatives \( \partial^2/\partial t_j^2 \) and \( \partial^2/\partial t_j \partial t_k \) as well as \( \partial^2/\partial t_j \partial t_k \) for \( j, k = 1, \ldots, \lambda \). We shall extend the notation for the eigenvalues and related quantities introduced in the \( 1 \)-dimensional case in an obvious way, without further comment.

We now state the \( \lambda \)-dimensional central limit theorem.

**Theorem 4.1.2.** Let \( \mathcal{X}_0, \mathcal{X}_1, \ldots \) be an indecomposable Markov chain of rank \( \lambda \), with state space \( S = \mathbb{R}^\lambda \), and satisfying

\[ (4.1.6) \quad \int_{\mathbb{R}^\lambda} |\mathcal{X}|^2 \beta(dx) = \int_{\mathbb{R}^\lambda} \cdots \int_{\mathbb{R}^\lambda} (y_1^2 + \cdots + y_{\lambda}^2) \beta(dy_1, \ldots, dy_{\lambda}) < \infty. \]

Put

\[ \nu_j := 1 - \frac{\lambda}{\lambda \xi_j} \tilde{\chi}_0(0, \ldots, 0), \]

\[ \nu_j := \left( \frac{\lambda}{\lambda \xi_j} \tilde{\chi}_0(0, \ldots, 0) \right)^2 - \frac{\lambda}{\lambda \xi_j} \tilde{\chi}_0(0, \ldots, 0) \]
\[ \tau_{jk} := \frac{2}{\beta t_j} \gamma_0(0, \ldots, 0) - \frac{2}{\beta t_k} \gamma_0(0, \ldots, 0) \]

Then the \( \nu_j \) and the \( \tau_{jk} \) are real constants, the \( \nu_j \) are nonnegative, and \( \eta_n(\epsilon_k) \), defined by (4.1.2), satisfies the relation

\[ \lim_{n \to \infty} e^{-\frac{\epsilon^2}{2n}} \eta_n(\frac{\epsilon}{\sqrt{n}}) = e^{-\frac{\epsilon^2}{2}} \sum_{j=1}^\infty \nu_j e^{-\frac{\epsilon^2}{2j}} \sum_{jk \neq k} \tau_{jk} e^{-\frac{\epsilon^2}{2j}}. \]

**PROOF.** The first part of the proof of theorem 3.2.9 can easily be extended to the \( d \)-dimensional case to yield relation (4.1.7). The proof of the assertions about the constants \( \nu_j \), \( \nu_j \) and \( \tau_{jk} \) also is an extension of the corresponding part of the proof of theorem 3.2.9, but on this point we give part of the details.

Considering vectors of the form \( \bar{\gamma}(0, \ldots, 0, \epsilon_j, 0, \ldots, 0) \), i.e. considering the \( j \)-th "marginal" characteristic function, reduces the problem to the 1-dimensional one of section 3.2. From theorem 3.2.9 it immediately follows that \( \nu_j \) is real and that \( \nu_j \) is nonnegative. Since now the right-hand side of (4.1.7) is the limit of a sequence of characteristic functions, continuous at \( \epsilon = 0 \), it is itself a characteristic function, say of the random vector \((Y_1, \ldots, Y_d)\). Taking the second order partial derivative \( \frac{\partial^2}{\partial t_j \partial t_k} \) for \( j \neq k \) at the point \( \epsilon = 0 \) we obtain

\[ \tau_{jk} = E Y_j Y_k = \text{cov}(Y_j, Y_k), \]

hence also \( \tau_{jk} \) is real.

Although the sequence \( X_j(1), X_j(2), \ldots \) for a fixed \( j \in \{1, \ldots, d\} \) is in general not necessarily a Markov chain, the marginal characteristic function satisfies the essential relations of the form (3.2.9). Hence the results of section 3.3 apply to \( \nu_j \) and \( \nu_j \). As the chain \( X(1), X(2), \ldots \) is assumed to be indistinguishable, it has an invariant (multivariate) distribution \( \eta \); the \( j \)-th marginal of \( \eta \) is denoted by \( \eta_j \). The following theorem is the "translation" of theorems 3.3.1 and 3.3.2. We now state it without further proof.

**THEOREM 4.1.3.** The relations

\[ \frac{1}{n} \sum_{j} \nu_j(x) \to \nu_j \ast \delta(x) \quad (n \to \infty) \]
and

\[(4.1.10) \quad \frac{1}{n} \var_{\mathcal{R}} S_j(n) = \nu_j + \mathcal{O}(1) \quad (n \to \infty)\]

hold for all \( j \in \{1, \ldots, d\} \) and \( \mathcal{R} \in \mathbb{R}^d \), while

\[(4.1.11) \quad \nu_j = \int_{\mathbb{R}^d} y g_j(dy) .\]

\[\square\]

The only new aspect of the \( d \)-dimensional case is the occurrence of covariance terms \( \pi_{jk} \). As the analogue to theorem 3.3.2 and (4.1.10) we have

**Theorem 4.1.4.** The relation

\[(4.1.12) \quad \frac{1}{n} \var (S_j(n), S_k(n)) = \pi_{jk} + \mathcal{O}(1) \quad (n \to \infty)\]

holds for all \( j \neq k \) and \( n \in \mathbb{N}^d \).

\[\square\]

**Proof.** It is no restriction to assume that \( d = 2 \) (consider the appropriate 2-dimensional marginal, cf. the remarks preceding theorem 4.1.3). On the one hand we have

\[(4.1.13) \quad \frac{d}{\delta \tau_1^2 \delta \tau_2^2} \varphi_n(0,0, | x_1, x_2) = \varphi_\mathcal{R} S_j(n) S_k(n) ,\]

and on the other hand (cf. (3.2.29), (3.3.5), (3.3.6) and (3.3.7))

\[(4.1.14) \quad \frac{d^2}{\delta \tau_1^2 \delta \tau_2^2} \varphi_n(0,0, | x_1, x_2) = (n-1)(n-2) \lambda_0 \nabla(1)(2) + (n-1) \lambda_0 \nabla(1)(2)^2 +\]

\[+ (n-1) \sum_{k=0}^{d-1} \lambda_k \nabla(1,k) \nabla(2,k) + \lambda_k \nabla(1,k)^2 \varphi_\mathcal{R} \nabla(1,k) + \lambda_k \nabla(2,k)^2 \varphi_\mathcal{R} \nabla(2,k) (n \to \infty) ,\]

where we used that \( \varphi_\mathcal{R}(0,0) = 1 \) for \( k = 0 \) and \( \varphi_\mathcal{R}(0,0) = 0 \) for \( k \neq 0 \), and where we have written \( \lambda_k \) for \( \lambda_k(0,0) \), \( \gamma_k(1) \) for \( \gamma_k(0,0) \), \( \gamma_k(2) \) for \( \gamma_k(0,0) \), and similar abbreviations for \( \varphi_\mathcal{R} \). Combining (4.1.13) and (4.1.14) we obtain (cf. (3.3.8)).
\[ (4.1.15) \quad - \sum S_1(n) \sum S_2(n) = \text{cov} \{ S_1(n), S_2(n) \} = \]
\[ = (n-1) \gamma_{0}^{(1,2)} + (n-1) (n-2) \gamma_{0}^{(1)} \gamma_{0}^{(2)} + \]
\[ + \sum_{k=0}^{n-2} \left\{ (n-k+1) \left[ \gamma_{0}^{(1)} - \gamma_{0}^{(2)} (n-k+1) \right] \left[ \gamma_{0}^{(2)} - \gamma_{0}^{(1)} (n-k+1) \right] \right\} + o(1) \quad (n \to \infty). \]

Relation (3.2.39) reads in the present context (for \( j = 1, 2 \))
\[ (4.1.15) \quad i \sum S_j(n) = (n-1) \gamma_{0}^{(j)} + w_{0}^{(j)} + \sum_{k=1}^{n-1} \gamma_{k}^{(j)} w_{k}^{(j)} + o(1) \quad (n \to \infty). \]

This yields, together with \( \tau := \gamma_{0}^{(1)} - \gamma_{0}^{(2)} = \gamma_{0}^{(1,2)}, \)
\[ (4.1.17) \quad \tau = \frac{1}{n-1} \sum \text{cov} \{ S_1(n), S_2(n) \} + \]
\[ + \sum_{k=1}^{n-2} \left\{ (n-k+1) \left[ \gamma_{0}^{(1)} - \gamma_{0}^{(2)} (n-k+1) \right] \left[ \gamma_{0}^{(2)} - \gamma_{0}^{(1)} (n-k+1) \right] \right\} + o(1) \quad (n \to \infty). \]

We have
\[ (4.1.18) \quad \text{cov} \{ S_1(n), S_2(n) \} = \sum \text{cov} \{ S_1(n-1), S_2(n-1) + S_2(n) \} + \]
\[ = \sum \text{cov} \{ S_1(n-1), S_2(n-1) \} + \sum \text{cov} \{ X_1(n), X_2(n) \} + \]
\[ + \sum_{k=1}^{n-1} \left( \text{cov} \{ X_1(k), X_2(n) \} + \text{cov} \{ X_1(n), X_2(k) \} \right). \]

Similar to (3.3.11) it is proved that \( \text{cov} \{ X_1(n), X_2(n) \} \) is bounded for \( n \in \mathbb{N}, \) and similar to (3.3.12), (3.3.13) and (3.3.15) it is proved that both \( \sum_{k=1}^{n-1} \text{cov} \{ X_1(k), X_2(n) \} \) and \( \sum_{k=1}^{n-1} \text{cov} \{ X_1(n), X_2(k) \} \) are bounded for \( n \in \mathbb{N}; \) in this instance integrals appear of the form \( \int f(y_1) dy_1 dy_2, \)
but this is not an essential difference. So, analogous to (3.3.16), we obtain
\[ (4.1.19) \quad \text{cov} \{ S_1(n), S_2(n) \} = \text{cov} \{ S_1(n-1), S_2(n-1) \} + o(1) \quad (n \to \infty). \]

It follows that the sum in the right-hand side of (4.1.17) is constant and hence equal to 0. This completes the proof. \( \Box \)
Another generalization of the central limit theorem for Markov chains of finite rank we formulate, combining the results in sections 3.2 and 3.3.

**Theorem 4.1.5.** Let $X_0, X_1, \ldots$ be an indecomposable Markov chain of finite rank and $h: S \to \mathbb{R}$ a measurable function satisfying

\[ \int_S |h(y)|^2 \, \mu(dy) = \infty. \tag{4.1.20} \]

Then, for all $x \in S$,

\[ \frac{\frac{1}{n} \left[ h(X_k) - \mathbb{E}_x h(X_k) \right]}{\left( \operatorname{var}_x \frac{1}{n} h(X_k) \right)^{1/2}} \] \[ \quad \xrightarrow{\text{as} \ n \to \infty} \quad \text{standard normal.} \tag{4.1.21} \]

**Proof.** It is easily verified that all assertions in sections 3.2 and 3.3, concerning the characteristic functions and the related quantities, remain valid when $e^{itx}$ is replaced by $e^{ith(x)}$.

**Example 4.1.6.** Define for a fixed set $A \subset S$ the function $h$ by

\[ h(x) = 1_A(x). \tag{4.1.22} \]

Then $\sum_{k=1}^n h(X_k)$ can be interpreted as the number of visits to the set $A$ after $n$ transitions. From theorem 4.1.5 it follows that this number is asymptotically normal.

**4.1.2. Continuous-time analogue**

If we assume that the kernel matrix $C$ is nonsingular, and define $\tilde{T}$ by

\[ \tilde{T} = T \cdot C^{-1}, \tag{4.1.23} \]

we can write (1.2.6) as

\[ P^{(n)}(A|B) = T^n_{B|A} C^n \mathbb{P}(A) \quad (n \geq 1) \tag{4.1.24} \]
(if $C^{-1}$ is not assumed to exist, then formulas become slightly more complicated). The difficulty in extending (4.1.24) to continuous time lies in the fact that (4.1.24) is not true for $n=0$: $I_0(x) \neq \frac{1}{n!} p^{(n)}(x) C^{-1} B(A)$. Still it is possible to obtain a kind of analogue as follows. Let the continuous-time transition $P_t$ be given by

\begin{equation}
(4.1.25) \quad P_t(A|x) = \sum_{n=0}^{\infty} e^{-\alpha t} \frac{(\alpha t)^n}{n!} p^{(n)}(x) \chi^n A \quad (t \geq 0),
\end{equation}

where $p^{(n)}$ satisfies (4.1.24). This means that $P_t$ describes a pseudo-Poisson process as considered, e.g., in FELLER [8], with an underlying process of finite rank.

Using (4.1.24) we can write (4.1.25) as

\begin{equation}
(4.1.26) \quad P_t(A|x) = e^{-\alpha t} I_0(x) + \sum_{n=1}^{\infty} e^{-\alpha t} \frac{(\alpha t)^n}{n!} T^n A(x) C^n B(A) =
\end{equation}

\begin{equation}
= e^{-\alpha t} I_0(x) + T(x) e^{-\alpha t} (e^{\alpha t C} - 1) B(A).
\end{equation}

Now for simplicity assume that $C$ has $r$ distinct eigenvalues $\lambda_0 = 1, \lambda_1, \ldots, \lambda_{r-1}$. Substituting the spectral decomposition of $C$ into the first line of (4.1.26) we obtain

\begin{equation}
(4.1.27) \quad P_t(A|x) = e^{-\alpha t} I_0(x) + T(x) e^{-\alpha t} \sum_{\lambda=0}^{r-1} \left( e^{\alpha t \lambda} - 1 \right) E_{\lambda} B(A),
\end{equation}

where we have used $E_{0=0} E_{\lambda} = I$ (cf. (4.1.5)).

From (4.1.26) and (4.1.27) it is clear that $P_t$ can be computed with the same ease as $P^{(k)}$ in (4.1.24). From (4.1.27), for example, it is clear that

\begin{equation}
(4.1.28) \quad \lim_{t \to \infty} P_t(A|x) = T(x) E_0 B(A) = G(A),
\end{equation}

as in the discrete-time case, cf. theorem 3.1.3.

Considering $P_t$ as an operator, with

\begin{equation}
(4.1.29) \quad (P_t f)(x) = \int f(y) P_t(dy|x),
\end{equation}

for the infinitesimal operator $G$ of $P_t$ we find
\[ (4.1.30) \quad Gf = \lim_{t \to 0} \frac{P_t f - f}{t} = a(t \gamma) f = a(x) \sum_{\ell=1}^{r-1} \left( \lambda_\ell - 1 \right) E_{\ell} \int f(y) B(dy). \]

Generally, if a transition kernel \( P_t \) is "of finite rank", i.e. if it has the form
\[ (4.1.31) \quad P_t(X|\omega) = \lambda(t) a(x) + \gamma a(x) \gamma B(A), \]
then, assuming differentiability, the generator \( G \) takes the form
\[ (4.1.32) \quad Gf = a'(0) f + \gamma \gamma C'(0) \mathbb{B}[f] \]
with \( \mathbb{B}[f] \) as in (1.4.3). It follows that diffusion processes, which have differential operators as generators (see e.g. BREEDEN [11]) cannot be of finite rank in this way. However, consistent with the fact that diffusion processes can be obtained as limits of pseudo-Poisson processes, diffusion kernels can sometimes be approximated by Markov chains of finite rank.

As an example we consider the Ornstein-Uhlenbeck process\(^*\), where the transition from \((t_1, x)\) to \((t_2, y)\) is governed by the normal distribution with expectation \( \gamma x \exp[-\lambda(t_2 - t_1)] \) and variance \( \sigma^2 (1 - \exp[-2\lambda(t_2 - t_1)]) \), cf. FELLER [8] and CRAMER [2]. If for simplicity we choose \( \lambda = \sigma = 1 \), the transition density is
\[ (4.1.33) \quad P_t(y|x) = \frac{1}{\sqrt{2\pi (t_1 - e^{-2t})}} \exp \left[ -\frac{1}{2(t_1 - e^{-2t})} (y - x e^{-t})^2 \right] = \frac{1}{\sqrt{2\pi}} \sum_{k=0}^{\infty} \frac{1}{k!} H_k(x) e^{-yt/2}, \]
where \( H_k \) denotes the \( k \)-th Hermite polynomial, defined by
\[ (4.1.34) \quad H_k(x) := (-1)^k e^{x^2/2} \frac{d^k}{dx^k} (e^{-x^2/2}) \quad \text{for } k = 0, 1, 2, \ldots. \]
We remark that \( H_k(x) \) is an \( e^{-x^2} \)-eigenfunction of \( P_t(y|x) \), i.e.
\[ (4.1.35) \quad \int_{\mathbb{R}} H_k(y) P_t(y|x) dy = e^{-kt} H_k(x) \quad \text{for all } x \in \mathbb{R} \text{ and } k \geq 0. \]

Apparently, \( P_t(y|x) \) can be approximated by a finite sum of products (for \( t = 0 \) the density has to be interpreted as a \( \delta \)-function), hence the

\(^*\) This was suggested to us by prof. J. Keilson.
corresponding kernel can be approximated by an expression similar to (6.1.26). This is, of course, not an easy approximation, if only because the first r terms in (4.1.33) do not sum to a nonnegative function.

4.2. Approximation by kernels of lower rank

As Markov chains of low rank are easy to handle, the question naturally arises to what extent more general Markov chains can be approximated by chains of finite rank, and how chains of high finite rank can be approximated by chains of lower rank. General answers are very hard to come by, so we shall mainly suggest possibilities, point out difficulties, and treat some special cases and examples; only few theorems will be given.

In the first subsection we consider approximation under preservation of the invariant distribution, in the second one the optimization of correlation coefficients.

4.2.1. Approximation and invariant distribution

How well can a transition kernel $P(A|x)$, together with its iterates $P^{(n)}(A|x)$, be approximated by a kernel $P_r(A|x)$ of rank r? This general question is too hard to answer.

If $P$ itself is a kernel of finite rank $N$ with distinct eigenvalues one may proceed as follows. Let $U$ be such that (cf. (A.1.19))

\begin{equation}
U^{-1} \mathbf{c} \mathbf{u} = \begin{bmatrix}
\lambda_1 & \cdots & 0 \\
0 & \ddots & \vdots \\
0 & \cdots & \lambda_N
\end{bmatrix} =: \mathbf{c}^\pi.
\end{equation}

Then putting

\begin{equation}
\bar{P}(x) := T_{\mathbf{c}}(x)U, \quad \bar{c}(y) = U^{-1} \mathbf{c}(y)
\end{equation}

we obtain

\begin{equation}
P^{(n+1)}(A|x) = T_{\mathbf{c}}(x)U(c^\pi)^nU^{-1} \mathbf{c}(A) = \sum_{j=1}^{N} \lambda_j^n \mathbf{c}_j(x) \mathbf{c}(A).
\end{equation}

Assume, without restriction, the the $\lambda_j$ are ordered according to decreasing absolute value. We may retain the r largest $\lambda_j$'s and approximate $P$ by
\[ (4.2.4) \quad P_r(A|x) = \sum_{j=1}^{\infty} \lambda_j \theta_j(x) \xi_j(A). \]

As by (4.2.1) and (4.2.2) \( \theta_j \) is orthogonal to \( \xi_k \) in the sense that
\[ (4.2.5) \quad \int \theta_j(x) \xi_k(d\lambda) = \lambda_j \delta_{jk}, \]

(cf. (A.1.2)) we also have
\[ (4.2.6) \quad P_r^{(n+1)}(A|x) = \sum_{j=1}^{\infty} \lambda_j^n \theta_j(x) \xi_j(A) \quad \text{for} \ n \geq 0. \]

In particular, if all \( \lambda_j \) of modulus 1 are retained, then
\[ (4.2.7) \quad P_r^{(n+1)}(A|x) = P_r^{(n+1)}(A|x) + o(e^n) \quad (n \to \infty) \]

for some \( \epsilon \) with \( 0 < \epsilon < 1 \) (cf. Lemma 2.1.8), and the invariant distribution (see section 3.1) is preserved.

More general kernels \( P \) with a discrete spectrum (compare the symmetric Fredholm kernels as discussed e.g. in Tricomi [31]) might be approximated in a similar way. However, although (4.2.4) and (4.2.6) might provide reasonable approximations, there is the problem that in general \( P_r \) will not be a transition kernel; it may be negative if not complex.

Now let \( P \) be any transition kernel with invariant distribution \( G \).

Trivially, there always exists a chain of finite rank having \( G \) as its invariant distribution: a sequence of independent random variables with common distribution \( G \) is a chain of rank \( r = 1 \) (see section 1.3). But for any prescribed rank \( r \geq 1 \) a chain with invariant distribution \( G \) can be obtained, as is shown in the following.

**Theorem 4.2.1.** For every distribution function \( G \) and for every \( r \geq 1 \) there is a Markov chain of (exact) rank \( r \), with respect to which \( G \) is invariant.

Furthermore, the transition kernel \( P_r(A|x) \) of this Markov chain can be chosen such that
\[ (4.2.8) \quad P_r(A|x) = \sum_{j=1}^{\infty} \theta_j(x) \xi_j(A) \]

the \( \theta_j \) are nonnegative. \( \square \)
PROOF. Let $S_1, S_2, \ldots, S_r$ be a measurable partition of the state space $\mathbb{R}$ with $S_j \neq \emptyset$ for all $j$. Then $G$ can be written as

$$
(4.2.9) \quad G(A) = \frac{1}{r} \sum_{j=1}^{r} G(S_j) \mathbb{B}_j(A),
$$

where

$$
(4.2.10) \quad \mathbb{B}_j(A) := \frac{G(A \cap S_j)}{G(S_j)} \quad \text{if } G(S_j) > 0,
$$

and $\mathbb{B}_j$ is an arbitrary probability measure on $S_j$ if $G(S_j) = 0$. Clearly, $\mathbb{B}_1, \ldots, \mathbb{B}_r$ are linearly independent. We define linearly independent $a_1, \ldots, a_r$ by

$$
(4.2.11) \quad a_j(x) := 1_{S_j}(x).
$$

It is easily verified that the kernel $P_{\varepsilon}(A|x)$ in $(4.2.8)$ with $a_j(x)$ and $\mathbb{B}_j(A)$ as defined above satisfies

$$
(4.2.12) \quad G(A) = \int_{\mathbb{R}} P_{\varepsilon}(A|x) G(dx),
$$

and clearly the $a_j$ are nonnegative. \[\]

REMARK 4.2.2. If $G$ has at least $r$ points of increase, then we may choose $S_1, \ldots, S_r$ such that $G(S_j) > 0$ for all $j$. Generally our Markov chain can be defined on any state space having at least $r$ points and containing $\text{supp}(G)$. \[\]

COROLLARY 4.2.3. For every $N$-state Markov chain on $\{s_1, \ldots, s_N\}$ with invariant distribution $G$ and every $r < N$ there exists a Markov chain of rank $r$ on $\{s_1, \ldots, s_N\}$ with invariant distribution $G$. \[\]

REMARK 4.2.4. In the corollary above $N$ may be countably infinite. \[\]

REMARK 4.2.5. In the proof of theorem 4.2.1 a special choice is made for the $a_j$ and the $\mathbb{B}_j$. In most cases there will be a large amount of freedom in choosing the $a_j$ and $\mathbb{B}_j$. This freedom may be used to satisfy other requirements than preservation of an invariant distribution. In the next subsection the optimization of correlation coefficients is considered for $r = 2$ in a slightly different context. \[\]
Every kernel of the form

\[ P(A|x) = \sum_{j=1}^{\infty} a_j(x) B_j(A) \]

with nonnegative \( a_j \) and probability measures \( B_j \) can be approximated by just retaining \( r \) terms and renormalizing. Take for example the Markov chain on \((0,1)\) with kernel

\[ P(y|x) = 2\pi \left( \frac{2-x}{2-xy} \right) \sum_{j=0}^{\infty} \left( \frac{x}{y} \right)^j y^{2r+1} \text{ for } x, y \in (0,1) . \]

We obtain the approximations

\[ P_r(y|x) = \left( 1 - \frac{x}{y} \right) \sum_{j=0}^{r-1} \left( \frac{x}{y} \right)^j y^{2j+1} c_r(x) , \]

with

\[ c_r(x) = \left( 1 - \frac{x}{y} \right)^{r-1} . \]

In general, invariant distributions will not be preserved and little can be said about the quality of approximation, but in practice it should not be too hard to find a reasonable approximation of rank 2 or 3, say, to a given kernel of nonfinite rank. Of course, one always has the possibility of discretizing the state space and to obtain a finite Markov chain approximation, but this would probably lead to approximations of higher dimension, i.e. involving larger matrices.

### 4.2.2. Correlation coefficients

Let \( F(x,y) = P(X \leq x, Y \leq y) \) be any bivariate distribution function with marginal distribution functions \( G_1(x) = F(x,\cdot) \) and \( G_2(y) = F(\cdot,y) \).

If \( \mu \) is a measure on \((\mathbb{R},\mathcal{B}(\mathbb{R}))\) such that \( P(x,y) \) for each \( y \) and \( G_1(x) \) are absolutely continuous with respect to \( \mu \), with respective derivatives \( f_1(x,y) \) and \( g_1(x) \), then through

\[ P(y|x) := P(Y \leq y | X = x) = \frac{f_1(x,y)}{g_1(x)} \text{ for } g_1(x) \neq 0 \]

a transition kernel is obtained for a stationary Markov chain with state space \( \mathbb{R} \). It is clear from (4.2.17) that this chain is of finite rank if
and only if $F(x,y)$ is of the form

$$(4.2.18) \quad F(x,y) = \sum_{j=1}^{k} f_j(x) \beta_j(y).$$

On the other hand, if a chain with kernel $P(y|x)$ has an invariant distribution $G(y)$, then a bivariate distribution function $F(x,y)$ can be defined by

$$(4.2.19) \quad F(x,y) := \int_{-\infty}^{x} \int_{-\infty}^{y} P(y|x) G(dx) G(dy).$$

In this case both marginals are equal to the invariant distribution $G$.

We consider the class $F_r(G)$ of all bivariate functions of the form

$$(4.2.18)$$

with both marginals equal to $G$. As in the case of kernels of finite rank we may assume that $r$ in (4.2.18) is minimal, i.e. that the $f_j(x)$ and the $\beta_j(y)$ are linearly independent, and that the $\beta_j$ are distribution functions.

The case $r = 1$ is trivial: $F_1(G)$ only contains $F(x,y) = G(x)G(y)$, the distribution of two independent, identically distributed random variables.

For $r \geq 2$ one may ask for those $F(x,y) \in F_r(G)$ yielding a smallest or largest correlation coefficient, in order to get an impression of the restrictiveness of the assumption $F(x,y) \in F_r(G)$. This problem has been solved by RENNER and STEUTEL [30] for $r = 2$; we give an account of their results. The problem for $r \geq 3$ seems hard to tackle, but an example will be given at the end of this section.

First we make some general remarks. If $F(x,y)$ is any bivariate distribution function with marginals $G_1$ and $G_2$, then according to FRECHET [11]

$$(4.2.20) \quad \max(G_1(x) + G_2(y) - 1,0) \leq F(x,y) \leq \min(G_1(x), G_2(y)),$$

where the bounding functions are distribution functions. For the covariance, if it exists, one can easily prove, using Fubini's theorem, that

$$(4.2.21) \quad \text{cov}(X,Y) = \iint xy F(dx,dy) - \int x G_1(dx) \int y G_2(dy) =$$

$$= \iint (F(x,y) - G_1(x)G_2(y))dx dy.$$

From (4.2.21) and (4.2.20) it is seen that among all bivariate distributions with marginals $G_1$ and $G_2$ the functions $\max(G_1(x) + G_2(y) - 1,0)$ and
\( \min(G_1(x), G_2(y)) \) yield the smallest and the largest correlation coefficient, respectively. However,

\[
\begin{align*}
H_0(x,y) &:= \max(G(x) * G(y) - 1, 0) \\
K_0(x,y) &:= \min(G(x), G(y))
\end{align*}
\]
do not belong to \( \mathcal{F}_r(G) \), so this does not solve our problem. Moreover, functions \( H_{\alpha,\tau}(x,y) \in \mathcal{F}_r(G) \) and \( K_{\alpha,\tau}(x,y) \in \mathcal{F}_r(G) \) such that

\[
(4.2.23) \quad H_{\alpha,\tau}(x,y) \leq \mathcal{F}(x,y) \leq K_{\alpha,\tau}(x,y) \quad \text{for all } \mathcal{F}(x,y) \in \mathcal{F}_r(G).
\]
do not exist, as one can easily verify.

Now let \( r = 2 \) and let the kernel \( P(y|x) \) be given by

\[
(4.2.24) \quad P(y|x) = a_1(x) B_1(y) + (1 - a_1(x)) B_2(y) = B_2(y) + a_1(x)(B_1(y) - B_2(y)).
\]

If an invariant distribution \( G(y) \) exists, it equals

\[
(4.2.25) \quad G(y) = \gamma B_1(y) + (1 - \gamma) B_2(y) = B_2(y) + \gamma(B_1(y) - B_2(y))
\]

and we can write

\[
(4.2.26) \quad P(y|x) = G(y) + \langle a_1(x) - \gamma \rangle (B_1(y) - B_2(y))
\]

so we have

**Lemma 4.2.6.** If \( P(y|x) \) is a kernel of rank 2 with invariant distribution \( G \), then

\[
(4.2.27) \quad P(y|x) = G(y) + p(x) q(y),
\]

where

\[
(4.2.28) \quad q(\omega) = q(\omega) = 0, \quad q(y) \neq 0
\]

and

\[
(4.2.29) \quad \int p(x) G(dx) = 0.
\]

\[ \square \]
**Proof.** With \( p(x) := a_1(x) - y \) and \( q(y) := b_1(y) - b_2(y) \) relation (4.2.27) is the same as (4.2.26). The assertions about \( q \) are trivial, and (4.2.29) follows from

\[
(4.2.30) \quad G(y) = \int P(y|x)G(dx) = \int (G(y) + p(x)q(y))G(dx) =
\]

\[
= G(y) + q(y) \int p(x)G(dx).
\]

From here on we work under the following

**Assumption.** The distribution function \( G \) is continuous and has a finite second moment.

For the covariance we find, using (4.2.27),

\[
(4.2.31) \quad \text{cov}(X,Y) = \int \int xy P(dy|x)G(dx) - \left( \int xG(dx) \right)^2 =
\]

\[
= \int xp(x)G(dx) \int yq(dy).
\]

The problem can now be formulated as follows: for a given \( G \) find functions \( p \) and \( q \) satisfying (4.2.28) and (4.2.29) such that the product of

\[
(4.2.32) \quad I_1(p) := \int xp(x)G(dx)
\]

and

\[
(4.2.33) \quad I_2(q) := \int yq(dy)
\]

is extremal.

**Lemma 5.2.7.** Let \( q \) be a function satisfying (4.2.28). Then there are \( p_1 \geq 0 \) and \( p_2 \geq 0 \) such that \( P_p(y) := G(y) + pq(y) \) is a distribution function if and only if \( p \in [-p_1, p_2] \).

**Proof.** The only concern is the monotonicity of \( P_p(y) \). Suppose \( p_0 > 0 \); it is easily verified that \( P_{p_0}(y) \) is nondecreasing if and only if \( P_p(y) \) is nondecreasing for each \( p \) with \( 0 \leq p \leq p_0 \). A similar statement holds for \( p_0 < 0 \). Hence we find that \( P_p(y) \) is a distribution function if and only if \( p \) is an element of a closed interval. Clearly, this interval contains \( p = 0 \) and is bounded.
REMARK 4.2.8. If \( q(y) \) is such that in lemma 4.2.7 we find \( p_1 = 0 \), then in 
(4.2.27) we must have \( p(x) > 0 \) for all \( x \). Using (4.2.29) it then follows 
that \( I_1(p) = 0 \) (cf. (4.2.32)), hence the covariance (and the correlation 
coefficient) we obtain equals 0, just as in the rank 1 case. Similarly, 
\( p_2 = 0 \) leads to a zero covariance. Therefore it is no restriction to assume 
that \( q(y) \) yields \( p_1 > 0 \) and \( p_2 > 0 \). 

\[ \text{Lemma 4.2.9. For a fixed function } q(y), \text{ the function } p^*(x) \text{ satisfying} \]
(4.2.29), for which \( I_1(p^*) \) is maximal, is given by 
\[ p^*(x) = \begin{cases} 
-p_1 & \text{for } x \leq \xi \\
p_2 & \text{for } x > \xi 
\end{cases}, \]
where \( \xi \) satisfies 
\[ G(\xi) = \frac{p_2}{p_1 + p_2}. \]

\[ \text{Proof. By the previous lemma } \psi(p|x) = G(y) \ast p^*(x)q(y) \text{ is a distribution} \]
function for each \( x \). Now let \( p(x) \) be an arbitrary function satisfying 
(4.2.29) and such that \( G(y) \ast p(x)q(y) \) is a distribution function. Then by 
lemma 4.2.7 we have 
\[ -p_1 \leq p(x) \leq p_2 \quad \text{for all } x \in \mathbb{R} \]
hence 
\[ I_1(p^*) - I_1(p) = \int_{-\infty}^{\infty} x(p^*(x) - p(x))G(dx) = \]
\[ = -\int_{-\infty}^{\xi} x(p(x) - p^*(x))G(dx) + \int_{\xi}^{\infty} x(p^*(x) - p(x))G(dx) \geq \]
\[ \geq \xi \int_{-\infty}^{\xi} (p(x) - p^*(x))G(dx) + \xi \int_{\xi}^{\infty} (p^*(x) - p(x))G(dx) = 0 \]
by (4.2.29), from which it also follows that
\[(4.2.38) \quad -p_1 G(\xi) + p_2 (1 - G(\xi)) = 0,\]

which yields \((4.2.35)\).

Remark 4.2.10. By the assumed continuity of \(G\) the existence of at least one \(\xi\) satisfying \((4.2.35)\) is guaranteed; of course, \(\xi\) need not be uniquely determined by \((4.2.35)\).

In lemma 4.2.9 the function \(q(y)\) determines the values of \(p_1\) and \(p_2\), and nothing else. We now reverse the situation, in order to find an optimal \(q(y)\) where \(p(x)\) is a fixed function.

**Lemma 4.2.11.** Let \(p^*(x)\) and \(\xi\) be given by \((4.2.34)\) and \((4.2.35)\), with \(p_1, p_2 > 0\). Let \(Q(p_1, p_2)\) be the set of all functions \(q(y)\) that in lemma 4.2.7 yield the interval \([-p_1, p_2]\). Define

\[
q^*(y) := \min \left( \frac{1 - G(y)}{p_1}, \frac{G(y)}{p_2} \right),
\]

\[
q_\xi(y) := \min \left( \frac{1 - G(y)}{p_2}, \frac{G(y)}{p_1} \right).
\]

Then \(q^*(y) \in Q(p_1, p_2)\), \(q_\xi(y) \in Q(p_1, p_2)\), \(q^*\) maximizes \(I_2(q)\) over \(Q(p_1, p_2)\) and \(q_\xi\) minimizes \(I_2(q)\) over \(Q(p_1, p_2)\).

**Proof.** Since \(G(y) - p_1 q(y)\) and \(G(y) + p_2 q(y)\) are distribution functions we have

\[(4.2.40) \quad 0 \leq G(y) - p_1 q(y) \leq 1, \quad 0 \leq G(y) + p_2 q(y) \leq 1\]

and hence

\[(4.2.41) \quad q^*(y) \leq q(y) \leq q_\xi(y)\]

It is easily verified that \(q^*(y), q_\xi(y) \in Q(p_1, p_2)\). Furthermore,

\[
(4.2.42) \quad I_2(q) = \int q(dy) - \int_0 q(y) dy + \int_0 (-q(y)) dy = -\int q(y) dy.
\]

From \((4.2.41)\) and \((4.2.42)\) it is clear that \(I_2(q)\) is maximal for \(q = q^*\) and minimal for \(q = q_\xi\).
The main result of this subsection is

**Theorem 4.2.12.** Let \( G \) be a continuous distribution function with expectation 0 and variance 1. The largest correlation coefficient (if it exists) yielded by a bivariate distribution function of the form (4.2.18) with \( r = 2 \) and both marginals equal to \( G \), is produced (cf. (4.2.19)) by a kernel of the form

\[
P^*(y|x) = \begin{cases} 
G(y) - p_1 q^*(y) & \text{for } x \leq \xi \\
G(y) + p_2 q^*(y) & \text{for } x > \xi , 
\end{cases}
\]

where \( p_1, p_2 > 0 \), \( \xi \) satisfies (4.2.35), and \( q^* \) is given by (4.2.39). The correlation coefficient is (the maximum over \( \xi \in \mathbb{R} \) of)

\[
\rho^*(\xi) = \frac{\left( \int_{-\infty}^{\xi} x G(dx) \right)^2}{G(\xi)(1 - G(\xi))}.
\]

The smallest correlation coefficient (if it exists) is produced by a kernel of the form

\[
P_\tau(y|x) = \begin{cases} 
G(y) - p_1 q_\tau(y) & \text{for } x \leq \xi \\
G(y) + p_2 q_\tau(y) & \text{for } x > \xi . 
\end{cases}
\]

The correlation coefficient is (the minimum over \( \xi, \tau \in \mathbb{R} \) of)

\[
\rho_\tau(\xi) = -\frac{\int_{-\infty}^{\xi} x G(dx) \int_{-\infty}^{\infty} y \delta(dy)}{G(\xi)(1 - G(\xi))},
\]

where \( \tau \) satisfies \( G(\tau) = 1 - G(\xi) \).

**Proof.** As \( G \) has expectation 0 and variance 1 the correlation coefficient corresponding to \( P(y|x) \) of the form (4.2.27) equals \( 1_{p_1}L_2(q) \), cf. (4.2.31). From lemmas 4.2.9 and 4.2.11 it follows that we have to consider kernels \( P^*(y|x) \) and \( P_\tau(y|x) \) of the form (4.2.43) and (4.2.45). We have, using that \( \int_{-\infty}^{\xi} x G(dx) = 0 \),

\[
\frac{\int_{-\infty}^{\xi} x G(dx) \int_{-\infty}^{\infty} y \delta(dy)}{G(\xi)(1 - G(\xi))}.
\]
\( I_1(p^*) = \int_{-\infty}^{\xi} x p^*(x) G(dx) = -p_1 \int_{-\infty}^{\xi} x G(dx) + p_2 \int_{\xi}^{\infty} x G(dx) = -p_1 \int_{-\infty}^{\xi} x G(dx) . \)

Using (4.2.35) we obtain

\[
q^*(y) = \begin{cases} 
\frac{G(y)}{p_2} & \text{for } y \leq \xi \\
1 - \frac{G(y)}{p_1} & \text{for } y > \xi
\end{cases}
\]

and hence

\[
I_2(q^*) = \int_{-\infty}^{\infty} y q^*(y) = \int_{-\infty}^{\xi} y q^*(y)
\]

\[
= -\int_{-\infty}^{\xi} y \frac{G(dy)}{p_2} + \int_{\xi}^{\infty} y \frac{G(dy)}{p_1} =
\]

\[
= -\frac{1}{p_2} \int_{-\infty}^{\xi} x G(dx) - \frac{1}{p_1} \int_{\xi}^{\infty} x G(dx) =
\]

\[
= -\frac{p_1 + p_2}{p_1 p_2} \int_{-\infty}^{\xi} x G(dx) .
\]

From (4.2.31), (4.2.47) and (4.2.49) it now follows that the covariance equals

\[
I_1(p^*) I_2(q^*) = \frac{(p_1 + p_2)^2}{p_1 p_2} \left( \int_{-\infty}^{\xi} x G(dx) \right)^2 =
\]

\[
= \frac{1}{G(\xi)(1 - G(\xi))} \left( \int_{-\infty}^{\xi} x G(dx) \right)^2 ,
\]
and this is the correlation coefficient \( \rho^*(\xi) \), since \( G \) has expectation 0 and variance 1.

Similarly we find that

\[
q_*(y) = \begin{cases} 
\frac{G(y)}{p_1} & \text{for } x \leq \xi \\
\frac{1 - G(y)}{p_2} & \text{for } x > \xi ,
\end{cases}
\]

where \( \xi \) satisfies \( G(\xi) = 1 - G(\xi) = p_1/(p_1 + p_2) \), so that

\[
I_2(q_*^*) = \frac{p_1 + p_2}{P_1 P_2} \int_{-\infty}^{\xi} y G(dy)
\]

and

\[
I_1(p^*) I_2(q_*^*) = -\frac{1}{G(\xi) (1 - G(\xi))} \int_{-\infty}^{\xi} x G(dx) \int_{-\infty}^{\xi} y G(dy) ,
\]

which proves (4.2.46).

**Remark 4.2.13.** For any \( \alpha > 0 \) we may write, instead of (4.2.43),

\[
q^*(y|\alpha) = \begin{cases} 
G(y) - \alpha p_1 q^*(y|\alpha) & \text{for } x \leq \xi \\
G(y) + \alpha p_2 q^*(y|\alpha) & \text{for } x > \xi .
\end{cases}
\]

Things can be somewhat simplified by a special choice of \( \alpha \). For instance, by taking \( \alpha = (p_1 p_2)^{-1} \) in (4.2.54), it is seen that in (4.2.43) we may without loss of generality assume that \( p_2 = p_1^{-1} \), in which case \( q^* \) is given by

\[
q^*(y) = -\min \left( \frac{1 - G(y)}{p_1} , p_1 G(y) \right).
\]

**Example 4.2.14.** If we take

\[
G(x) = \frac{1}{2\sqrt{3}} (x + \sqrt{3}) \quad \text{for } -\sqrt{3} < x < \sqrt{3} ,
\]
i.e. if $G$ is the standardized uniform distribution, then we have

$$
\int_{-1}^{1} x G(dx) = \frac{1}{2} (3 - \xi^2)^2
$$

and we find

$$
\phi^*(\xi) = \frac{1}{2} (3 - \xi^2)
$$

$$
\phi_\alpha(\xi) = -\frac{1}{2} (3 - \eta^2)
$$

The largest correlation coefficient $\phi^* = \frac{1}{2}$ is obtained for $\xi = 0$, the smallest correlation coefficient $\phi_\alpha = -\frac{1}{2}$ is obtained for $\eta = \xi = 0$ ($\eta = 0$ implies $\xi = 0$).

For the same choice of $G$ the distributions $W_p(x,y)$ and $K_p(x,y)$, given in (4.2.22), lead to the extreme correlation coefficients $1$ and $-1$, respectively. If, on the other hand, we take Morgenstern's bivariate distributions, as discussed in section 1.4.2, we obtain a largest correlation coefficient $1/3$ and a smallest correlation coefficient $-1/3$ (see below (4.4.16)).

Regarding large correlation coefficients example 4.2.14 can be extended to ranks higher than 2. First we note that for rank 2 the optimal kernel $P^*(y|x)$ can be written as

$$
P^*(y|x) = \begin{cases}
\left(1 - \frac{p_1}{p_2}\right) G(y) & \text{for } x \leq \xi, y \leq \xi \\
1 & \text{for } x \leq \xi, y > \xi \\
0 & \text{for } x > \xi, y \leq \xi \\
\left(1 + \frac{p_1}{p_2}\right) G(y) - \frac{p_2}{p_1} & \text{for } x > \xi, y > \xi
\end{cases}
$$

From (4.2.59) it is easily seen that the corresponding bivariate distribution is concentrated on the sets $(-\infty, \xi] \times (-\infty, \xi]$ and $(\xi, \infty) \times (\xi, \infty)$. This suggests that, in order to obtain a kernel of rank $r > 2$ with a large correlation coefficient, one should determine $r - 1$ distinct points $\xi_1, \ldots, \xi_{r-1}$ and take a bivariate distribution that is concentrated on the $r$ sets $(-\infty, \xi_1] \times (-\infty, \xi_1]$, $(\xi_1, \xi_2] \times (\xi_1, \xi_2]$, $\ldots$, $(\xi_{r-2}, \infty) \times (\xi_{r-2}, \infty)$.
As an example we consider the uniform distribution on \((0,1)\). With \(\xi_0 = 0\) and \(\xi_r = 1\) we take

\[
(4.2.60) \quad f_r(x, y) = \begin{cases} 
\frac{1}{\xi_j - \xi_{j-1}} & \text{for } \xi_{j-1} < x, y \leq \xi_j, \quad j = 1, \ldots, r \\
0 & \text{otherwise}
\end{cases}
\]

as the bivariate density function, and we obtain

\[
(4.2.61) \quad \text{cov}(X, Y) = \text{EXY} - \text{E}X\text{EY} = 
= \frac{1}{r} \sum_{j=1}^{r} \frac{1}{\xi_j - \xi_{j-1}} \left( \xi_j^2 - \xi_{j-1}^2 \right)^2 - 1 = 
= \frac{1}{r} \sum_{j=1}^{r} \left( \xi_j^2 - \xi_{j-1}^2 \right) \left( \xi_j - \xi_{j-1} \right) - 1 = 
= \frac{1}{r} \sum_{j=1}^{r} \xi_j^3 - \frac{1}{r} \sum_{j=1}^{r} \xi_{j-1}^3 + \frac{1}{r} \sum_{j=1}^{r} \left( \xi_j^2 \xi_{j-1} \right) - \xi_j \xi_{j-1}^2 - \frac{1}{r} = 
= \frac{1}{r} \sum_{j=1}^{r} \left( \xi_j^2 \xi_{j-1} - \xi_{j-1}^2 \xi_j \right).
\]

If we take \(\xi_j = j/r\) for \(j = 0, \ldots, r\), then we get

\[
(4.2.62) \quad \text{cov}(X, Y) = \frac{1}{r} \sum_{j=1}^{r} \left( \frac{j^2}{r-1} \right) - \left( \frac{j-1}{r} \right)^2 - \frac{1}{r} = 
= \frac{1}{4r^3} \left( \frac{r^3}{j=1} j^2 - \frac{r}{j=1} j \right) = 
= \frac{1}{4r^3} \frac{1}{3} (r-1)r(r+1) = \frac{1}{12} \left( 1 - \frac{1}{r} \right),
\]

and hence

\[
(4.2.63) \quad \rho(X, Y) = 1 - \frac{1}{r^2},
\]

which agrees with example 4.2.14, where \(r = 2\). That this is the largest correlation coefficient available for bivariate densities of the form

\[
(4.2.60)
\]

can be shown as follows. For fixed \(\xi_0, \xi_1, \ldots, \xi_{r-1}, \xi_{r+1}, \ldots, \xi_r\), the covariance in \((4.2.61)\) attains a maximum if
\[ (4.2.64) \quad \xi_t^2 \xi_{t-1} - \xi_{t+1}^2 \xi_t + \xi_{t+1}^2 \xi_{t-1} - \xi_t^2 \xi_{t+1} = \]
\[ - \xi_t^2 (\xi_{t+1} - \xi_{t-1}) + \xi_t (\xi_{t+1}^2 - \xi_{t-1}^2) \]

attains a maximum, i.e. if \( \xi_t = (\xi_{t+1} + \xi_{t-1})/2 \). It follows that \( \xi_j = j/r \)
for \( j = 0, \ldots, r \) is the optimal choice.

4.3. Applicability

Markov chains of finite rank that are not finite Markov chains, e.g.
chains with a continuous state space, do not appear to arise naturally in
practical problems. Though the model seems attractive enough, we did not
find any examples of its use in the applied literature. The usefulness of
these chains would depend on their use as a model as described in section
1.3, applications as given in BUNNENBURG [28] and [29], and on their use as
approximations to more general Markov chains.

A somewhat related model occurs in inventory control: if \( X_n \) (\( n = 0, 1, \ldots \))
is the size of the inventory of a commodity just after the \( n \)-th sale, then
for \( j = 1, \ldots, r \)
\[ (4.3.1) \quad P(X_{n+1} \leq y \mid X_n = x) = I - F(x-y+c_j) \text{ if } x_{j-1} < x \leq x_j, \]
where \( F \) is the distribution function of the sales and \( c_j \) is the amount
ordered (and instantaneously delivered) when the inventory drops to a level
between \( x_{j-1} \) and \( x_j \). We can write (4.3.1) as
\[ (4.3.2) \quad P(y|x) = \frac{1}{j} \sum_{j=1}^{r} a_j(x)D_j(y-x), \]
which is rather similar to (1.1.1). However, in the 2-step transition
kernel
\[ (4.3.3) \quad P^{(2)}(y|x) = \sum_{j=1}^{r} \sum_{k=1}^{r} a_j(x) \int D_k(u-y) a_k(u)D_j(du) \]
the factor \( D_k \), unlike \( D_k \) in (1.2.3), depends on \( u \) too, and the analysis is
less simple than in our model.
We conclude with a brief look at Markov chains of finite rank with rewards. Chains with rewards were introduced in HÒWARD [13] and have been studied extensively by many others, see e.g. VAN DER WAL [32]. We follow Howard as closely as possible.

On a transition from \( x \) to \( y \) a reward \( r(x,y) \) is received. We define

\[
q(x) := \int r(x,y)P(dy|x) = E_r r(X_n, X_{n+1})
\]

For a chain of finite rank we obtain

\[
q(x) = \sum_{k=1}^{\infty} \tau_k(x) \int r(x,y)B(dy)
\]

The expected earnings \( v_n(x) \) in \( n \) steps are given by

\[
v_n(x) := \mathbb{E}_x \left( \sum_{k=1}^{n} r(X_{k-1}, X_k) \right) = \\
= \sum_{k=1}^{n} \int q(y)P^{k-1}(dy|x) = \\
= q(x) + \sum_{k=2}^{n} \tau_k(x) c^{k-2} \int q(y)B(dy)
\]

and \( v_n \) is easily seen to satisfy the relation

\[
v_{n+1}(x) = q(x) + \int v_n(y)P(dy|x)
\]

If we restrict to indecomposable chains, we obtain from theorem 3.1.2 and from (4.3.6), in analogy to Howard,

\[
g := \lim_{n \to \infty} \frac{v_n(x)}{n} = \int q(y)G(dy)
\]

If we further restrict to the monomorphic case, we obtain from theorem 3.1.3 and from (4.3.6) and (4.3.7) that

\[
v_{n+1}(x) = (n+1)g - q(x) - g + \int \left( v_n(y) - ng \right)P(dy|x) = \\
= q(x) - g + \sum_{k=1}^{n} \tau_k(x) c^{k-1} - \int q(y)B(dy)
\]

has a limit \( v(x) \) for \( n \to \infty \), which satisfies
\[ (4.3.10) \quad v(x) + g = q(x) + \int v(y)P(dy|x). \]

Howard's policy iteration method would now consist of solving \( g \) and \( v(x) \) from (4.3.8) and (4.3.10) for given \( q \) and \( P \), and then obtaining a new \( q \) and \( P \) by maximizing the right-hand side of (4.3.10).

In the special case that the reward \( r(x,y) = r(y) \) is independent of the initial point \( x \), we obtain from (4.3.4) and (4.3.10)

\[ (4.3.11) \quad q(x) = \frac{1}{J} \sum_{j=1}^{J} q_j a_j(x), \quad v(x) = \frac{1}{J} \sum_{j=1}^{J} v_j a_j(x). \]

Now (4.3.10) is equivalent to

\[ (4.3.12) \quad v_i + g = q_i + \sum_{j=1}^{J} c_{ij} v_j \quad \text{for} \ i = 1, \ldots, r, \]

which is exactly the same relation as (4.1) in Howard [15]. Solving (4.3.10) with given \( a_j(x) \) is equivalent to solving (4.3.12). Also, if in our model the \( a_j(x) \) would be fixed and nonnegative, then maximizing the right-hand side of (4.3.10) with respect to \( q(x) \) and \( P(.|x) \) would be equivalent to maximizing \( q + Cy \) with respect to \( q \) and \( C \), i.e., with respect to the possible kernel matrices that can be obtained for the given \( a_j \) and the admissible \( B_j \) as determined by the problem at hand. In this way a continuous-type Markov programming problem could be reduced to a discrete one.
A.1. The spectral decomposition of a matrix

In Chapters 2 and 3 we use detailed information about the spectral properties of several matrices closely related to transition matrices. In order not to distract unnecessarily from the main theme, a number of known results on spectral behaviour are collected in this appendix.

Somewhat unusually, we start from the well-known Jordan theorem about the normal form of a matrix, which we do not prove, and derive the properties we need from this (very strong) result. A standard book on matrix theory proceeds, of course, in the opposite direction: first elementary results are derived and finally Jordan's theorem is proved. The purpose of this section, however, is only to clarify and ascertain the correctness of what we use in our work. The advantage of our presentation here is a very explicit formulation of the required results.

We first give Jordan's theorem.

Theorem A.1.1. For every complex $n \times n$ matrix $M$ there exists a nonsingular $n \times n$ matrix $U$ such that $M' := U^{-1}MU$ has the form

\[(A.1.1) \quad M' = \begin{bmatrix} \lambda_1 & 1 & & & \\ & \ddots & \ddots & \ddots & \\ & & \ddots & \ddots & 1 \\ & & & \ddots & \lambda_n \\ & & & & \lambda_1 \end{bmatrix} \]
Each block has a fixed complex number $\lambda_k$ on its main diagonal and elements 1 on its first superdiagonal; all other elements in each block as well as all elements of $H^*$ outside these blocks are 0. The size of a block may be $1 \times 1$, in which case the block consists of a single element $\lambda_k$. $H$ is unique up to permutation of the blocks.

A proof of theorem A.1.1 can be found in standard texts on matrix theory, see for instance Lancaster [21]. Further a lot of detailed information can be found in Van Dantzig [3].

We stipulate that in (A.1.1) the numbers $\lambda_0, \ldots, \lambda_q$ are distinct. A block with $\lambda_k$ on the diagonal is called a $\lambda_k$-block. By $g_k$ we denote the number of $\lambda_k$-blocks and by $n_k$ the total number of $\lambda_k$'s on the diagonal of $H^*$. Let $D_M(\lambda) := \det(M - \lambda I)$ be the characteristic polynomial of $M$. We have

**Lemma A.1.2.** The characteristic polynomials of $M$ and $H^*$ satisfy

$$D_M(\lambda) = D_{H^*}(\lambda) = \prod_{k=1}^{q} (\lambda_k - \lambda)^{g_k}.$$  

**Proof.** The first equality follows from the similarity of $M$ and $H^*$, the second one from the explicit diagonal form of $H^*$ and the definition of $n_k$. $lacksquare$

We see that the $\lambda_k$ are the eigenvalues of $M$ and of $H^*$, and that $n_k$ is the algebraic multiplicity of $\lambda_k$. We now show that $g_k$ is the geometric multiplicity of $\lambda_k$, i.e. the dimension of the $\lambda_k$-eigenspace.

Let $e_k$ be the $k$-th unit vector ($k = 1, \ldots, r$), i.e. $e_k = (0, \ldots, 0, 1, 0, \ldots, 0)$ with the element 1 at position $k$.

**Lemma A.1.3.** Suppose a $\lambda$-block has upper left corner at position $(k_1, k_1)$ and lower right corner at position $(k_2, k_2)$ in $H^*$. Then $e_{k_1}$ is a right $\lambda$-eigenvector and $e_{k_2}$ is a left $\lambda$-eigenvector.

**Proof.** Follows from (A.1.1) by an easy computation. $lacksquare$

**Lemma A.1.4.** There is no right $\lambda$-eigenvector of $H^*$ independent of the right $\lambda$-eigenvectors given by Lemma A.1.3 for the various $\lambda$-blocks. The same is true for left $\lambda$-eigenvectors.

**Proof.** The statements are easily verified from (A.1.1). $lacksquare$
Corresponding to each $\lambda$-block there is exactly one (up to scalar multiples) right and one left $\lambda$-eigenvector, and these vectors form a basis for the right and for the left $\lambda$-eigenspace, respectively. Hence $g_k$ is the geometric multiplicity of $\lambda_k$. Obviously, $g_k \leq m_k$ and $g_k = m_k$ only if all $\lambda_k$-blocks are of size 1. In that case $\lambda_k$ is called nondegenerate.

**Lemma A.1.5.** If $u_k$ is a left $\lambda_k$-eigenvector and $v_k$ a right $\lambda_k$-eigenvector of $M^*$, then

$$S_k u_k v_k = 0 \quad \text{if} \quad \lambda_k \neq \lambda_k.$$ 

**Proof.** Follows directly from the explicit form of the eigenvectors.

For $k = 0, \ldots, s$ we define the $\tau \times \tau$ matrices $E_k^*$ and $N_k^*$ as follows: $E_k^*$ is obtained from $M^*$ by replacing all elements $\lambda_k$ by 1 and all other elements by 0; $N_k^*$ is obtained from $M^*$ by retaining the elements 1 on the superdiagonals of all $\lambda_k$-blocks and replacing all other elements by 0. For $M^*$ we now have

**Lemma A.1.6.**

$$M^* = \sum_{k=0}^{s} \left( \lambda_k E_k^* + N_k^* \right),$$

where $E_k^*$ and $N_k^*$ have the following properties:

$$E_k^* = 1,$$

$$E_k^* E_k^* = 1,$$

$$N_k^* = 1,$$

$$N_k^* E_k^* = 1,$$

$$E_k^* N_k^* = 1,$$

$$E_k^* E_k^* = 1 \quad \text{if} \quad \lambda_k \neq \lambda_k.$$ 

**Proof.** All properties easily follow from the definition of $E_k^*$ and $N_k^*$.
LEMMA A.1.7. For a given matrix \( M^* \) the matrices \( E_k^* \) and \( N_k^* \) are uniquely determined by (A.1.4)–(A.1.9).

PROOF. Suppose we have \( M^* = \sum_{k=0}^{\infty} \left( \lambda_k E_k^* + N_k^* \right) = \sum_{k=0}^{\infty} \left( \lambda_k E_k^* + N_k^* \right), \) \( E_k^* \) and \( N_k^* \) also satisfying (A.1.5)–(A.1.9). From (A.1.4), (A.1.6) and (A.1.9) it follows that

\[
(A.1.10) \quad N_k^* = (M^* - \lambda_k I)E_k^* = E_k^*(M^* - \lambda_k I),
\]

and hence, using (A.1.6), that

\[
(A.1.11) \quad (N_k^*)^n = (M^* - \lambda_k I)^n E_k^* = E_k^*(M^* - \lambda_k I)^n \quad \text{for all } n \geq 1.
\]

By (A.1.7) \( N_k^* \) is nilpotent, so that for \( n \) sufficiently large we obtain

\[
(A.1.12) \quad 0 = \begin{bmatrix} X & \cdots & \cdots \\ \vdots & \ddots & \vdots \\ \vdots & \vdots & X \end{bmatrix} = \begin{bmatrix} E_k^* \cdots \cdots \\ \vdots \ddots \vdots \\ \vdots \cdots E_k^* \end{bmatrix} = \begin{bmatrix} E_k^* \\ \vdots \\ E_k^* \end{bmatrix} 
\]

where \( X \) and \( Y \) are triangular matrices with elements of the form \((\lambda_k - \lambda_k)^n \neq 0\) on the main diagonal, and where all other parts contain only 0's. Clearly \( X \) and \( Y \) are nonsingular and therefore \( E_k^* \) has the form (only 0's in the blank parts)

\[
(A.1.13) \quad E_k^* = \begin{bmatrix} E_k^* \\ \vdots \\ E_k^* \end{bmatrix}.
\]

This yields

\[
(A.1.14) \quad E_k^* E_k^* = 0 \quad \text{if } k \neq \ell.
\]

Using (A.1.5) and (A.1.14) we get

\[
(A.1.15) \quad E_k^* = E_k^* \left( \sum_{k=0}^{\infty} \left( \lambda_k E_k^* + N_k^* \right) \right) = E_k^* E_k^* = \left( \sum_{k=0}^{\infty} \left( \lambda_k E_k^* + N_k^* \right) \right) E_k^* - E_k^*,
\]

and hence by (A.1.10) also \( N_k^* = N_k^* \).
We conclude our observations of $M^*$ by a remark on the matrices $E_k^*$. Define the $r \times n_k$ matrix $E_k^*$ and the $m_k \times r$ matrix $L_k^*$ by

\begin{equation}
E_k^* := \begin{bmatrix}
0 & \ldots & 0 & 1 & \ldots & 0 & 0 & \ldots & 0
\end{bmatrix},
\end{equation}

\begin{equation}
L_k^* := \begin{bmatrix}
0 & \ldots & 0
\vdots & \ddots & \vdots
0 & \ldots & 0
\end{bmatrix}.
\end{equation}

Then we have

\begin{equation}
E_k^* = R_k^* L_k^*,
\end{equation}

and

\begin{equation}
L_k^* R_k^* = I_{m_k}.
\end{equation}

If $\lambda_k$ is nondegenerate, then the columns of $R_k^*$ are linearly independent right $\lambda_k$-eigenvectors of $M^*$ and the rows of $L_k^*$ are (transposed) linearly independent left $\lambda_k$-eigenvectors.

Now let $M$ be an arbitrary complex $r \times r$ matrix with Jordan matrix $M^* = U^{-1} M U$, so that

\begin{equation}
M^* = U M U^{-1}.
\end{equation}

We have

**Theorem A.1.8.** For each $r \times r$ matrix $M$ with distinct eigenvalues $\lambda_0, \ldots, \lambda_n$, there exist matrices $E_k$ and $N_k$, such that

\begin{equation}
M = \sum_{k=0}^{n} (\lambda_k E_k + N_k),
\end{equation}

and such that $E_k$ and $N_k$ have the same properties (A.1.5)–(A.1.9) as the $E_k^*$ and $N_k^*$. For a given $M^*$, i.e., for a given $U$, the $E_k$ and $N_k$ are uniquely determined by these conditions.

**Proof.** Take $E_k^*$ and $N_k^*$ as in lemma A.1.6, and define

\begin{equation}
E_k := U E_k^* U^{-1}, \quad N_k := U N_k^* U^{-1}.
\end{equation}

From lemma A.1.6 and (A.1.19) it is easily verified that these $E_k$ and $N_k$ fulfill the requirements, while their uniqueness follows from lemma A.1.7. \[\square\]
The expression for \( M \) given by (A.1.20) is called a spectral decomposition of the matrix \( M \). Its use is evident when we consider powers of \( M \).

**Theorem A.1.2.** Let \( M \) be a matrix with spectral decomposition (A.1.20). Then

\[
M^n = \sum_{k=0}^{\infty} (\lambda_k^N \mathbf{z}_k + N_k) \mathbf{z}_k \mathbf{z}_k^T
\]

for all \( n \geq 1 \).

**Proof.** The relation (A.1.22) follows directly from the special properties of \( \mathbf{z}_k \) and \( N_k \).

The number of terms of the inner sum in (A.1.22) is \( n_k - \hat{c}_k \). Hence this sum vanishes if \( \lambda_k \) is nondegenerate, while the number of terms is always less than \( r \).

For eigenvectors the relation between \( M \) and \( M^* \) is as follows. The vector \( \mathbf{v} \) is a right \( \lambda \)-eigenvector of \( M \) if and only if \( U^{-1} \mathbf{v} \) is a right \( \lambda \)-eigenvector of \( M^* \), whereas \( \mathbf{u} \) is a left \( \lambda \)-eigenvector of \( M \) if and only if \( U \mathbf{u} \) is a left \( \lambda \)-eigenvector of \( M^* \). This follows easily from (A.1.19). We therefore have

**Lemma A.1.10.** If \( \mathbf{y}_k \) is a left \( \lambda_k \)-eigenvector and \( \mathbf{x}_k \) a right \( \lambda_k \)-eigenvector of \( M \), then

\[
\mathbf{x}_k^T \mathbf{y}_k = 0 \quad \text{if} \quad \lambda_k \neq \lambda_k'.
\]

**Proof.** Use the argument given above and apply Lemma A.1.5.

Define the matrices \( \mathbf{R}_k \) and \( \mathbf{L}_k \) by

\[
\mathbf{R}_k := U^{-1} \mathbf{R}_k^* \quad \text{and} \quad \mathbf{L}_k := \mathbf{L}_k^* U
\]

where \( \mathbf{R}_k^* \) and \( \mathbf{L}_k^* \) are given by (A.1.16). One now easily verifies the following results:
**Lemma A.1.11.** If \( \lambda_k \) is nondegenerate, then the matrix \( \mathbf{E}_k \) can be written as

\[
\mathbf{E}_k = \mathbf{R}_k \mathbf{L}_k,
\]

where the \( r \times n \) matrix \( \mathbf{R}_k \) and the \( n \times r \) matrix \( \mathbf{L}_k \) are given by (A.1.24). The columns of \( \mathbf{R}_k \) are linearly independent right \( \lambda_k \)-eigenvectors of \( \mathbf{M}_k \), the rows of \( \mathbf{L}_k \) are (transposed) linearly independent left \( \lambda_k \)-eigenvectors of \( \mathbf{M}_k \), and

\[
\mathbf{L}_k \mathbf{R}_k = \mathbf{I}_{\mathbf{M}_k}.
\]

**Corollary A.1.12.** If \( \lambda_k \) is an eigenvalue with algebraic multiplicity 1, then

\[
\mathbf{E}_k = \mathbf{Q}_k \mathbf{Q}_k^T,
\]

where \( \mathbf{Q}_k \) and \( \mathbf{Q}_k^T \) are any right and left \( \lambda_k \)-eigenvectors.

**Proof.** Since \( \lambda_k \) is nondegenerate, Lemma A.1.11 applies. In this case the matrices \( \mathbf{R}_k \) and \( \mathbf{L}_k \) are in fact two vectors. If \( \mathbf{v}_k \) and \( \mathbf{u}_k \) are arbitrary right and left eigenvectors, then \( \mathbf{v}_k = \alpha \mathbf{E}_k \) and \( \mathbf{u}_k = \beta \mathbf{L}_k \) for some nonzero numbers \( \alpha \) and \( \beta \), and we have \( \mathbf{v}_k^T \mathbf{u}_k = \alpha \beta \).

**Lemma A.1.13.** Let \( \mathbf{v} \) be an arbitrary vector. If \( |\lambda_k| < 1 \), then for any \( \rho \) with \( |\lambda_k| < \rho < 1 \)

\[
(\lambda_k \mathbf{E}_k + \rho \mathbf{I})^n \mathbf{v} = \mathcal{O}(\rho^n) \quad (n \to \infty).
\]

**Proof.** We have, cf. (A.1.22),

\[
| (\lambda_k \mathbf{E}_k + \rho \mathbf{I})^n \mathbf{v} | \leq | \lambda_k^n \mathbf{E}_k \mathbf{v} | + \sum_{k=1}^{n-1} | \lambda_k^{n-k} \rho^k \mathbf{E}_k \mathbf{v} | \leq
\]

\[
= | \lambda_k^n | | \mathbf{E}_k \mathbf{v} | + | \sum_{k=1}^{n-1} \lambda_k^{n-k} \rho^k | | \mathbf{E}_k \mathbf{v} | =
\]

\[
= \mathcal{O}(\rho^{n-1}) | \lambda_k | \mathcal{O}(\rho) \quad (n \to \infty).
\]

If \( |\lambda_k| < \rho < 1 \), then the last term is \( \mathcal{O}(\rho^n) \) (n \to \infty).
For kernel matrices we have the following situation (cf. theorem 2.1.4 and lemma 2.1.5): \( \lambda_0, \ldots, \lambda_{d-1} \) are nondegenerate eigenvalues of modulus 1 and \( \lambda_0, \ldots, \lambda_{d} \) satisfy \( |\lambda_k| < c < 1 \) for some \( c \). We have

**Lemma A.1.16.** Let \( \mathcal{E}_k \) be the vector space spanned by the left \( \lambda_k \)-eigenvectors with \( k = 0, \ldots, d-1 \) of a kernel matrix \( C \). If \( v \in \mathcal{E}_k \) then

\[
C^n v = o(c^n) \quad (n = \infty) .
\]

**Proof.** We use the spectral decomposition of \( C \). As \( L_k v = 0 \) for \( k = 0, \ldots, d-1 \), we have

\[
\sum_{k=0}^{d-1} \lambda_k^n x_k v = \sum_{k=0}^{d-1} \lambda_k^n L_k v = 0 ,
\]

hence, using lemma A.1.13,

\[
C^n v = \sum_{k=0}^{d-1} (\lambda_k^n x_k + x_k) v = o(c^n) \quad (n = \infty) .
\]

**Lemma A.1.17.** If \( C \) is a kernel matrix and \( v \) a vector satisfying

\[
\lim_{n \to \infty} C^n v = 0 ,
\]

then

\[
C^n v = o(c^n) \quad (n = \infty) .
\]

**Proof.** Let \( (v_1, \ldots, v_n) \) be a basis of left eigenvectors for the space \( \mathcal{E}_\lambda \), defined in the previous lemma. We can write \( v \) in the form

\[
v = \sum_{j=1}^{n} a_j v_j + a_0 v_0 ,
\]

with \( v_0 \in \mathcal{E}_\lambda \) and where the \( a_j \) are complex numbers. We get

\[
C^n v = \sum_{j=1}^{n} a_j C^n v_j + a_0 C^n v_0 =
\]

\[
= \sum_{j=1}^{n} a_j \lambda_j^n v_j + o(c^n) \quad (n = \infty) ,
\]
with \( \ell(j) \in \{0, \ldots, d-1\} \) for all \( j \). From (A.1.33) it follows that

\[(A.1.37) \quad \lim_{n \to \infty} \sum_{j=1}^{m} a_j \lambda^n(j) v_j = 0,\]

and hence, by the linear independence of the \( v_j \),

\[(A.1.38) \quad \lim_{n \to \infty} a_j \lambda^n(j) = 0 \quad \text{for all } j.\]

Since \( |\lambda(j)| = 1 \) we must have \( a_j = 0 \) for all \( j \), and (A.1.34) follows from (A.1.36).

\[\Box\]

### A.2. Miscellaneous

In this section of the appendix we prove the various results that we use in the main text.

**Lemma A.2.1.** Suppose the \( r \times r \) matrix \( M(t) \) has a first derivative (element-wise) with respect to \( t \in J \), \( J \) an open subset of \( \mathbb{R} \). If all eigenvalues \( \lambda(t) \) of \( M(t) \) satisfy \( |\lambda(t)| \leq \epsilon_0 < 1 \) for all \( t \in J \), then

\[(A.2.1) \quad \frac{d}{dt} M^0(t) = o(c^n) \quad (n \to \infty)\]

for any \( \epsilon \) with \( \epsilon_0 < \epsilon < 1 \) and all \( t < J \).

**Proof.** Let \( \| \cdot \| \) denote the matrix norm defined by

\[(A.2.2) \quad \|M\| = \max_{i,j} \left| \frac{1}{n} \right| M_{ij} \| .\]

Take a fixed \( \epsilon \) with \( \epsilon_0 < \epsilon < 1 \) and then an arbitrary \( \epsilon_1 \) with \( \epsilon_0 < \epsilon_1 < \epsilon \).

Using Lemma A.1.3 we find that

\[(A.2.3) \quad \|M^n(t)\| = o(c^n) \quad (n \to \infty) \quad \text{for all } t < J,\]

hence for each \( t < J \) there is an integer \( N > 1 \), such that

\[(A.2.4) \quad \|M^n(t)\| \leq \epsilon_1^n \quad \text{for all } n \geq N.\]

From
\begin{align}
  &\frac{d}{dt} M^n(t) = \sum_{k=0}^{n-1} M^k(t) M^n(t) M^{n-k-1}(t) \\
  \text{we obtain for all } n \geq 2N + 1 \tag{A.2.5} &
  \frac{d}{dt} M^n(t) \leq \sum_{k=0}^{n-1} \|M^k(t)\| \|M^n(t)\| \|M^{n-k-1}(t)\| \\
  \leq & \|M^n(t)\| \left( \sum_{k=0}^{n-1} \|M^k(t)\| \|M^{n-k-1}(t)\| + \sum_{k=N+1}^{n-1} \varepsilon_k^{n-1} + \sum_{k=N+1}^{n-1} \varepsilon_k^{N-k-1}(t) \right). \tag{A.2.6}
\end{align}

\begin{align}
  \frac{d}{dt} M^n(t) = \mathcal{O}(m^N) = o(\varepsilon^N) \quad (n \to \infty), \tag{A.2.7}
\end{align}

and a fortiori that (A.2.1) holds.

**Lemma A.2.2.** If the matrix \(M(t)\) of Lemma A.2.1 has a second derivative, then also

\begin{align}
  \frac{d^2}{dt^2} M^n(t) = o(\varepsilon^N) \quad (n \to \infty). \tag{A.2.8}
\end{align}

**Proof.** For the second derivative of \(M^n(t)\) we have, cf. (A.2.5),

\begin{align}
  \frac{d^2}{dt^2} M^n(t) = & \sum_{k=0}^{n-1} M^k(t) M^n(t) M^{n-k-1}(t) \\
  &+ \sum_{k=N+1}^{n-1} \varepsilon_k^{n-1} \sum_{l=0}^{k-N-1} M^l(t) M^n(t) M^{n-l-1}(t) M^{n-k-l-2}(t). \tag{A.2.9}
\end{align}

This yields for each \(t \in \mathcal{J}\) and all \(n \geq 2N + 1\) an inequality analogous to (A.2.6), and it follows that

\begin{align}
  \frac{d^2}{dt^2} M^n(t) = \mathcal{O}(m^N) = o(\varepsilon^N) \quad (n \to \infty). \tag{A.2.10}
\end{align}

It is well known that eigenvalues with modulus 1 of transition matrices are nondegenerate. The following lemma is a generalization of this result; our proof is based on the proof given by Fritz, Huppert and Willems [12] for transition matrices.
Lemma A.2.3. Suppose that all eigenvalues \( \lambda \) of the \( r \times r \) matrix \( M \) satisfy \( |\lambda| \leq 1 \) and that \( M^n \) is bounded for \( n \in \mathbb{N} \). Then all eigenvalues \( \lambda \) with \( |\lambda| = 1 \) are nondegenerate.

Proof. First we note that \( \|M^n\| \) is bounded for \( n \in \mathbb{N} \) (i.e., as in (A.2.2)). Suppose that \( \lambda \) with \( |\lambda| = 1 \) is a degenerate eigenvalue, i.e., in the Jordan form \( M = U^{-1}NU \) of \( M \) there is a \( \lambda \)-block

\[
\begin{pmatrix}
\lambda & 1 & \cdots & 0 \\
0 & \ddots & \ddots & \vdots \\
\vdots & \ddots & \ddots & 1 \\
0 & \cdots & 0 & \lambda
\end{pmatrix} =: Q
\]

(A.2.11)
of size at least \( 2 \times 2 \). An easy computation yields that \( (M^n) \) contains the block

\[
\begin{pmatrix}
\lambda^n & 0 & \cdots & 0 \\
0 & \ddots & \ddots & \vdots \\
\vdots & \ddots & \ddots & \lambda^n \\
0 & \cdots & 0 & \lambda^n
\end{pmatrix}
\]

(A.2.12)

We only consider the second last row of \( Q \). We get

\[
(A.2.13) \quad \|Q^n\| = \|\lambda^n\| = \|\alpha^n\| = 1 + n \quad \text{for all } n \geq 1.
\]

On the other hand we have

\[
(A.2.14) \quad \|Q^n\| = \|U^{-1}NU^n\| = \|U^{-1}N^nU\| = \mathcal{O}(1)
\]

\[\leq \mathcal{O}(1) \cdot \|M^{2n}\| \quad \text{for all } n \geq 1.
\]

The combination of (A.2.13) and (A.2.14) contradicts the boundedness of \( M^n \). Hence \( \lambda \) is nondegenerate.

Lemma A.2.4. If \( f_1, \ldots, f_r \) are linearly independent complex-valued functions, then there exists at least one \( r \)-tuple \( x_1, \ldots, x_r \) such that
\[
\begin{pmatrix}
\varepsilon_1(x_1) & f_1(x_2) & \ldots & f_1(x_r) \\
\varepsilon_2(x_1) & f_2(x_2) & \ldots & f_2(x_r) \\
\vdots & \vdots & \ddots & \vdots \\
\varepsilon_r(x_1) & f_r(x_2) & \ldots & f_r(x_r)
\end{pmatrix} \neq 0.
\]

**Proof.** We first note that none of the \( f_i \) can vanish identically, as the \( f_j \) would then be linearly dependent. We proceed by induction.

If \( r = 1 \) the assertion is trivial.

Now let \( r \geq 2 \) and suppose we can construct an \((r-1)\)-tuple \( x_1, \ldots, x_{r-1} \) such that the \((r-1) \times (r-1)\) matrix \( D = (d_{jk}) \) with elements \( d_{jk} := f_k(x_j) \) for \( j, k \in \{1, \ldots, r-1\} \) satisfies the condition \( \det D \neq 0 \). Let \( y \in \mathbb{C}^{r-1} \) be the vector with \( j \)-th component \( y_j = f_j(x_j) \). The equation \( Dy = y \) has exactly one solution vector \( y = x^* \in \mathbb{C}^{r-1} \). From the linear independence of \( f_1, \ldots, f_r \) we know that there is at least one \( x_r \) such that

\[
f_r(x_r) \neq \sum_{j=1}^{r-1} f_j(x_r) y_j.
\]

One easily verifies that the \( r \)-tuple \( x_1, \ldots, x_r \) satisfies (A.2.15).

**Lemma A.2.5.** Let \( \lambda_k \) be given by

\[
\lambda_k = e^{2\pi i k/d} \quad \text{for } k = 0, \ldots, d-1.
\]

If \( a_0, \ldots, a_{d-1} \) are complex numbers such that

\[
\sum_{k=1}^{d-1} a_k \lambda_k^n = a_0 \quad \text{for all } n \geq 1,
\]

then \( a_0 = a_1 = \ldots = a_{d-1} = 0. \)

**Proof.** The polynomial

\[
A(x) := \sum_{k=1}^{d-1} a_k x^k - a_0
\]

is of degree \( d-1 \), but has \( d \) zeros: \( x_k = e^{2\pi i (k/d)} \) for \( k = 0, \ldots, d-1 \). It follows that \( A(x) \equiv 0 \), i.e. \( a_0 = a_1 = \ldots = a_{d-1} = 0. \)
REFERENCES


[25] DNICESCU, O. and Gh. MEHOC (1940), Propriétés analytiques des chaînes de Markoff étudiées à l'aide de la fonction caractéristique, Mathematica (Cluj) XVI, 13-43.


SAMENVATTING

Een Markov-keten is een Markov-proces met een discrete tijd parameter, d.w.z. een rij stochastische grootheden $X_0, X_1, X_2, \ldots$ met de eigenschap dat voor alle $A \in \mathcal{B}(\mathbb{R})$

$$P(X_{n+1} \in A \mid X_0 = x_0, \ldots, X_n = x_n) = P(X_{n+1} \in A \mid X_n = x_n).$$

Een Markov-keten beschrijft een systeem dat op gegeven tijdstippen $t_1, t_2, \ldots$ van de ene toestand in de volgende overgaat, waarbij de kansverdeling van "de toekomst" alleen afhangt van "het heden" en niet van "het verleden".

De eenvoudigste Markov-ketens zijn eindige Markov-ketens, waarbij de verzameling $S$ van mogelijke toestanden, de toestandsruimte, eindig is: $S = \{s_1, \ldots, s_r\}$. Zo'n proces wordt beschreven door een $r \times r$ overgangsmatrix $P$ gedefinieerd door

$$P_{jk} = P(X_{n+1} = s_k \mid X_n = s_j),$$
waarbij voor de $k$-staps overgangskansen geldt

$$(1) \quad P(X_{n+k} = s_k \mid X_n = s_j) = P_k^{(k)} = (P^k)_{jk}. $$

In dit proefschrift worden Markov-ketens van eindige rang bestudeerd. Dit zijn Markov-ketens met een willekeurige toestandsruimte, maar met een zeer speciale overgangsstructuur, nl. (in vectormatrix)

$$P(X_{n+1} \in A \mid X_n = x) = \sum_{j=1}^{r} a_j(x)B_j(A) = \sum_{j=1}^{r} a_j(x)B_j(A).$$

Dit heeft tot gevolg dat voor de $k$-staps overgangskansen geldt, analoog aan (1):

$$P(X_{n+k} \in A \mid X_n = x) = \sum_{j=1}^{r} a_j(x)C_{jk}(A),$$
waarbij de $r \times r$ matrix $C$, de covariantmatrix, gedefinieerd is door

$$C_{jk} = \int a_k(x)B_j(dx).$$
Markov-ketens van eindige rang, die in 1960 door Runnenburg werden geïntroduceerd, hebben het voordoen een grote algemeenheid toe te laten, terwijl de rekentechnische problemen in wenen niet groter zijn dan die bij eindige Markov-ketens. De structuur van deze processen wordt onderzocht in hoofdstuk 1. In de hoofdstukken 2, 3 en 4 wordt deze specifieke structuur uitgebreid en worden met behulp van eenvoudige matrix-theorie allerlei resultaten afgeleid. Alle stellingen gelden in het bijzonder voor eindige Markov-ketens, die als speciale gevallen van Markov-ketens van eindige rang kunnen worden opgevat. Sommige stellingen kunnen eveneens worden als speciale gevallen van reeds bekende resultaten; het voordeel van ons model ligt in de expliciete vorm van de resultaten en de relatieve eenvoud van de bewijzen.

In hoofdstuk 2 wordt de eigenvaardestructuur van kornmatrices bestudeerd. Analoog aan de situatie bij eindige Markov-ketens blijkt dat de multipliciteit van de eigenwaarde 1 overeenkomt met het aantal disjuncte absorbierende deelverzamelingen van de toestandsruimte. Na enige beschouwingen over recurrente en transiente verzamelingen wordt de Hopf-decompositie voor ketens van eindige rang bewezen. Voorts blijkt dat, als de eigenwaarde 1 enkelvoudig is, \( z \in \mathbb{C} \mid z = \exp(i \pi t)/d, t = 0, \ldots, d-1 \) voor een zekere \( d \geq 1 \) de verzameling van eigenwaarden met absolute waarde 1 is. Deze eigenwaarden corresponderen met een cykel van \( d \) disjuncte verzamelingen \( S_0 = S_1 = \ldots = S_{d-1} = S_0 \); er is, ingeval \( d \geq 2 \), sprake van periodicitie.

Het zwaarte punt van dit proefschrift ligt bij de limietstellingen van hoofdstuk 3. Onder gebruikmaking van de resultaten van hoofdstuk 2 en van de speciale eigenschappen van matrizes (die bewezen worden in een appendix) wordt eerst het bestaan van een invariantverdeling onderzocht. Vervolgens worden een centrale-limietstelling voor ketens van eindige rang en enige verwante resultaten bewezen. Verder blijkt voor ketens van eindige rang ook een analoge bestaan van de klassieke extreme-waardenstelling voor rijen van onafhankelijke stochastische variabelen. Bovendien geldt hierbij: als \( G \) de invariantverdelingsfunctie van de Markov-ketens is en \( F \) een niet-ontaarde verdelingsfunctie, dan is \( \lim_{n \to \infty} P(\max(X_1, \ldots, X_n) \leq a_n y + b_n) = F(y) \) voor alle \( y \) als \( \lim_{n \to \infty} c_n(a_n y + b_n) = F(y) \). Tenslotte wordt enige elementaire vervangingsetheorie voor Markov-ketens van eindige rang beschouwd. Bij alle bewijzen spelen machten van \( r \times r \) matrizes een essentiële rol; deze matrizes zijn verwant met de kernmatrix en hebben evenals die kernmatrix een dominante eigenwaarde. Een voorbeeld van dat soort matrizes is de matrix \( \mathbb{C}(t) \) die een essentiële rol speelt bij het bewijm van de centrale-limietstelling.
\[ \tilde{c}_{jk}(t) = \int s_k(x)e^{itx}s_j(dx), \]
met \( \tilde{c}(0) = c. \)

In hoofdstuk 4 wordt een aantal generalisaties gegeven en de mogelijkheid van toepassen bekeken. De centraal-limietstelling van hoofdstuk 3 wordt gegeneraliseerd tot hogere dimensies en tot functies van Markov-ketens van eindige rang. Over de mogelijkheid van een generalisatie tot Markov-processen met een continue tijdpараметer worden enige opmerkingen gemaakt. Op de vraag in hoeverre algemene Markov-ketens benaderd kunnen worden door ketens van eindige rang kan geen algemeen antwoord gegeven worden, maar wel worden enkele mogelijkheden en moeilijkheden aangewezen. Speciaal wordt onderzocht welke waarden de correlatiecoëfficiënt van twee opvolgende variabelen kan hebben in een Markov-keten van rang 2; voor ketens van rang groter dan 2 wordt een voorbeeld gegeven. Tot slot wordt, zeer in het kort, dynamische programmering in Markov-ketens van eindige rang besproken. Een speciaal optimaliseringsprobleem van continue aard kan in het geval van een Markov-keten van eindige rang worden teruggebracht tot een soortgelijk probleem van discrete aard.
CURRICULUM VITAE


Vanaf augustus 1982 is de schrijver als docent wiskunde verbonden aan het Lorentzlyceum te Eindhoven en aan de Katholieke Leerlingen te Tilburg.
STELLINGEN
bij het proefschrift

ON MARKOV CHAINS OF FINITE BARK.

van

Aigle Hoekstra
Als $X_1, X_2, \ldots$ een submartingaal is, dan zijn de drie beweringen

(a) $\sup_{n \in \mathbb{N}} \mathbb{E}|X_n| < \infty$.

(b) $\liminf_{n \to \infty} \mathbb{E}|X_n| < \infty$.

(c) $\liminf_{n \to \infty} \mathbb{E} X_n < \infty$.

equivalent.

II

Laat $X_1, X_2, \ldots$ een rij onafhankelijke, gelijkeverdeelde stochastische variabelen zijn met $P(X_1 = 1) = P(X_1 = -1) = \frac{1}{2}$. Zij $S_n := X_1 + \ldots + X_n$ voor $n \geq 1$. In 1976 stelde Freudenthal, dat hij geen waarde voor $n_0$ zou kunnen geven waarvoor geldt

$$P\left(\sup_{n > n_0} \frac{|S_n|}{n} \leq 0.04\right) \geq 0.95.$$ 

deze ongelijkheid is juist voor $n_0 = 6610$.

tekeningen, DWO, Utrecht.


III

Laat $X_1, X_2, \ldots$ een rij onafhankelijke, standaardverdeelde stochastische variabelen zijn en laat $S_n := X_1 + \ldots + X_n$ voor $n \geq 1$.

Voor elke $\varepsilon > 0$ en elke $N \in \mathbb{N}$ met $N > \varepsilon^{-2} \log 4$ geldt voor alle $n_0 \geq 1$

$$P\left(\sup_{n > n_0} \frac{S_n}{n} \geq \varepsilon\right) \leq P\left(\frac{S_{n_0}}{n_0} > \varepsilon\right).$$
Zij voor $n \geq 1$ de verdelingsfunctie $G_n$ gegeven door

$$
G_n(x) = \begin{cases} 
0 & \text{voor } x < 0 \\
1 & \text{voor } 0 \leq x \leq n \\
1 & \text{voor } x > n.
\end{cases}
$$

Dan geldt

$$
l_v (n \cdot x(V(x)) = V(x) \text{ voor alle } x \in \mathbb{R},
$$

waarin $V(x)$ een niet-omstaande verdelingsfunctie is, terwijl $(nx)^{n-1}$, alleen voor $x = 0$ een begrensde rij is. Dit voorbeeld is in strijd met een uitspraak in FELLER [2] die een onderdeel vormt van het bewijs van Lemma VIII.2.1; dit lemma wordt in FELLER [1] correct bewezen.


Als $\lambda_1, \ldots, \lambda_n$ een $\nu$-tal reële getallen is, niet noodzakelijk verschillend, met $\lambda_1 = 1$, $|\lambda_j| \leq 1$ (j = 1, ..., n) en $\lambda_1 \cdot \lambda_2 \cdot \ldots \cdot \lambda_n = -1$ voor elke deelverzameling $\{\lambda_1, \lambda_2, \ldots, \lambda_k\}$ van $\{1, \ldots, n\}$, dan bestaat er een $n \times n$ Markov-matrix met eigenwaarden $\lambda_1, \ldots, \lambda_n$. 

VII

In het Promotiereglement Technische Hogeschool Eindhoven 1981 wordt niet duidelijk aangegeven wat er bedoeld wordt met een "stelling". Speciaal bij wiskundigen kan dit tot verwarring leiden, mede vanwege de geïsoleerde mogelijkheid tot discussie over stellingen.

VIII

Gevolgd de moderne ontwikkelingen op het gebied van de apparatuur voor tijdssynchronisatie is het denkbaar dat een uitdrukking als "met de klok mee" door toekomstige generaties wiskundigen niet begrepen zal worden.