Embedding discrete- into continuous-time Markov chains

Bachelor thesis

submitted by Leon V. S. Roschig to the

Department of Mathematics and Computer Science

Philipps-University Marburg

September 28, 2015

Supervised by

Prof. Dr. Hajo Holzmann

Contents

Introduction (English)

Markov chains are stochastic processes with certain distribution characteristics, that are used in various fields of research and modeling. Applications include automatic speech recognition, valuation of web pages by search engines, algorithmic music composition and, as will be discussed more intensively in this bachelor thesis, forecasting of the credit-worthiness of a debtor.

Markov chains operate on spaces of so-called *states*. In this thesis, only processes with finitely many states will be discussed. Markov chains can be divided into processes in *discrete* and *continuous time*. The defining characteristic of a discrete-time Markov chain is its transition matrix P, which specifies the probability of changing between the states of the process in a given time unit. The natural powers of P allow prediction of the behavior of the chain over integer multiples of this time unit. The construction of a continuous-time Markov chain is based on a so-called generator matrix Q. Using power series expansions, we can generalize the exponential function to matrices and make sense of the expression $P(t) := \exp(tQ)$ for $t \geq 0$. This notion will be used to define the transition probabilities of a Markov chain in continuous time.

Hence, sufficient knowledge of the generator Q enables us to forecast the course of the process over any given time horizon. The resulting increase in flexibility of prediction is highly valuable to many practical applications. To profit from this advantage, one tries to *embed* a discrete- into a continuous-time Markov chain. This means that we assume a given transition matrix P of a process in discrete time to be $P(1)$ in a continuous-time model. We will then have to solve the central equation

$$P = \exp(Q) \ ,$$

where P and Q are matrices with certain properties.

This identity motivates us to search for an inverse to the matrix exponential, a *matrix logarithm*. We approach this task by drawing on the theory of the real and complex logarithm function. The naive technique of substituting a matrix symbolically for a complex variable does not provide a suitable framework for this idea. Therefore, we will resort to a concept of greater scope, the so-called *holomorphic functional calculus*. Briefly worded, we will explain the application of functions to matrices by following the well-known *Cauchy integral formula* from complex analysis, which states that

$$f(z) = \int_\Gamma \frac{f(\zeta)}{\zeta - z} \, d\zeta$$

for a *holomorphic* function f, a complex argument z, an appropriately defined *line integral* and a suitable *cycle* Γ.

It is a priori not clear, whether the central equation will be solvable at all. Hence, we will concern ourselves with the task of finding necessary and sufficient conditions for the existence of an exact or approximate solution, which became known as the *embedding problem.* Furthermore, we do not know from the outset, if a solution to the identity will be unique. Thus, in case multiple different solutions exist, we will have to discuss which is the "correct" one for our application. This question is commonly referred to as the *identification problem.*

We will use the results of these investigations to formulate algorithms for solving the above equation exactly or approximately. In the end, the developed methods will be tested and compared with the help of numerical examples from the real credit rating process.

The main source for this thesis is the paper [IRW], where certain results about the holomorphic functional calculus are assumed to be known and, building on this, a great part of the theory for the subsequent chapters is developed.

Introduction (German)

Markov-Ketten sind stochastische Prozesse mit bestimmten Verteilungseigenschaften, die in verschiedenen Gebieten der Forschung und Modellierung verwendet werden. Anwendungsbeispiele umfassen automatische Spracherkennung, Bewertung von Internetseiten durch Suchmaschinen, algorithmische Musikkomposition und, wie in dieser Bachelorarbeit intensiv diskutiert wird, Prognostizierung der Kreditwürdigkeit eines Schuldners.

Markov-Ketten operieren auf sogenannten *Zustandsräumen*. In dieser Arbeit werden nur Prozesse mit endlichem Zustandsraum betrachtet. Markov-Ketten können in Prozesse in *diskreter* und *stetiger Zeit* unterteilt werden. Das definierende Charakteristikum einer zeitdiskreten Markov-Kette ist ihre Übergangsmatrix P, welche die Wahrscheinlichkeit angibt, innerhalb einer gegebenen Zeiteinheit zwischen den *Zuständen* des Prozesses zu wechseln. Die natürlichen Potenzen von P erlauben eine Vorhersage der Entwicklung des Prozesses über ganzzahlige Vielfache der Zeiteinheit. Die Konstruktion einer zeitstetigen Markov-Kette basiert auf einer sogenannten *Erzeugermatrix* Q. Mithilfe von Potenzreihendarstellungen lässt sich die Exponentialfunktion auf Matrizen verallgemeinern und daher sinnvoll von dem Ausdruck $P(t) := \exp(tQ)$ für $t \geq 0$ sprechen. Dieser kann dann verwendet werden, um die Übergangswahrscheinlichkeiten einer Markov-Kette in stetiger Zeit zu definieren.

Somit gestattet uns eine hinreichende Kenntnis des Erzeugers Q, den Verlauf des Prozesses über einen beliebigen Zeithorizont vorauszusagen. Die dadurch gesteigerte Flexibilität bei der Vorhersage stellt für viele praktische Anwendungen eine wertvolle Bereicherung dar. Um von diesem Vorteil zu profitieren, versucht man die zeitdiskrete in eine zeitstetige Markov-Kette *einzubetten*. Dies bedeutet, dass man die Übergangsmatrix P eines Prozesses in diskreter Zeit als $P(1)$ in einem zeitstetigen Modell annimmt. Es gilt dann, die zentrale Gleichung

$$P = \exp(Q)$$

zu lösen, wobei P und Q Matrizen mit bestimmten Eigenschaften bezeichnen.

Das motiviert uns, nach einer Umkehrfunktion zum Matrixexponential zu suchen, einem *Matrixlogarithmus*. Wir wollen diese Aufgabe dadurch angehen, dass wir auf die Theorie der reellen und komplexen Logarithmusfunktion zurückgreifen. Die naive Technik, eine Matrix symbolisch für eine komplexe Variable zu substituieren, bietet keinen geeigneten Rahmen für diese Idee. Deshalb bedienen wir uns eines Konzeptes mit weitreichenderer Anwendungsmöglichkeit, des sogenannten *holomorphen Funktionalkalküls*. Kurz gefasst werden wir die Anwendung von Funktionen auf Matrizen in Anlehnung an die bekannte *Cauchy'sche Integralformel* aus der Funktionentheorie erklären, welche besagt, dass für eine *holomorphe* Funktion f, ein komplexes Argument z, ein entsprechend definiertes *Kurvenintegral* und einen geeigneten *Zyklus* Γ gilt:

$$f(z) = \int_\Gamma \frac{f(\zeta)}{\zeta - z} \, d\zeta \ .$$

Es ist a priori nicht klar, ob die zentrale Gleichung überhaupt lösbar sein wird. Daher werden wir uns mit der Aufgabe beschäftigen, notwendige und hinreichende Bedingungen für die Existenz einer exakten oder näherungsweisen Lösung zu finden, welche als *Einbettungsproblem* bekannt geworden ist. Weiterhin können wir im Vorfeld nicht von der Eindeutigkeit einer Lösung ausgehen. Daher muss im Falle der Existenz mehrerer Lösungen die Frage geklärt werden, welche darunter die "richtige" für unsere Anwendung ist. Diese Thematik bezeichnet man üblicherweise als *Identifikationsproblem*.

Die Ergebnisse unserer Nachforschungen können dann genutzt werden, um Algorithmen zur exakten oder näherungsweisen Lösung der obigen Gleichung zu formulieren. Zum Schluss werden wir die dort entwickelten Methoden anhand von numerischen Beispielen des realen Kredit-Ratingprozesses testen und vergleichen.

Die Hauptquelle für diese Bachelorarbeit ist die Publikation [IRW], in der bestimmte Resultate über den holomorphen Funktionalkalkül als bekannt vorausgesetzt werden und darauf aufbauend ein Großteil der Theorie der nachfolgenden Kapitel entwickelt wird.

Chapter 1

Holomorphic functional calculus

1.1 Preliminaries

As announced in the introduction, the first chapter will be dedicated to the holomorphic functional calculus. This is a technique of applying functions of a complex argument to other objects, namely to elements of so-called *Banach algebras*. We will see later that this repertoire comprises the square matrices, which we are interested in.

To formulate some basic definitions and results about Banach algebras, let $|.|$ denote the complex absolute value and write $B_r(z_0) := \{z \in \mathbb{C} \mid |z-z_0| < r\}$ for the open ball of radius $r > 0$, centered at $z_0 \in \mathbb{C}$.

Definition 1.1.1 (Banach space)**.** A $\mathbb{C}$-*Banach space* is a complete, normed $\mathbb{C}$-vector space. More explicitly, this means that, if X is a set and $+ : X \times X \to X$, $\cdot : \mathbb{C} \times X \to X$ and $\|.\|_X : X \to [0, \infty)$, then we call $(X, +, \cdot, \|.\|_X)$ a $\mathbb{C}$-Banach space if:

(A1) *Associativity*: $x_1 + (x_2 + x_3) = (x_1 + x_2) + x_3 \quad \forall\, x_1, x_2, x_3 \in X$.

(A2) *Identity element*: $\exists\, 0_X \in X \;\forall\, x \in X :\; x + 0_X = x = 0_X + x$.

(A3) *Inverse element*: $\forall\, x \in X \;\exists - x \in X :\; x + (-x) = 0_X = (-x) + x$.

(A4) *Commutativity*: $x_1 + x_2 = x_2 + x_1 \quad \forall\, x_1, x_2 \in X$.

(S1) *Distributive law for vectors*: $\alpha \cdot (x_1 + x_2) = \alpha \cdot x_1 + \alpha \cdot x_2 \quad \forall\, \alpha \in \mathbb{C},\, x_1, x_2 \in X$.

(S2) *Distributive law for scalars*: $(\alpha + \beta) \cdot x = \alpha \cdot x + \beta \cdot x \quad \forall\, \alpha, \beta \in \mathbb{C},\, x \in X$.

(S3) *Associativity*: $\alpha \cdot (\beta \cdot x) = (\alpha\beta) \cdot x \quad \forall\, \alpha, \beta \in \mathbb{C},\, x \in X$.

(S4) *Identity element*: $1_{\mathbb{C}} \cdot x = x \quad \forall\, x \in X$.

(N1) *Definiteness*: $\|x\|_X = 0_{\mathbb{R}} \iff x = 0_X$.

(N2) *Homogeneity*: $\|\alpha \cdot x\|_X = |\alpha| \|x\|_X \quad \forall\, \alpha \in \mathbb{C},\, x \in X$.

(N3) *Triangle inequality*: $\|x_1 + x_2\|_X \leq \|x_1\|_X + \|x_2\|_X \quad \forall\, x_1, x_2 \in X$.

(N4) *Completeness*: Every sequence in X which is a *Cauchy-sequence* with respect to $\|.\|_X$ converges.

Definition 1.1.2 (Banach algebra)**.** A *unital* $\mathbb{C}$*-Banach algebra* is a $\mathbb{C}$-Banach space together with a structure of an *unital algebra over* $\mathbb{C}$, where the multiplication is continuous. So, if $\mathcal{A}$ is a set and $+ : \mathcal{A}\times\mathcal{A} \to \mathcal{A}$, $\cdot : \mathbb{C}\times\mathcal{A} \to \mathcal{A}$, $\circ : \mathcal{A}\times\mathcal{A} \to \mathcal{A}$ and $\|.\|_{\mathcal{A}} : \mathcal{A} \to [0,\infty)$, then we say that $(\mathcal{A}, +, \cdot, \circ, \|.\|_{\mathcal{A}})$ is a (unital) $\mathbb{C}$-Banach algebra, if $(\mathcal{A}, +, \cdot, \|.\|_{\mathcal{A}})$ satisfies (A1)-(A4), (S1)-(S4) and (N1)-(N4) and additionally:

(M1) *Associativity*: $(A \circ B) \circ C = A \circ (B \circ C) \quad \forall\, A, B, C \in \mathcal{A}$.

(M2) *Additivity*: $A \circ (B + C) = A \circ B + A \circ C$, $(A + B) \circ C = A \circ C + B \circ C \quad \forall\, A, B, C \in \mathcal{A}$.

(M3) *Homogeneity*: $\alpha \cdot (A \circ B) = (\alpha \cdot A) \circ B = A \circ (\alpha \cdot B) \quad \forall\, \alpha \in \mathbb{C},\, A, B \in \mathcal{A}$.

(M4) *Identity element*: $\exists\, I_{\mathcal{A}} \in \mathcal{A}\, \forall\, A \in \mathcal{A} :\; A \circ I_{\mathcal{A}} = A = I_{\mathcal{A}} \circ A$.

(N5) *Submultiplicativity*: $\|A \circ B\|_{\mathcal{A}} \leq \|A\|_{\mathcal{A}} \|B\|_{\mathcal{A}} \quad \forall A, B \in \mathcal{A}$.

(N6) *Identity element*: $\|I_{\mathcal{A}}\|_{\mathcal{A}} = 1$.

Since we will not be discussing Banach spaces or algebras over any other field than the complex numbers, we omit the prefix "$\mathbb{C}$-" most often. The same applies to the attribute "unital" for Banach algebras. For convenience, we often leave out the operators for scalar and algebra multiplication, as in $\alpha A := \alpha \cdot A$, $AB := A \circ B$. Also, we will sometimes write the scalar multiplication with inverted order of the arguments, i. e. $A\alpha := \alpha A$. When the operations on the set are clear, we often refer to $\mathcal{A}$ as the Banach algebra instead of $(\mathcal{A}, +, \cdot, \circ, \|.\|)$. Also, if we have fixed a single Banach algebra $\mathcal{A}$, we write the identity as $I := I_{\mathcal{A}}$ and the norm as $\|.\| := \|.\|_{\mathcal{A}}$.

Remark 1.1.3 (Operator algebra)**.** In functional analysis, Banach algebras are typically not taken as given, but instead, they are constructed from Banach spaces by the following means: let $(X, +, \cdot, \|.\|_X)$ be a $\mathbb{C}$-Banach space. We call a linear mapping $L : X \to X, x \mapsto Lx$ a *linear operator on* X. We say that L is *bounded*, if there exists $M > 0$ such that

$$\|Lx\|_X \leq M \|x\|_X \quad \forall\, x \in X \; . \tag{1.1}$$

Notice that a bounded linear operator is in general not a bounded function, but rather a so-called *locally bounded function*. Write $B(X)$ for the set of all bounded linear operators on X become. Now, we want to define suitable operations on $B(X)$, so that this set becomes a Banach algebra. First, the Banach space structure of X allows us to introduce natural pointwise addition "$\oplus$" and scalar multiplication "$\odot$", i. e. $(L_1 \oplus L_2)x := L_1x + L_2x$, $(\alpha \odot L)x := \alpha \cdot Lx$. We define the algebra multiplication on $B(X)$ as the composition "$\circ$" given by $(L_1 \circ L_2)x := L_1(L_2x)$. Finally, we set the norm $\|.\|_{B(X)}$ by $\|L\|_{B(X)} := \inf\{M > 0 \mid \|Lx\|_X \leq M\|x\|_X \ \forall x \in X\}$, where condition (1.1) ensures that $\|L\|_{B(X)}$ is finite for every $L \in B(X)$. It can be shown that with these definitions, $B(X)$ becomes a Banach algebra, called the *algebra of bounded linear operators on* X or, more concisely, the *operator algebra of* X.
Applying this procedure to the Banach space of n-dimensional complex column vectors $\mathbb{C}^n$ gives rise to an operator algebra $B(\mathbb{C}^n)$ which can easily be identified with the algebra of square matrices with complex entries $\mathbb{C}^{n\times n}$, that we are interested in. To keep the theory a bit more general, however, we consider not just operators algebras $B(X)$ constructed from Banach spaces X, but any given Banach algebras $\mathcal{A}$.

From now on for the rest of this chapter, let $\mathcal{A}$ be a fixed Banach algebra. The term "invertible" shall always be meant with respect to the algebra multiplication. The set of all invertible elements of $\mathcal{A}$ forms a group and so we can import some basic results from algebra. For example, we know that, in case of existence, the inverse is unique (we denote it by A^{-1} for an invertible $A \in \mathcal{A}$) or that $(AB)^{-1} = B^{-1}A^{-1}$ if $A, B \in \mathcal{A}$ are invertible. As long as we de not violate non-commutativity of the algebra multiplication, we can also conveniently write $\frac{1}{A}$ for the inverse of A.

Definition 1.1.4 (Spectrum and resolvent)**.** Let $A \in \mathcal{A}$. The set $\sigma(A) := \{\lambda \in \mathbb{C} \mid (\lambda I - A)$ is not invertible$\}$ is called the *spectrum of* A. The complement $\rho(A) := \mathbb{C} \setminus \sigma(A)$ of the spectrum is referred to as the *resolvent set of* A. For $z \in \rho(A)$, we call $(zI - A)^{-1}$ the *resolvent of* A *at* z and $R_A : \rho(A) \to \mathcal{A}, z \to (zI - A)^{-1}$ the *resolvent mapping of* A.

For operator algebras $B(X)$ of finite-dimensional vector spaces X, the spectrum of an element coincides with the set of *eigenvalues* of that element according to the definition of classical linear algebra (i. e. $\lambda \in \mathbb{C}$ is an eigenvalue of $L \in B(X)$, if there exists an *eigenvector* $x \neq 0_X$ such that $Lx = \lambda x$). In the infinite-dimensional case, we still find that every eigenvalue is an element of the spectrum, but the converse may become false. The most commonly given counterexample for this (see e. g. [Swan]) is the Banach-space ℓ^1 of *sequences whose series converge absolutely* together with the right-shift operator $R : \ell^1 \to \ell^1, (x_1, x_2, \ldots) \mapsto (0, x_1, x_2, \ldots)$ and the complex number 0.

Lemma 1.1.5 (First resolvent identity)**.** For every $A \in \mathcal{A}$ and $z_1, z_2 \in \rho(A)$

$$(z_1 I - A)^{-1} - (z_2 I - A)^{-1} = (z_1 I - A)^{-1}(z_2 - z_1)(z_2 I - A)^{-1} \ .$$

Proof. Multiply both sides of $(z_2 I - A) - (z_1 I - A) = (z_2 I - z_1 I)$ with $(z_1 I - A)^{-1}$ from the left and $(z_2 I - A)^{-1}$ from the right. $\square$

Definition 1.1.6 (Holomorphy and analyticity)**.** Let $D_f \subseteq \mathbb{C}$ and $f : D_f \to \mathcal{A}$. We call f *holomorphic on an open set* $U \subset D_f$, if the limit

$$f'(z_0) := \lim_{z \to z_0} (z - z_0)^{-1}(f(z) - f(z_0))$$

exists (with respect to $\|.\|$) for every $z_0 \in U$. We say that f is *holomorphic at* $z_0 \in D_f$, if there exists an open neighborhood $V \subseteq D_f$ of z_0 such that f is holomorphic on V.
We call f analytic at $z_0 \in D_f$, if there exists an open neighborhood $V \subseteq D_f$ of z_0 and a sequence of elements of the Banach Algebra $(a_n)_{n \in \mathbb{N}_0} \subseteq \mathcal{A}$ such that $\sum_{n=0}^{\infty} a_n (z - z_0)^n$ converges to $f(z)$ (with respect to $\|.\|$) for every $z \in V$. We say that f is analytic on an open set $U \subseteq D_f$, if f is analytic at every $z_0 \in U$.

We will refer to functions taking values in a Banach space or algebra as vector-valued functions. Just like in the complex case, every vector-valued holomorphic function is continuous. Similarly, a generalization of the well-known *Cauchy's integral formula* can be used to show that a vector valued function is holomorphic on an open set if and only if it is analytic on that set and that the coefficients for the power series expansion are given by *Taylor's theorem*.

To prepare for the next proposition, acknowledge this short definition: for $A \in \mathcal{A}$, the term $\sum_{n=0}^{\infty} A^n$ (note that this expression makes sense in a Banach algebra) is called the *Neumann series of* A. The properties of this series resemble those of the *geometric series* for real numbers closely in terms of the symbolism: it can be shown that $\sum_{n=0}^{\infty} A^n$ converges, if $\|A\| < 1$, and that, in case of convergence, the limit is $(I - A)^{-1}$. It has also been found out that for $\|A\| < 1$

$$\|(I - A)^{-1}\| \leq \frac{1}{1 - \|A\|} \ . \tag{1.2}$$

Proposition 1.1.7 (Properties of spectrum and resolvent)**.** Let $A \in \mathcal{A}$.

(*i*) The resolvent set $\rho(A)$ is open and the resolvent mapping R_A is holomorphic on $\rho(A)$.

(*ii*) The spectrum $\sigma(A)$ is a non-empty, compact subset of the complex plane $\mathbb{C}$.

Proof. We follow the standard proof of this result in functional analysis (see e. g. [Garr]):

(i): Fix $z_0 \in \rho(A)$ and let $z \in \mathbb{C}$. We set $a_n := (-1)^n((z_0 I - A)^{-1})^{n+1} \in \mathcal{A}$ and consider the series

$$\sum_{n=0}^{\infty} a_n (z - z_0)^n = (z_0 I - A)^{-1} \sum_{n=0}^{\infty} ((z_0 - z)(z_0 I - A)^{-1})^n \ .$$

If $|z - z_0| < 1/\|(z_0 I - A)^{-1}\| =: \varepsilon$, then (N2) implies $\|(z_0 - z)(z_0 I - A)^{-1}\| < 1$ and thus, the right-hand side converges to

$$(z_0 I - A)^{-1}(I - (z_0 - z)(z_0 I - A)^{-1})^{-1} = ((I - (z_0 - z)(z_0 I - A)^{-1})(z_0 I - A))^{-1} = (zI - A)^{-1} \ .$$

This implies that $B_\varepsilon(z_0) \subseteq \rho(A)$, so $\rho(A)$ is open, and $R_A(z) = \sum_{n=0}^{\infty} a_n (z - z_0)^n$ for every $z \in B_\varepsilon(z_0)$, so R_A is holomorphic on $\rho(A)$.

(ii): Because of the *Heine-Borel theorem*, we can prove compactness by showing that $\sigma(A)$ is bounded and closed. The latter follows from (i), since $\sigma(A) = \mathbb{C} \setminus \rho(A)$. To see that $\sigma(A)$ is bounded by $\|A\|$, notice that $|z| > \|A\|$ implies $\|Az^{-1}\| < 1$ and thus

$$z^{-1} \sum_{n=0}^{\infty} (Az^{-1})^n = z^{-1}(I - Az^{-1})^{-1} = ((I - Az^{-1})z)^{-1} = (zI - A)^{-1} \ ,$$

meaning $z \notin \sigma(A)$. Finally, if $\sigma(A)$ was empty, then R_A would be holomorphic on $\mathbb{C}$, a so-called *entire function*. Furthermore, R_A is bounded, as it is continuous wherever $|z| \leq \|A\|$, and, for $|z| > \|A\|$, (M5) and the inequality (1.2) yield

$$\begin{aligned} \|R_A(z)\| &= \|z^{-1}(I - Az^{-1})^{-1}\| = |z|^{-1}\|(I - Az^{-1})^{-1}\| \\ &\leq \frac{|z|^{-1}}{1 - \|A\||z|^{-1}} = \frac{1}{|z| - \|A\|} \to 0, \quad |z| \to \infty \ . \end{aligned}$$

But then, the vector-valued version of *Liouville's theorem* says that R_A must be constant and, as $R_A(z) \to 0_{\mathcal{A}}$, $|z| \to \infty$, even $R_A \equiv 0_{\mathcal{A}}$, which obviously is a contradiction.

□

1.2 Integration in Banach spaces and algebras

It was mentioned in the introduction that the formulation of the holomorphic functional calculus will involve a notion of a *line integral* for vector-valued functions. There are many different methods for constructing such an integral; most of them involve mechanisms which we already know from the real- or complex-valued case. For this thesis, I chose a generalization of the *Riemann-Stieltjes integral*, because it is intuitive to handle and allows easy proofs of the important results. The following definitions lay the groundwork for this theory:

Let $\mathcal{A}$ be a given $\mathbb{C}$-Banach algebra with norm $\|.\|$. If not declared otherwise, assume $m, n, M, N \in \mathbb{N}, a, b, c, d \in \mathbb{R}$, $a < b, c < d, D, D' \subseteq \mathbb{C}$, let $|.|$ denote the complex absolute value and write $\operatorname{dist}(D, D') := \inf_{z \in D, z' \in D'} |z - z'|$ for the distance of D and D'.

Definition 1.2.1 (Partition)**.** An $(n+1)$-tuple $P = (t_0, \ldots, t_n)$ is called a *partition of the interval* $[a, b]$ if $a = t_0 < t_1 < \ldots < t_n = b$. If $P = (t_0, \ldots, t_n)$ and $Q = (s_0, \ldots, s_m)$ are partitions of the same interval, and for every $i \in \{1, \ldots, n\}$ there exists $j_i \in \{1, \ldots, m\}$ such that $t_i = s_{j_i}$, then Q is called a *refinement of the partition* P. Using the same notation as above, we understand the *union* $P \cup Q$ *of two partitions* P *and* Q as the tuple obtained from ordering the elements of the set $\{t_0, \ldots, t_n\} \cup \{s_0, \ldots, s_n\}$ (the essential aspect here is that elements which are found in both P and Q, only appear once in $P \cup Q$). The tuple $P \cup Q$ is always a partition as well as a refinement of both P and Q and is referred to as the *common refinement of* P *and* Q. We set the *norm of a partition* P as $\|P\| := \max_{i=1,\ldots,n} t_i - t_{i-1}$.
We call an n-tuple $T = (\tau_1, \ldots, \tau_n)$ *tags for the partition* P, if $t_{i-1} \leq \tau_i \leq t_i,\ i = 1, \ldots, n$. We will sometimes rephrase this by saying that T *tags* P. We name $T(P) := (\tau_1(P), \ldots, \tau_n(P))$, where $\tau_i(P) := (t_{i-1} + t_i)/2,\ i = 1, \ldots, n$, the *centered tags for the partition* P.
We call a sequence of partitions $(P_n)_{n \in \mathbb{N}}$ of the same interval *admissible*, if $\|P_n\| \to 0$, $n \to \infty$. We refer to $(T_n)_{n \in \mathbb{N}}$ as *corresponding tags for* $(P_n)_{n \in \mathbb{N}}$, if T_n tags P_n for every $n \in \mathbb{N}$.
Finally, for functions $g : [a, b] \to \mathcal{A}$, $\gamma : [a, b] \to \mathbb{C}$, a partition $P = (t_0, \ldots, t_n)$ and tags $T = (\tau_1, \ldots, \tau_n)$ for P, we define the corresponding *Riemann-Stieltjes sum* as

$$S(g, \gamma, P, T) := \sum_{i=1}^{n} g(\tau_i)(\gamma(t_{i-1}) - \gamma(t_i)) \ . \tag{1.3}$$

Definition 1.2.2 (Riemann-Stieltjes integral)**.** Let $\gamma : [a, b] \to \mathbb{C}$. We call a function $g : [a, b] \to \mathcal{A}$ *Riemann-Stieltjes integrable with respect to* γ if there exists an element $I(g, \gamma) \in \mathcal{A}$ of the Banach algebra satisfying the following condition: for every $\varepsilon > 0$ there exists a partition P_ε such that for every refinement P of P_ε and any tags T for P we find

$$\|S(g, \gamma, P, T) - I(g, \gamma)\| < \varepsilon \ .$$

If g is Riemann-Stieltjes integrable with respect to γ, we write

$$\int_a^b g(t)\ d\gamma(t) := I(g, \gamma) \ .$$

To see that the Riemann-Stieltjes integral is well-defined, assume $I(g, \gamma), J(g, \gamma) \in \mathcal{A}$, any $\varepsilon > 0$ and partitions $P_{\varepsilon/2}, Q_{\varepsilon/2}$ satisfying the criterion of the definition. Then, $P_{\varepsilon/2} \cup Q_{\varepsilon/2}$ is a refinement of $P_{\varepsilon/2}$ and hence $\|S(g, \gamma, P_{\varepsilon/2} \cup Q_{\varepsilon/2}, T(P_{\varepsilon/2} \cup Q_{\varepsilon/2})) - I(g, \gamma)\| < \varepsilon/2$. Apply the same reasoning to $J(g, \gamma)$ and $Q_{\varepsilon/2}$ and find $\|I(g, \gamma) - J(g, \gamma)\| < \varepsilon$.

Definition 1.2.3 (Curve)**.** We call a continuous function $\gamma : [a, b] \to D$ a *curve (in D)*. We denote the image of a curve γ as $[\gamma]$. If $\gamma : [a, b] \to D$ is a curve, $\phi : [c, d] \to [a, b]$ is a strictly increasing, surjective function (note that such a mapping is automatically continuous and bijective) and $\gamma' = \gamma \circ \phi : [c, d] \to D$, then we call γ' a *reparameterization of* γ *to the interval* $[c, d]$. Every curve $\gamma : [a, b] \to D$ can always be reparameterized to any interval $[c, d]$ by means of

$$\phi : [c, d] \to [a, b],\ t \mapsto \frac{(b-a)t + ad - bc}{d - c} \ .$$

When speaking about a curve $\gamma : [a, b] \to D$, we refer to $\gamma(a)$ as the *initial* and $\gamma(b)$ as the *terminal point.* If $\gamma_1 : [a, b] \to D$ and $\gamma_2 : [b, c] \to D$ are two curves such that the terminal point of γ_1 coincides with the initial point of γ_2, we can define a curve $\gamma_1 + \gamma_2$ called the *sum of* γ_1 *and* γ_2 by setting

$$\gamma_1 + \gamma_2 : [a, c] \to D,\ t \mapsto \begin{cases} \gamma_1(t), & a \le t \le b \\ \gamma_2(t), & b < t \le c \end{cases} \ .$$

If $b \neq c$, $\gamma_2 : [c, d] \to D$ and $\gamma_1(b) = \gamma_2(c)$, set $\gamma_1 + \gamma_2 := \gamma_1 + \gamma_2'$, where γ_2' is a reparameterization of γ_2 to the interval $[b, c]$. Also, we define *inverse of a curve* $\gamma : [a, b] \to D$ as

$$-\gamma : [a, b] \to D, t \mapsto \gamma(a + b - t) \ .$$

We say that a curve γ is *piecewise continuously differentiable (or piecewise C^1)*, if there exists a finite number of points $a = a_0 < a_1 < \ldots < a_n = b$ such that $\gamma|_{[a_{i-1},a_i]}$ is continuously differentiable for every $i \in \{1, \ldots, n\}$. We define the *length of a curve* γ as

$$L(\gamma) := \sup_{a=t_0<t_1<\ldots<t_n=b} \sum_{i=1}^{n} |\gamma(t_i) - \gamma(t_{i-1})| \in [0, \infty]$$

and call γ *rectifiable* if $L(\gamma) < \infty$.
A curve γ is said to be *simple* if for all $s, t \in [a, b]$ the conditions $\gamma(s) = \gamma(t)$ and $(s, t) \notin \{(a, b), (b, a)\}$ together imply $s = t$ and *closed* if $\gamma(a) = \gamma(b)$. A simple, closed curve is called a *Jordan curve.*

It can be shown that every piecewise C^1 curve is rectifiable.

Definition 1.2.4 (Chain)**.** If $\gamma_1, \ldots, \gamma_n$ are a finite number of curves in D and $k_1, \ldots, k_n \in \mathbb{N}$, we call a formal linear combination

$$\Gamma = k_1\gamma_1 \oplus \ldots \oplus k_n\gamma_n$$

a *chain (in D).* The image of a chain Γ is given by $[\Gamma] := \bigcup_{i=1}^{n} [\gamma_i]$. We understand an addition "$\oplus$" on the set of chains in D as the addition of the coefficients of the linear combinations, inserting zeros, where necessary. The *inverse of a chain* Γ is defined as $-\Gamma := k_1(-\gamma_1) \oplus \ldots \oplus k_n(-\gamma_n)$. Set the *length of a chain* Γ to be

$$L(\Gamma) = \sum_{i=1}^{n} k_i L(\gamma_i) \ .$$

When $\gamma_1, \ldots, \gamma_n$ are closed, we call Γ a *cycle (in D).*

I remark that in the algebraic field of *homology theory*, chains are usually defined a little differently, allowing the coefficients to be negative integers as well in order to generate a group structure. However, this would have not provided any additional value for our theory and could have led to ambiguous notation, which justifies the choice of the above definition. We can now establish one of the main results of this section:

Theorem & Definition 1.2.5 (Line integral)**.** Assume $D_f \subseteq \mathbb{C}$, a continuous function $f : D_f \to \mathcal{A}$ and a rectifiable curve $\gamma : [a, b] \to D_f$ in D_f. Then, the function $g := f \circ \gamma : [a, b] \to \mathcal{A}$ is Riemann-Stieltjes integrable with respect to γ and

$$\int_a^b g(t)\ d\gamma(t) = \lim_{n\to\infty} S(g, \gamma, P_n, T_n) \tag{1.4}$$

for every admissible sequence of partitions $(P_n)_{n\in\mathbb{N}}$ and any corresponding tags $(T_n)_{n\in\mathbb{N}}$. In this case, we define the *line integral of f along γ* as

$$\int_\gamma f(z)dz := \int_a^b (f\circ\gamma)(t)\ d\gamma(t)\ .$$

We generalize the line integral to chains $\Gamma = k_1\gamma_1 \oplus \ldots \oplus k_n\gamma_n$ of rectifiable curves by

$$\int_\Gamma f(z)\ dz := \sum_{i=1}^n k_i \int_{\gamma_i} f(z)\ dz\ .$$

Proof. We follow the approach seen in [Bals, p. 16f.]. First, let us do some preliminary calculations: Assume $\varepsilon > 0$. As $f : D \to \mathcal{A}$ and $\gamma : [a,b] \to D$ are continuous, so is $f\circ\gamma : [a,b] \to \mathcal{A}$. By virtue of a slight generalization of *Heine's theorem*, we find that $f\circ\gamma$ is even uniformly continuous. As γ is rectifiable, we obtain a real number δ such that $\sigma, \tau \in [a,b], |\tau - \sigma| < \delta \implies \|f(\gamma(\tau)) - f(\gamma(\sigma))\| < \varepsilon/(2L(\gamma))$. Let $P_\varepsilon = (t_0, \ldots, t_n)$ be any partition of $[a,b]$ such that $\|P_\varepsilon\| < \delta$ and write $T_\varepsilon = (\tau_1, \ldots, \tau_n)$ for arbitrary tags corresponding to P_ε. Assume additionally a refinement $P = (s_0, \ldots, s_m)$ of P_ε and tags $T = (\sigma_1, \ldots, \sigma_m)$ for P. For every $i \in \{1, \ldots, n\}$ we find $j_i, k_i \in \{1, \ldots, m\}$ such that $t_{i-1} = s_{j-1} < s_j < \ldots < s_k = t_i$. We obtain

$$\begin{aligned}\left\| f(\gamma(\tau_i)(\gamma(t_i) - \gamma(t_{i-1})) - \sum_{l=j_i}^{k_i} f(\gamma(\sigma_l)(\gamma(s_l) - \gamma(s_{l-1}))\right\| &= \left\| \sum_{l=j_i}^{k_i} (f(\gamma(\tau_i)) - f(\gamma(\sigma_l))(\gamma(s_l) - \gamma(s_{l-1}))\right\| \\ &\le \sum_{l=j_i}^{k_i} \|(f(\gamma(\tau_i)) - f(\gamma(\sigma_l))\|\,|(\gamma(s_l) - \gamma(s_{l-1}))| \\ &< \frac{\varepsilon}{2L(\gamma)} L(\gamma|_{[t_{i-1},t_i]})\ .\end{aligned}$$

Convince yourself that the curve length is additive and deduce

$$\begin{aligned}\|S(f\circ\gamma, \gamma, P, T) - S(f\circ\gamma, \gamma, P_\varepsilon, T_\varepsilon)\| &\le \sum_{i=1}^n \left\| f(\gamma(\tau_i)(\gamma(t_i) - \gamma(t_{i-1})) - \sum_{l=j_i}^{k_i} f(\gamma(\sigma_l)(\gamma(s_l) - \gamma(s_{l-1}))\right\| \\ &< \frac{\varepsilon}{2L(\gamma)} \left(L(\gamma|_{[t_0,t_1]}) + \ldots + L(\gamma|_{[t_{n-1},t_n]})\right) = \frac{\varepsilon}{2}\ .\end{aligned}$$

Now, we can establish convergence of each sequence of Riemann-Stieltjes sums:

To this end, let $(P_n)_{n\in\mathbb{N}}$ be any admissible sequence of partitions and $(T_n)_{n\in\mathbb{N}}$ any corresponding tags. Since $\mathcal{A}$ is complete, it suffices to show that $(S(f\circ\gamma, \gamma, P_n, T_n))_{n\in\mathbb{N}} \subseteq \mathcal{A}$ is a Cauchy sequence. To do so, choose $n_0 \in \mathbb{N}$ such that $\|P_n\| < \delta \,\forall\, n \geq n_0$ and assume $m, n \geq n_0$. Since $P_m \cup P_n$ is a refinement of both P_m and P_n, the above calculations yield

$$\|S(f \circ \gamma, \gamma, P_m, T_m) - S(f \circ \gamma, \gamma, P_m \cup P_n, T(P_m \cup P_n))\| < \frac{\varepsilon}{2} ,$$
$$\|S(f \circ \gamma, \gamma, P_m \cup P_n, T(P_m \cup P_n)) - S(f \circ \gamma, \gamma, P_n, T_n)\| < \frac{\varepsilon}{2} .$$

Thus, we can conclude

$$\begin{aligned}\|S(f \circ \gamma, \gamma, P_m, T_m) - S(f \circ \gamma, \gamma, P_n, T_n)\| &\leq \|S(f \circ \gamma, \gamma, P_n, T_n) - S(f \circ \gamma, \gamma, P_m \cup P_n, T(P_m \cup P_n))\| \\ &\quad + \|S(f \circ \gamma, \gamma, P_m \cup P_n, T(P_m \cup P_n)) - S(f \circ \gamma, \gamma, P_m, T_m)\| \\ &< \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon .\end{aligned}$$

We finish the proof by deducing Riemann-Stieltjes integrability of $f \circ \gamma$:
Assume $\varepsilon, \delta, (P_n)_{n\in\mathbb{N}}, n_0$ and $(T_n)_{n\in\mathbb{N}}$ as above. Write $I(f \circ \gamma, \gamma)$ for the limit of $(S(f \circ \gamma, \gamma, P_n, T_n))_{n\in\mathbb{N}}$, choose $n_1 \in \mathbb{N}$ such that $\|S(f \circ \gamma, \gamma, P_n, T_n) - I(f \circ \gamma, \gamma)\| < \varepsilon/2 \,\forall\, n \geq n_1$ and set $P_\varepsilon := P_{max\{n_0,n_1\}}$. Finally, it follows that for any refinement P of P_ε and any tags T for P

$$\begin{aligned}\|S(f \circ \gamma, \gamma, P, T) - I(f \circ \gamma, \gamma)\| &\leq \|S(f \circ \gamma, \gamma, P, T) - S(f \circ \gamma, \gamma, P_\varepsilon, T_\varepsilon)\| \\ &\quad + \|S(f \circ \gamma, \gamma, P_\varepsilon, T_\varepsilon) - I(f \circ \gamma, \gamma)\| \\ &< \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon .\end{aligned}$$

□

Remark 1.2.6 (Line integral)**.** We will see later that, in the situation of the holomorphic functional calculus, we will always integrate along piecewise C^1 curves. Hence, we could have chosen a different approach to the above definition that looks more familiar: we could have set the line integral along a continuously differentiable curve γ as $\int_\gamma f(z)\,dz := \int_a^b f(\gamma(t))\gamma'(t)\,dt$, where the right-hand side denotes the Riemann-Stieltjes integral with respect to the identity $\mathrm{id}_{[a,b]}$ on $[a,b]$, or, more simply, the Riemann integral of $(f \circ \gamma) \cdot \gamma'$. A more general definition for piecewise C^1 curves could then have been obtained by linear extension. As outlined in [Bals, p. 17], however, this would have led to the exact same notion of a line integral. Therefore, I chose the more universal definition for this thesis, which also allows easier proofs of the important properties of the integral.

Proposition 1.2.7 (Properties of line integral concerning curve)**.**

Assume $D_f \subseteq \mathbb{C}$, a continuous function $f : D_f \to \mathcal{A}$ and a rectifiable curve $\gamma : [a, b] \to D_f$.

(i) *Invariance to reparameterization*: If $\gamma' : [c, d] \to D_f$ is a reparameterization of γ, then

$$\int_\gamma f(z)\, dz = \int_{\gamma'} f(z)\, dz\ .$$

(ii) *Additivity*: If $\gamma_1 : [a, b] \to D_f$ and $\gamma_2 : [c, d] \to D_f$ are rectifiable curves such that $\gamma_1(b) = \gamma_2(c)$, then $\gamma_1 + \gamma_2$ is a rectifiable curve in D_f and

$$\int_{\gamma_1+\gamma_2} f(z)\, dz = \int_{\gamma_1} f(z)\, dz + \int_{\gamma_2} f(z)\, dz\ .$$

(iii) *Inversion*: $-\gamma : [a, b] \to D_f$ is a rectifiable curve in D_f and

$$\int_{-\gamma} f(z)\, dz = -\int_\gamma f(z)\, dz\ .$$

Proof. (i): Let $\gamma' = \gamma \circ \phi$ for a strictly increasing, surjective function $\phi : [c, d] \to [a, b]$. Since γ and ϕ are continuous, so is $\gamma' = \gamma \circ \phi$ and hence γ' is a curve. Monotonicity of ϕ gives

$$\begin{array}{ccccccccccccc} & c & = & t_0 & < & t_1 & < & \dots & < & t_n & = & d & \\ \Longrightarrow & a & = & \phi(t_0) & < & \phi(t_1) & < & \dots & < & \phi(t_n) & = & b & . \end{array}$$

Deduce that γ' is rectifiable.

Now, let $(P_n)_{n\in\mathbb{N}}$ be an admissible sequence of partitions of $[c, d]$ and let $(T_n)_{n\in\mathbb{N}}$ be any corresponding tags. Interpreted in a slightly different way, the above results say that, if $P_n = (t_0, \dots, t_N)$, then $(\phi(P_n))_{n\in\mathbb{N}}$ with $\phi(P_n) := (\phi(t_0), \dots, \phi(t_N))$ is a sequence of partitions of $[a, b]$. Uniform continuity of ϕ ensures that $\|\phi(P_n)\| \to 0$, $n \to \infty$, so $(\phi(P_n))_{n\in\mathbb{N}}$ is admissible. If $T_n = (\tau_1, \dots, \tau_N)$, then $(\phi(T_n))_{n\in\mathbb{N}}$ with $\phi(T_n) := (\phi(\tau_1), \dots, \phi(\tau_N))$ corresponds to $(\phi(P_n))_{n\in\mathbb{N}}$. Verify that

$$S(f \circ \gamma, \gamma, \phi(P_n), \phi(T_n)) = S(f \circ \gamma', \gamma', P_n, T_n)\ .$$

The assertion then follows from the definition, when letting $n \to \infty$.

(ii): By virtue of (i), it suffices to prove the results in the case, where $\gamma_2 : [b, c] \to D_f$. As the curve length is additive, we find $L(\gamma_1 + \gamma_2) = L(\gamma_1) + L(\gamma_2) < \infty$ and $\gamma_1 + \gamma_2$ is rectifiable.

Now, let $(P_n)_{n\in\mathbb{N}}, (Q_n)_{n\in\mathbb{N}}$ be admissible sequences of partitions of $[a, b]$ and $[b, c]$ with corresponding tags $(T_n)_{n\in\mathbb{N}}, (S_n)_{n\in\mathbb{N}}$. Then, similar to ($i$), $(P_n \cup Q_n)_{n\in\mathbb{N}}$ is a sequence of partitions of $[a, c]$ and since

$\|P_n \cup Q_n\| = \max\{\|P_n\|, \|Q_n\|\} \to 0$, $n \to \infty$, we find that $(P_n \cup Q_n)_{n\in\mathbb{N}}$ is admissible. Define the union of $T_n = (\tau_1, \ldots, \tau_N)$ and $S_n = (\sigma_1, \ldots, \sigma_M)$ as the $(N+M)$-tuple $T_n \cup S_n := (\tau_1, \ldots, \tau_N, \sigma_1, \ldots, \sigma_M)$ (i. e. keeping the possible doubling $\tau_N = \sigma_1$, unlike what we specified for partitions). Then, $(T_n \cup S_n)_{n\in\mathbb{N}}$ corresponds to $(P_n \cup Q_n)_{n\in\mathbb{N}}$. Convince yourself that

$$S(f \circ (\gamma_1 + \gamma_2), \gamma_1 + \gamma_2, P_n \cup Q_n, T_n \cup S_n) = S(f \circ \gamma_1, \gamma_1, P_n, T_n) + S(f \circ \gamma_2, \gamma_2, Q_n, S_n)$$

and let $n \to \infty$.

(iii): As $[-\gamma] = [\gamma]$ and $L(-\gamma) = L(\gamma)$, we find that $-\gamma$ is a rectifiable curve in D_f.

Now, fix $n \in \mathbb{N}$. The definition of the integral allows us to specifically choose the equidistant partition $P_n := (t_0, \ldots, t_N)$, where $t_i := a + i/N\,(b-a)$, $i = 0, \ldots, N$. Again, set $T_n := T(P_n)$ and write $T_n = (\tau_1, \ldots, \tau_n)$. Verify elementarily that $a+b-t_i = t_{N-i}$, $i = 0, \ldots, N$ and $a+b-\tau_i = \tau_{N-i+1}$, $i = 1, \ldots, N$. The axiom (A4) allows us to perform an index transformation, yielding

$$\begin{aligned}
S(f \circ (-\gamma), -\gamma, P_n, T_n) &= \sum_{i=1}^{N} f(\gamma(a+b-\tau_i))(\gamma(a+b-t_i) - \gamma(a+b-t_{i-1})) \\
&= \sum_{i=1}^{N} f(\gamma(\tau_{N-i+1}))(\gamma(t_{N-i}) - \gamma(t_{N-i+1})) \\
&= -\sum_{i=1}^{N} f(\gamma(\tau_{N-i+1}))(\gamma(t_{N-i+1}) - \gamma(t_{(N-i+1)-1})) \\
&= -\sum_{i=1}^{N} f(\gamma(\tau_i))(\gamma(t_i) - \gamma(t_{i-1})) = -\,S(f \circ \gamma, \gamma, P_n, T_n)\ .
\end{aligned}$$

Letting $n \to \infty$ finishes the proof.

□

Remark 1.2.8 (Addition operators)**.** The last proposition implies that the two operators "+" and "$\oplus$" for adding curves defined in this thesis are to some extent compatible: if $\gamma_1 : [a,b] \to D$ and $\gamma_2 : [c,d] \to D$ are two rectifiable curves such that $\gamma_1(b) = \gamma_2(c)$, then $\gamma_1 + \gamma_2$ and $\gamma_1 \oplus \gamma_2$ are equivalent in the sense that

$$\int_{\gamma_1+\gamma_2} f(z)\,dz = \int_{\gamma_1\oplus\gamma_2} f(z)\,dz$$

for every function f which is continuous on $[\gamma_1+\gamma_2] = [\gamma_1\oplus\gamma_2]$. It is, however, not redundant to introduce both concepts, as the result of "+" is a function and thus can be handled more intuitively, but "$\oplus$" is the more general operator, that can be applied to any two curves, not just to the ones satisfying the above condition.

Proposition 1.2.9 (Properties of line integral concerning integrand)**.** Throughout this proposition, assume $D_f \subseteq \mathbb{C}$ and $f : D_f \to \mathcal{A}$ continuous.

(i) *Linearity*: If $D_g \subseteq \mathbb{C}$, $g : D_g \to \mathcal{A}$ continuous and $\Gamma = k_1\gamma_1 \oplus \ldots \oplus k_n\gamma_n$ is a chain of rectifiable curves in $D_f \cap D_g$, defining $f + g : D_f \cap D_g \to \mathcal{A}, z \mapsto f(z) + g(z)$ gives us a continuous function satisfying

$$\int_\Gamma f(z) + g(z)\; dz = \int_\Gamma f(z)\; dz + \int_\Gamma g(z)\; dz\ .$$

Similarly, if $\alpha \in \mathbb{C}$ and $\Gamma = k_1\gamma_1 \oplus \ldots \oplus k_n\gamma_n$ is a chain of rectifiable curves in D_f, we set $\alpha f : D_f \to \mathcal{A}, z \mapsto \alpha f(z)$ and obtain

$$\int_\Gamma \alpha f(z)\; dz = \alpha \int_\Gamma f(z)\; dz\ .$$

From now on, additionally let $\Gamma = k_1\gamma_1 \oplus \ldots \oplus k_n\gamma_n$ be a chain of rectifiable curves in D_f.

(ii) *Estimation lemma*:

$$\left\| \int_\Gamma f(z)\; dz \right\| \leq ML(\Gamma)\ ,$$

where $M := \max_{z\in[\Gamma]} \|f(z)\|$.

(iii) *Compatibility with multiplication*: For every $A \in \mathcal{A}$, the functions $Af : D_f \to \mathcal{A}, z \mapsto Af(z)$ and $fA : D_f \to \mathcal{A}, z \mapsto f(z)A$ are continuous and we find

$$\int_\Gamma Af(z)\; dz = A\left(\int_\Gamma f(z)\; dz\right)\ ,$$
$$\int_\Gamma f(z)A\; dz = \left(\int_\Gamma f(z)\; dz\right) A\ .$$

The result also holds verbatim for continuous, complex-valued functions $f : D_f \to \mathbb{C}$, where the right-hand side is explained by applying the theory of this section to the Banach algebra $\mathbb{C}$.

(iv) *Compatibility with dual space*: We call $\mathcal{A}' := \{\varphi : \mathcal{A} \to \mathbb{C} \mid \varphi \text{ continuous, } \mathbb{C}\text{-linear}\}$ the *continuous dual space of* $\mathcal{A}$. For every $\varphi \in \mathcal{A}'$

$$\varphi\left(\int_\Gamma f(z)\; dz\right) = \int_\Gamma \varphi(f(z))\; dz\ ,$$

where the right-hand side is meant as in the complex-valued case of (iii).

(v) *Continuity with respect to compact convergence*: Let $(f_n)_{n\in\mathbb{N}}$ be a sequence of functions which are continuous on $D \subseteq \mathbb{C}$ and let $f_n \to f$ uniformly over compact subsets of D. Then, f is continuous on D and for every chain Γ of rectifiable curves in D

$$\int_\Gamma f_n(z)\ dz \to \int_\Gamma f(z)\ dz\ .$$

Proof. All the results are established only for a single rectifiable curve $\gamma : [a, b] \to D_f$. The generalization to chains $\Gamma = k_1\gamma_1 \oplus \ldots \oplus k_n\gamma_n$ then always follows by invoking axioms of the Banach algebra or other assumptions as mentioned in the respective parts of the proof.

Throughout, let $(P_n)_{n\in\mathbb{N}}$ be any admissible sequence of partitions of $[a, b]$ and write $T_n := T(P_n)$.

(i): By virtue of the axiom (S1) for every $n \in \mathbb{N}$

$$S((f+g)\circ\gamma, \gamma, P_n, T_n) = S(f\circ\gamma,\gamma,P_n,T_n) + S(g\circ\gamma,\gamma,P_n,T_n)\ .$$

Letting $n \to \infty$ yields the desired equation. The same idea works for homogeneity with (S1) and (S3). Transfer the results to chains using (A4), (S1) and (S2).

(ii): Following the proof in [Berg], we observe that the function $\|.\| \circ f \circ \gamma : [a, b] \to \mathbb{R}$ is continuous and hence, the maximum in $\max_{z\in[\gamma]} \|f(z)\| = \max_{t\in[a,b]} (\|.\|\circ f\circ\gamma)(t)$ is attained. For $n \in \mathbb{N}$, we write $P_n = (t_0, \ldots, t_N), T_n = (\tau_1, \ldots, \tau_N)$ and obtain

$$\|S(f\circ\gamma,\gamma,P_n,T_n)\| \le \sum_{i=1}^{N} \|f(\gamma(\tau_i))\||\gamma(t_i)-\gamma(t_{i-1})| \le M\sum_{i=1}^{N} |\gamma(t_i)-\gamma(t_{i-1})| \le ML(\gamma)\ .$$

Extend this to chains by invoking the triangle inequality and homogeneity of $\|.\|$ as well as the distributive law for $\mathbb{R}$.

(iii): Continuity of Af and fA follows from continuity of f with the help of axioms (M1) and (N5). In the same way as in (i), deduce from (M1) that for every $n \in \mathbb{N}$

$$S((Af)\circ\gamma,\gamma,P_n,T_n) = A\ S(f\circ\gamma,\gamma,P_n,T_n)\ ,$$
$$S((fA)\circ\gamma,\gamma,P_n,T_n) = S(f\circ\gamma,\gamma,P_n,T_n)\ A$$

and let $n \to \infty$. The result for general chains follows with the help of (M1). The proof in the complex-valued case is analogous with (S2) and (N2) playing the role of (M1) and (N5).

(iv): Using the notation of (ii), linearity of $\varphi \in \mathcal{A}'$ gives

$$\varphi(S(f\circ\gamma,\gamma,P_n,T_n)) = \sum_{i=1}^{N} \varphi(f(\gamma(\tau_i)))(\gamma(t_i)-\gamma(t_{i-1}))\ .$$

We identify the right-hand side as a Riemann-Stieltjes-sum which can be used to define the line integral of $\varphi \circ f$ along γ in the sense of theorem & definition 1.2.5 for $\mathcal{A} = \mathbb{C}$. Hence, as $n \to \infty$, this term converges to $\int_\gamma \varphi(f(z))\, dz$ in the sense of the proposition. Continuity of φ makes the left-hand side converge to $\varphi\left(\int_\gamma f(z)\, dz\right)$ and we obtain the desired equation. Reduce the case of chains to the one of a single curve by linearity of φ.

(v): From topology and calculus we know that the limit function f in this setting has to be continuous. $[\Gamma]$ is a finite union of images of compact intervals under continuous function and is thus a compact subset of D. Apply (i) and (ii) to find

$$\left\| \int_\Gamma f_n(z)\ dz - \int_\Gamma f(z)\ dz \right\| = \left\| \int_\Gamma f_n(z) - f(z)\ dz \right\| \leq L(\Gamma) \max_{z \in [\Gamma]} \|f_n(z) - f(z)\| \to 0, \quad n \to \infty\ .$$

□

Next, we want to prove that there always exists a suitable cycle for the definition of our future calculus. To do so, we need to shed some light on the assumptions of Cauchy's integral formula mentioned in the introduction:

Definition 1.2.10 (Winding number and admissible cycle)**.** Let $\gamma : [a, b] \to \mathbb{C} \setminus \{0\}$ be a closed curve. As seen in [Down, p. 4f.], we can always find a continuous function $\theta : [a, b] \to \mathbb{R}$, such that $\gamma(t) = |\gamma(t)|\, e^{i\theta(t)}$ for every $t \in [a, b]$. We define the *winding number of* γ *about the origin* as

$$\mathrm{ind}_\gamma(0) := \frac{\theta(b) - \theta(a)}{2\pi}\ .$$

Well-definedness of this notion follows from the fact that any two choices for θ can only differ by a constant integer multiple of 2π (see [Down, p. 1ff.] for the proof). Note that since γ is closed, $\theta(b) - \theta(a) \in 2\pi\mathbb{Z}$ and hence, $\mathrm{ind}_\gamma(0) \in \mathbb{Z}$. We generalize this to $z \in \mathbb{C} \setminus [\gamma]$ by setting $\mathrm{ind}_\gamma(z) = \mathrm{ind}_{\gamma - z}(0)$, where $\gamma - z : [a, b] \to \mathbb{C} \setminus \{0\}, t \mapsto \gamma(t) - z$. It is shown in [Down, p. 7f.] that for a closed piecewise C^1 curve γ the winding number can be computed more easily by

$$\mathrm{ind}_\gamma(z) = \frac{1}{2\pi i} \int_\gamma \frac{d\zeta}{\zeta - z}\ ,$$

where the right-hand side again denotes the line integral of the complex-valued function $f(\zeta) := (\zeta - z)^{-1}$ along γ, explained by applying the theory of this section to the Banach algebra $\mathbb{C}$. For a cycle $\Gamma = k_1\gamma_1 \oplus \ldots \oplus k_n\gamma_n$ and $z \in \mathbb{C} \setminus [\Gamma]$, we set the *winding number of* Γ *about* z as

$$\mathrm{ind}_\Gamma(z) := \sum_{i=1}^{n} k_i \, \mathrm{ind}_{\gamma_i}(z) \ .$$

If $\gamma_1, \ldots, \gamma_n$ are piecewise C^1 this coincides with

$$\mathrm{ind}_\Gamma(z) = \frac{1}{2\pi i} \int_\Gamma \frac{d\zeta}{\zeta - z} \ .$$

With these definitions we can restate Cauchy's integral formula precisely as

$$f(z) = \int_\Gamma \frac{f(\zeta)}{\zeta - z} \, d\zeta$$

for a function f which is holomorphic on an open set $U \subseteq \mathbb{C}$, a complex argument $z \in U$ and a cycle Γ consisting of finitely many rectifiable Jordan curves in $U \setminus \{z\}$ such that $\mathrm{ind}_\Gamma(z) = 1$ and $\mathrm{ind}_\Gamma(\zeta) = 0$ for every $\zeta \in \mathbb{C} \setminus U$. This motivates the following definition:
If $K \neq \emptyset$ is a non-empty, compact subset of an open set $U \subseteq \mathbb{C}$, we call a cycle Γ *admissible for K and U*, if Γ consists of finitely many rectifiable Jordan curves in $U \setminus K$ such that $\mathrm{ind}_\Gamma(z) = 1 \ \forall z \in K$ and $\mathrm{ind}_\Gamma(z) = 0 \ \forall z \in \mathbb{C} \setminus U$.

As seen in [Born, p. 63], using the estimation lemma 1.2.9 (ii) for a cycle Γ of piecewise C^1 curves and $z, z' \in \mathbb{C} \setminus [\Gamma]$ yields

$$|\,\mathrm{ind}_\Gamma(z) - \mathrm{ind}_\Gamma(z')| = \frac{1}{2\pi} \left| \int_\Gamma \frac{z - z'}{(\zeta - z)(\zeta - z')} \, d\zeta \right| \leq \frac{|z - z'] \ L(\Gamma)}{2\pi \, \mathrm{dist}(\{z, z'\}, [\Gamma])^2} \ ,$$

so $\mathrm{ind}_\Gamma : \mathbb{C} \setminus [\Gamma] \to \mathbb{Z}$ is continuous.

Lemma 1.2.11 (Saks-Zygmund)**.** Let $K \neq \emptyset$ be a non-empty, compact subset of an open set $U \subseteq \mathbb{C}$. Then, there always exists an admissible cycle Γ for K and U. In fact, we can even assume the curves appearing in Γ to be piecewise C^1.

Proof. We elaborate on the sketch of the proof given in [Born, p. 76]:
From topology we know that the distance $\mathrm{dist}(K, \mathbb{C} \setminus U)$ of the compact set K and the closed set $\mathbb{C} \setminus U$ is positive. We cover the complex plane with a grid of width $\delta := \mathrm{dist}(K, \mathbb{C} \setminus U)/2 > 0$ by setting $z_{j,k} := \delta \ (j + ik)$, $j, k \in \mathbb{Z}$. Write

$$S_{j,k} := \{z \in \mathbb{C} \mid j\delta \leq \mathrm{Re}(z) \leq (j+1)\delta,\, k\delta \leq \mathrm{Im}(z) \leq (k+1)\delta\},\, j,k \in \mathbb{Z}$$

for the square with the vertices $z_{j,k}, z_{j+1,k}, z_{j+1,k+1}$ and $z_{j,k+1}$.

Since K is non-empty and bounded, finitely many squares, but at least one, intersect with K. More precisely, if $|z| \leq M\ \forall\, z \in K$, then $K \cap S_{j,k} \neq \emptyset$ for no more than $4(\lceil M/\delta \rceil)^2$ distinct pairs $(j,k) \in \mathbb{Z} \times \mathbb{Z}$. Rename the corresponding squares as $S_1, \ldots, S_\ell$. To see that $\bigcup_{i=1}^{\ell} S_i \subseteq U$, verify that for every $i \in \{1, \ldots, \ell\}$, $z \in S_i$ there exists $z' \in K \cap S_i$ satisfying $|z - z'| \leq \sqrt{2}\delta < \mathrm{dist}(K, \mathbb{C} \setminus U) \implies z \notin \mathbb{C} \setminus U$.

Now, for $z, w \in \mathbb{C}$, let $\overrightarrow{zw}$ denote the oriented line segment with initial point z and terminal point w, so

$$\overrightarrow{zw} : [0,1] \to \mathbb{C}, t \mapsto (1-t)z + tw\ .$$

Define ∂S_i as the boundary curve of the square S_i, oriented in the positive sense (i. e. counter-clockwise). This means, if $S_i = S_{j,k}$, then

$$\partial S_i := \overrightarrow{z_{j,k} z_{j+1,k}} + \overrightarrow{z_{j+1,k} z_{j+1,k+1}} + \overrightarrow{z_{j+1,k+1} z_{j,k+1}} + \overrightarrow{z_{j,k+1} z_{j,k}}\ .$$

We denote those oriented line segments appearing in $\partial S_1 \oplus \ldots \oplus \partial S_\ell$, whose images are edges of exactly one S_i, by $\tilde{\gamma}_1, \ldots, \tilde{\gamma}_m, m \leq 4l$. These are curves in $U \setminus K$, since for every $\tilde{\gamma}_h$ we find $[\tilde{\gamma}_h] \subseteq [\partial S_i \oplus \ldots \oplus \partial S_\ell] = \bigcup_{i=1}^{\ell} [\partial S_i] \subseteq \bigcup_{i=1}^{\ell} S_i \subseteq U$, but also $K \cap [\tilde{\gamma}_h] = \emptyset$, because otherwise $K \cap [-\tilde{\gamma}_h] = K \cap [\tilde{\gamma}_h] \neq \emptyset$ and so K would also intersect with the square that has $-\tilde{\gamma}_h$ in its boundary and $[\tilde{\gamma}_h]$ would be an edge of two distinct squares, which contradicts the choice of $\tilde{\gamma}_h$.

Now, fix $h \in \{1, \ldots, m\}$ and consider $\tilde{\gamma}_h$. We claim that there always has to be an $h' \in \{1, \ldots, m\}$ such that the initial point of $\tilde{\gamma}_{h'}$ is the terminal point z of $\tilde{\gamma}_h$. To see this, note that for the structure of ∂S_i, you can always "turn left" at z and find another curve that appears in ∂S_i. If this line segment equals $\tilde{\gamma}_{h'}$ for some $h' \in \{1, \ldots, m\}$, we are done. Otherwise, the image of this curve has to be an edge of two squares S_i and $S_{i'}$ and there exists a line segment in $\partial S_{i'}$ "straight ahead" of z. Again, check if this curve appears in $\tilde{\gamma}_1, \ldots, \tilde{\gamma}_m$ and if not, then for the same reasons as above, there exists an option "to the right" of z. If we have not found a suitable line segment before, then this last curve has to be $\tilde{\gamma}_{h'}$ for some $h' \in \{1, \ldots, m\}$, because otherwise $[\tilde{\gamma}_h]$ was an edge of two distinct squares. If the choice at z should ever not be unique, go for the option "most to the left", i. e. if available, the curve "to the left" and if not, the one "straight ahead", but never the line segment "to the right". This way you avoid self-intersections in the curve. Following this procedure, we can always connect the curves in $\tilde{\gamma}_1, \ldots, \tilde{\gamma}_m$ and, as there are only finitely many of them, we will eventually return to the initial point of $\tilde{\gamma}_h$. By adding all the line segments that occurred in this process by "+", we obtain an obviously rectifiable and even piecewise C^1 Jordan curve, which we denote by γ_1. If there are line segments left in $\tilde{\gamma}_1, \ldots, \tilde{\gamma}_m$, repeat the process and obtain further piecewise C^1 Jordan curves $\gamma_2, \ldots, \gamma_n$. Set

$$\Gamma := \gamma_1 \oplus \ldots \oplus \gamma_n \ .$$

By definition, with every γ that appears in $\partial S_1 \oplus \ldots \oplus \partial S_\ell$, but not in Γ, also $-\gamma$ is a part of $\partial S_1 \oplus \ldots \oplus \partial S_\ell$, but not of Γ. Proposition 1.2.7 (*iii*) implies that for every function f that is continuous on $[\gamma]$

$$\int_{\gamma \oplus (-\gamma)} f(z)\, dz = \int_\gamma f(z)\, dz - \int_\gamma f(z)\, dz = 0_{\mathcal{A}} \ .$$

For every $z \in \mathbb{C} \setminus [\partial S_1 \oplus \ldots \oplus \partial S_\ell]$, the function $f(\zeta) = (\zeta - z)^{-1}$ is continuous on the image of all such γ and thus, it follows that

$$\operatorname{ind}_\Gamma(z) = \operatorname{ind}_{\partial S_1 \oplus \ldots \oplus \partial S_\ell}(z) \ .$$

By construction

$$\begin{aligned} \operatorname{ind}_{\partial S_1 \oplus \ldots \oplus \partial S_\ell}(z) = 1 \quad & \forall\, z \in K \setminus [\partial S_1 \oplus \ldots \oplus \partial S_\ell] \ , \\ \operatorname{ind}_{\partial S_1 \oplus \ldots \oplus \partial S_\ell}(z) = 0 \quad & \forall\, z \in \mathbb{C} \setminus (U \cup [\partial S_1 \oplus \ldots \oplus \partial S_\ell]) \ . \end{aligned}$$

Continuity of $\operatorname{ind}_\Gamma$ makes these properties hold for all $z \in K$ and $z \in \mathbb{C} \setminus U$, respectively.

□

As a last part of this section, we want to foreshadow well-definedness of our future calculus. We do so by generalizing the well-known *Cauchy integral theorem* from complex analysis to Banach space-valued functions. More precisely, we will use the so-called *homologous version* of this result, which involves the notion of the *interior of a cycle* Γ defined by $\operatorname{Int}(\Gamma) := \{z \in \mathbb{C} \setminus [\Gamma] \mid \operatorname{ind}_\Gamma(z) \neq 0\}$ and reads as follows: If $U \subseteq \mathbb{C}$ is open, $f : U \to \mathbb{C}$ is holomorphic and Γ is a cycle in U, that is *homologous to zero*, i. e. $\operatorname{Int}(\Gamma) \subseteq U$, then

$$\int_\Gamma f(z)\, dz = 0_{\mathbb{C}} \ .$$

A proof of this result for the complex case can be seen in [Bals, p. 21f.]. In this thesis, I will only render the idea from [Bals, p.22] on how this result can be extended (as could be many others) to functions taking values in a general Banach space.

Theorem 1.2.12 (Cauchy's integral theorem)**.** If $U \subseteq \mathbb{C}$ is open, $f : U \to \mathcal{A}$ is holomorphic, and Γ is a cycle in U, that is homologous to zero, then

$$\int_\Gamma f(z)\, dz = 0_\mathcal{A} \ .$$

Proof. By virtue of the well-known *Hahn-Banach theorem* from functional analysis, it suffices to show $\varphi(\int_\Gamma f(z)\, dz) = 0_\mathbb{C}$ for all $\varphi \in \mathcal{A}' = \{\varphi : \mathcal{A} \to \mathbb{C} \mid \varphi \text{ continuous, } \mathbb{C}\text{-linear}\}$ (see [Bals, p. 63]). Now, to prove that this condition is satisfied, invoke proposition 1.2.9 (*iv*) and find that for every $\varphi \in \mathcal{A}'$

$$\varphi\left(\int_\Gamma f(z)\, dz\right) = \int_\Gamma \varphi(f(z))\, dz \ .$$

Because φ is continuous and linear, the function $\varphi \circ f : U \to \mathbb{C}$ is holomorphic and so the complex version Cauchy's integral theorem says that the right-hand side equals $0_\mathbb{C}$,

□

Corollary 1.2.13 (Well-definedness)**.** Let $K \neq \emptyset$ be a non-empty, compact subset of an open set $U \subseteq \mathbb{C}$ and let Γ, Γ' be two admissible cycles for U and K. Then, for every f that is holomorphic in U

$$\int_\Gamma f(z)\, dz = \int_{\Gamma'} f(z)\, dz \ .$$

Proof. The sets U and $\mathbb{C} \setminus K$ are open and so is $U \cap (\mathbb{C} \setminus K) = U \setminus K$. $\Gamma \oplus (-\Gamma)$ is a cycle in $U \setminus K$ and the conditions of the Saks-Zygmund lemma give

$$\begin{aligned} \mathrm{ind}_{\Gamma\oplus(-\Gamma')}(z) &= 1 + (-1) = 0 \quad \forall z \in K \ , \\ \mathrm{ind}_{\Gamma\oplus(-\Gamma')}(z) &= 0 + 0 = 0 \quad\ \ z \in \mathbb{C} \setminus U \ . \end{aligned}$$

so $\mathrm{Int}_{\Gamma\oplus(-\Gamma')} \subseteq U \setminus K$ and $\Gamma \oplus (-\Gamma')$ is homologous to zero. Thus, the generalized version of Cauchy's integral theorem in 1.2.12 states that

$$\int_{\Gamma\oplus(-\Gamma')} f(z)\, dz = 0_\mathcal{A} \ ,$$

which, by virtue of proposition 1.2.7 (*iii*), is equivalent to

$$\int_\Gamma f(z)\, dz = \int_{\Gamma'} f(z)\, dz \ .$$

□

1.3 Holomorphic functional calculus

Still, let $\mathcal{A}$ be a Banach algebra with multiplicative identity I and norm $\|.\|$ and fix $A \in \mathcal{A}$.

Definition 1.3.1 (Algebra of holomorphic functions)**.** Write $H(A)$ for the set of all complex-valued functions $f : D_f \to \mathbb{C}$, $D_f \subseteq \mathbb{C}$ that are holomorphic in an open neighborhood $U_f \subseteq D_f$ of the spectrum $\sigma(A)$ of A. We establish the structure of a $\mathbb{C}$-algebra on $H(A)$ as follows: if $f : D_f \to \mathbb{C}$, $g : D_g \to \mathbb{C}$ are holomorphic on the open neighborhoods $U_f \subseteq D_f$, $U_g \subseteq D_g$ of $\sigma(A)$, respectively, then $f + g : D_f \cap D_g \to \mathbb{C}, z \mapsto f(z) + g(z)$ and $f \cdot g : D_f \cap D_g \to \mathbb{C}, z \mapsto f(z) \cdot g(z)$ are holomorphic on the open neighborhood $U_f \cap U_g$ of $\sigma(A)$. Similarly, $\alpha f : D_f \to \mathbb{C}, z \mapsto \alpha f(z)$ is holomorphic on the open neighborhood D_f of $\sigma(A)$ for every $\alpha \in \mathbb{C}$.

Proposition 1.1.7, the Saks-Zygmund lemma, corollary 1.2.13 and the above definition have paved the way, so that we can now finally formulate the holomorphic functional calculus:

Definition 1.3.2 (Holomorphic functional calculus)**.** The *holomorphic functional calculus for A* is the mapping

$$\begin{aligned} \Phi_A : H(A) &\to \mathcal{A} \ , \\ f &\mapsto \frac{1}{2\pi i} \int_\Gamma \frac{f(\zeta)}{\zeta I - A} \, d\zeta \ , \end{aligned}$$

where Γ is an admissible cycle for $\sigma(A)$ and the open set U_f, on which f is holomorphic. Typically, we write $f(A) := \Phi_A(f)$.

Theorem 1.3.3 (General properties of the calculus)**.**

(i) *Homomorphism property*: Φ_A is an algebra homomorphism from $H(A)$ onto $\mathcal{A}$, i. e. for every $\alpha \in \mathbb{C}$, $f, g \in H(A)$

$$\begin{aligned} (f+g)(A) &= f(A) + g(A) \ , \\ (\alpha f)(A) &= \alpha f(A) \ , \\ (f \cdot g)(A) &= f(A) \cdot g(A) \ . \end{aligned}$$

(*ii*) *Preserving commutativity*: Every element $B \in \mathcal{A}$ that commutes with A, meaning that $AB = BA$, also commutes with $f(A)$ for every $f \in H(A)$.

(*iii*) *Continuity with respect to compact convergence*: let $(f_n)_{n\in\mathbb{N}}$ be a sequence of functions which are holomorphic on an open neighborhood U of $\sigma(A)$, and let $f_n \to f$, $n \to \infty$ uniformly over compact subsets of U. Then, $f \in H(A)$ and

$$f_n(A) \to f(A), \quad n \to \infty \ .$$

Proof. (*i*): Compatibility with addition and scalar multiplication follows directly from Proposition 1.2.9 (*i*) with the help of (M1) and (S3). For the result concerning multiplication, we follow [Neun, p. 8f.]: let $f : D_f \to \mathbb{C}$ and $g : D_g \to \mathbb{C}$ be holomorphic on the open neighborhoods $U_f \subseteq D_f$, $U_g \subseteq D_g$ of $\sigma(A)$. The Saks-Zygmund lemma gives us an admissible cycle Γ_1 for $\sigma(A)$ and $U_f \cap U_g$. Now, $[\Gamma_1]$ is a finite union of images of compact intervals under continuous functions and thus, also $\sigma(A) \cup [\Gamma_1]$ compact. Because of this, we can also choose an admissible cycle Γ_2 for $\sigma(A) \cup [\Gamma_1]$ and $U_f \cap U_g$ (the need for this will become clear later on). Use the first resolvent identity 1.1.5 and proposition 1.2.9 (*iii*) to deduce

$$\begin{aligned}
f(A) \cdot g(A) &= \left(\frac{1}{2\pi i}\int_{\Gamma_1} \frac{f(\zeta)}{\zeta I - A}\, d\zeta\right)\left(\frac{1}{2\pi i}\int_{\Gamma_2} \frac{g(\omega)}{\omega I - A}\, d\omega\right) \\
&= \frac{1}{(2\pi i)^2}\int_{\Gamma_1}\int_{\Gamma_2} \frac{f(\zeta)g(\omega)}{(\zeta I - A)(\omega I - A)}\, d\omega\, d\zeta \\
&= \frac{1}{(2\pi i)^2}\int_{\Gamma_1}\int_{\Gamma_2} f(\zeta)g(\omega)\left(\frac{(\zeta I - A)^{-1} - (\omega I - A){-1}}{\omega - \zeta}\right) d\omega\, d\zeta \\
&= \frac{1}{(2\pi i)^2}\int_{\Gamma_1} \frac{f(\zeta)}{\zeta I - A}\left(\int_{\Gamma_2} \frac{g(\omega)}{\omega - \zeta}\, d\omega\right) d\zeta - \frac{1}{(2\pi i)^2}\int_{\Gamma_2} \frac{g(\omega)}{\omega I - A}\left(\int_{\Gamma_1} \frac{f(\zeta)}{\omega - \zeta}\, d\zeta\right) d\omega \ .
\end{aligned}$$

Now, the second term vanishes by virtue of Cauchy's integral theorem for complex-valued functions, because we have constructed the cycles to satisfy $[\Gamma_1] \cap [\Gamma_2] = \emptyset$ and thus, $\zeta \mapsto f(\zeta)(\omega - \zeta)^{-1}$ is holomorphic in an open neighborhood of $[\Gamma_1]$, that does not intersect with $[\Gamma_2]$. Therefore, the complex version of Cauchy's integral formula lets us complete the argument by

$$\begin{aligned}
f(A) \cdot g(A) &= \frac{1}{2\pi i}\int_{\Gamma_1} \frac{f(\zeta)}{\zeta I - A}\left(\frac{1}{2\pi i}\int_{\Gamma_2} \frac{g(\omega)}{\omega - \zeta}\, d\omega\right) d\zeta \\
&= \frac{1}{2\pi i}\int_{\Gamma_1} \frac{f(\zeta)g(\zeta)}{\zeta I - A}\, d\zeta = (f \cdot g)(A) \ .
\end{aligned}$$

(*ii*): Verify with plain algebra that every element which commutes with A, also commutes with $(\zeta I - A)^{-1}$ for every $\zeta \in \rho(A)$. Apply Proposition 1.2.9 (*iii*) and (S3) to finish the proof.

(*iii*): It is a standard result from complex analysis that the limit function f in this setting is holomorphic on U and hence an element of $H(A)$. Because R_A is bounded, the assumption implies that $f_n R_A/(2\pi i) \to f R_A/(2\pi i)$ uniformly over compact subsets of U. Close by applying proposition 1.2.9 (*v*). $\square$

Proposition 1.3.4 (Extension of polynomial and power series calculi)**.** Adopt the convention $A^0 := I$.

(i) *Polynomial case*: If $f(z) = \sum_{n=0}^{N} a_n z^n$, $N \geq 0$, $a_1, \ldots, a_N \in \mathbb{C}$, then $f \in H(A)$ and

$$f(A) = \sum_{n=0}^{N} a_n A^n \ .$$

(ii) *Power series case*: If $f \in H(A)$ and there exists z_0 such that $f(z) = \sum_{n=0}^{\infty} a_n (z - z_0)^n$, $(a_n)_{n \in \mathbb{N}} \subseteq \mathbb{C}$ for every z in an open neighborhood U of $\sigma(A)$, then

$$f(A) = \sum_{n=0}^{\infty} a_n (A - z_0 I)^n \ .$$

Proof. (i): I furnish the standard proof of this result in functional analysis as seen in [Schn]: by virtue of theorem 1.3.3 (i), it suffices to prove that $f(z) = z^k$, $k \geq 0$ implies $f(A) = A^k$. Verify that $\Gamma : [0,1] \to \mathbb{C}, t \mapsto r\exp(2\pi i t)$, $r > \|A\|$, is an admissible cycle for $\sigma(A)$ and $U_f = \mathbb{C}$, using the proof of proposition 1.1.7 (ii). Rearrange in the way that can be seen at the same place to find

$$f(A) = \frac{1}{2\pi i} \int_\Gamma \frac{\zeta^k}{\zeta I - A} \, d\zeta = \frac{1}{2\pi i} \int_\Gamma \sum_{n=0}^{\infty} \frac{A^n}{\zeta^{n+1-k}} \, d\zeta \ .$$

Because a power series converges uniformly to its limit over compact subsets of its disc of convergence, we can use proposition 1.2.9 (iii) and (v) to find

$$\frac{1}{2\pi i} \int_\Gamma \sum_{n=0}^{\infty} \frac{A^n}{\zeta^{n+1-k}} \, d\zeta = \sum_{n=0}^{\infty} A^n \left(\frac{1}{2\pi i} \int_\Gamma \frac{1}{\zeta^{n+1-k}} \, d\zeta \right) \ .$$

To show that this last term equals A^k, evaluate the complex-valued integral in the last line directly using the idea outlined in remark 1.2.6:

$$\frac{1}{2\pi i} \int_\Gamma \frac{1}{\zeta^{n+1-k}} \, d\zeta = \frac{1}{2\pi i} \int_0^1 \frac{1}{(r\exp(2\pi i t))^{n+1-k}} \, 2\pi i r \exp(2\pi i t) \, dt$$
$$= \int_0^1 (r\exp(2\pi i t))^{k-n} \, dt = \begin{cases} t|_0^1 = 1, & n = k \\ \frac{r^{k-n}}{2\pi i(k-n)} \exp(2\pi i(k-n)t)|_0^1 = 0, & n \neq k \end{cases} .$$

(ii): First, we want to show that $f(z) = (z - z_0)^n$ implies $f(A) = (A - z_0 I)^n$ for every $n \in \mathbb{N}$. Let $Z^k : \mathbb{C} \to \mathbb{C}, z \mapsto z^k$ denote the k-th monomial for $k \in \mathbb{N}$ and, only for this occasion identify z_0^{n-k} with the function defined by multiplying an element of $\mathcal{A}$ with this number for $k, n \in \mathbb{N}$. Use (i), the homomorphism property and the *binomial theorem* (first for $H(A)$, then for the commuting elements A and $z_0 I$ of $\mathcal{A}$) to deduce

$$\begin{aligned} f(A) &= (\sum_{k=0}^{n} Z^k z_0^{n-k})(A) \\ &= \sum_{k=0}^{n} Z^k(A) z_0^{n-k} \\ &= \sum_{k=0}^{n} A^k (z_0 I)^{n-k} = (A - z_0 I)^n \ . \end{aligned}$$

The assertion now follows from additivity and theorem 1.3.3 (*iii*) with the help of the above mentioned uniform convergence argument for power series.

□

Theorem 1.3.5 (Spectral mapping theorem & composition identity)**.**

(*i*) *Spectral mapping theorem*: For every $f \in H(A)$

$$\sigma(f(A)) = f(\sigma(A)) \ .$$

(*ii*) *Composition identity*: If $f \in H(A)$ and $g \in H(f(A))$, then

$$(g \circ f)(A) = g(f(A)) \ .$$

Proof. (*i*): The idea for this part of the proof is taken from [Neun, p. 10f.]. "$\subseteq$": we prove the equivalent contrapositive. Let $z \notin f(\sigma(A))$. Then, the function $g(\zeta) := (z - f(\zeta))^{-1}$ is holomorphic on the open neighborhood $\mathbb{C} \setminus f^{-1}(\{z\})$ of $\sigma(A)$. As $g(\zeta)(z - f(\zeta)) = 1$, the homomorphism property and proposition 1.3.4 give $g(A)(zI - f(A)) = I$, so $z \notin \sigma(f(A))$.
"$\supseteq$": let $f : D_f \to \mathbb{C}$ be holomorphic on U_f and fix $z \in D_f$. Define $g : D_f \to \mathbb{C}$ by

$$g(\zeta) := \begin{cases} \frac{f(z)-f(\zeta)}{z-\zeta}, & \zeta \neq z \\ f'(\zeta), & \zeta = z \end{cases} \ .$$

Then, g is holomorphic on U_f and $f(z) - f(\zeta) = (z - \zeta)g(\zeta)$ for every $\zeta \in \mathbb{C}$. So again, the previous results yield $f(z)I - f(A) = (zI - A)g(A)$. Now, if $f(z) \notin \sigma(f(A))$, then $zI - A$ is invertible with inverse $(f(z)I - f(A))^{-1}g(A)$ by virtue the above equation and commutativity of the multiplication in $H(A)$.

(ii): We elaborate on the proof from [Tayl]. Let $f : D_f \to \mathbb{C}$ be holomorphic in an open neighborhood U_f of $\sigma(A)$ and let $g : D_g \to \mathbb{C}$ be holomorphic in an open neighborhood U_g of $\sigma(f(A))$. The set $f^{-1}(U_g)$ is open and, as $f(\sigma(A)) \overset{(i)}{=} \sigma(f(A)) \subseteq U_g$, contains $\sigma(A)$. Because of this, we can define $U_f' := U_f \cap f^{-1}(U_g)$ to get an open neighborhood of $\sigma(A)$, on which f is holomorphic, satisfying $f(U_f') \subseteq U_g$ (see the rest of the proof for the use of this). Choose admissible cycles Γ_1 for $\sigma(A)$ and U_f' and Γ_2 for the compact set $f([\Gamma_1])$ and U_g. Now, Cauchy's integral formula for complex-valued functions states that for every $\zeta \in [\Gamma_1]$

$$(g \circ f)(\zeta) = g(f(\zeta)) = \frac{1}{2\pi i} \int_{\Gamma_2} \frac{g(\omega)}{\omega - f(\zeta)} \, d\omega \ .$$

Insert this into the definition

$$\begin{aligned}
(g \circ f)(A) &= \frac{1}{2\pi i} \int_{\Gamma_1} (g \circ f)(\zeta)(\zeta I - A)^{-1} \, d\zeta \\
&= \frac{1}{(2\pi i)^2} \int_{\Gamma_1} \left(\int_{\Gamma_2} \frac{g(\omega)}{\omega - f(\zeta)} \, d\omega \right) (\zeta I - A)^{-1} \, d\zeta \ .
\end{aligned}$$

Again, use proposition 1.2.9 (iii), reverse the order of integration (boundedness of the involved functions allows application of a generalized version of the *Fubini-Tonelli theorem*) and appeal to the proof of part (i) of this theorem to get

$$\begin{aligned}
(g \circ f)(A) &= \frac{1}{(2\pi i)^2} \int_{\Gamma_1} \left(\int_{\Gamma_2} \frac{g(\omega)}{\omega - f(\zeta)} \, d\omega \right) (\zeta I - A)^{-1} \, d\zeta \\
&= \frac{1}{2\pi i} \int_{\Gamma_2} g(\omega) \left(\frac{1}{2\pi i} \int_{\Gamma_1} \frac{1}{\omega - f(\zeta)} (\zeta I - A)^{-1} \, d\zeta \right) d\omega \\
&= \frac{1}{2\pi i} \int_{\Gamma_2} g(\omega)(\omega I - f(A))^{-1} \, d\omega = g(f(A)) \ .
\end{aligned}$$

□

Corollary 1.3.6 (Spectral radius formula)**.** We call the real number $r(A) := \sup\{|\lambda| \mid \lambda \in \sigma(A)\}$ the *spectral radius of* A. The following identity holds:

$$r(A) = \lim_{n\to\infty} \|A^n\|^{1/n} = \inf_{n\geq 1} \|A^n\|^{1/n} \ .$$

Notice that the left-hand side does not depend on the norm of the Banach algebra, whereas the terms in the middle and on the right do.

Proof. This proof is due to [Rud]. Fix $r > r(A)$ and write $\Gamma : [0,1] \to \mathbb{C}, t \mapsto r\exp(2\pi i t)$. Then, proposition 1.3.4 and the estimation lemma 1.2.9 (ii) give for every $n \in \mathbb{N}$

$$\|A^n\| = \frac{1}{2\pi}\left\|\int_\Gamma \frac{\zeta^n}{\zeta I - A}\, d\zeta\right\| \leq \frac{1}{2\pi}L(\Gamma)\max_{\zeta\in[\Gamma]}\|\zeta^n(\zeta I - A)^{-1}\| = r^{n+1}M \ .$$

where $M := \max_{\zeta\in[\Gamma]}\|(\zeta I - A)^{-1}\| < \infty$. Deduce $\limsup_{n\to\infty}\|A^n\|^{1/n} \leq r$ for every $r > r(A)$, so that

$$\limsup_{n\to\infty}\|A^n\|^{1/n} \leq r(A) \ .$$

But on the other hand, by virtue of the spectral mapping theorem, $\lambda \in \sigma(A)$ implies $\lambda^n \in \sigma(A^n)$ for every $n \in \mathbb{N}$. As seen in the proof of proposition 1.1.7 (*ii*), this means that $|\lambda^n| \leq \|A^n\|$. Finish the proof by deducing

$$r(A) \leq \inf_{n\geq 1}\|A^n\|^{1/n} \leq \liminf_{n\to\infty}\|A^n\|^{1/n} \ .$$

□

The last result of this chapter involves some notions that cannot be defined in every Banach algebra, but only in operator algebras of finite-dimensional Banach spaces. We know that every n-dimensional $\mathbb{C}$-vector space is isomorphic to $\mathbb{C}^n$. As discussed before, the corresponding operator algebra $B(\mathbb{C}^n)$ can be identified with the matrix algebra $\mathbb{C}^{n\times n}$. For this reason, let us directly formulate this proposition in terms of matrices:

Proposition 1.3.7 (Specific properties of the calculus for matrices)**.** Let $A \in \mathbb{C}^{n\times n}$.

(i) *Identity criterion using divisibility*: $f(A) = g(A)$ if $f - g$ is divisible (as a member of $H(A)$) by the characteristic polynomial χ_A of A.

(ii) *Identity criterion using eigenvalues*: If A has distinct eigenvalues, then $f(A) = g(A)$ if and only if $f(\lambda) = g(\lambda)$ for every eigenvalue λ of A.

(iii) *Surjectivity concerning commuting matrices*: If A has distinct eigenvalues, then every $n \times n$ matrix that commutes with A is $f(A)$ for some f.

(iv) *Real matrices*: If A is a real matrix and $f(\overline{z}) = \overline{f(z)}$ in a neighborhood of $\sigma(A)$, then f(A) is real.

(v) *Eigenvectors*: If v is an eigenvector of A for eigenvalue λ, then $f(A)v = f(\lambda)v$.

Proof. (i): The *Cayley-Hamilton theorem* from linear algebra states that $\chi_A(A) = 0_{\mathbb{C}^{n\times n}}$. Thus, if $f - g = \chi_A \cdot q$ for some $q \in H(A)$, then the homomorphism property gives $f(A) - g(A) = (f - g)(A) = (\chi_A \cdot q)(A) = \chi_A(A)\ q(A) = 0_{\mathbb{C}^{n\times n}}$, which is equivalent to $f(A) = g(A)$.

(*ii*): We know from linear algebra that a matrix with distinct eigenvalues is *diagonalizable*, i. e. $A = S\Lambda S^{-1}$, where $S \in \mathbb{C}^{n\times n}$ is invertible and $\Lambda = \operatorname{diag}(\lambda_1, \dots, \lambda_n)$ (a matrix, where all entries are zero, except for the diagonal entries, which are filled with the eigenvalues $\lambda_1, \dots, \lambda_n$ of A).
Now, we prove that $f(A) = f(S\Lambda S^{-1}) = Sf(\Lambda)S^{-1}$ for every $f \in H(A) = H(\Lambda)$ (notice that $\sigma(A) = \sigma(\Lambda)$). To do so, we verify that, if $\zeta I-\Lambda$ is invertible, then so is $\zeta I-S\Lambda S^{-1}$ with inverse $S(\zeta I-\Lambda)^{-1}S^{-1}$, and apply proposition 1.2.9 (*iii*). Multiplication with S and S^{-1} yields the intermediate result

$$f(A) = g(A) \iff f(\Lambda) = g(\Lambda)\ .$$

Next, we show that $f(\Lambda) = \operatorname{diag}(f(\lambda_1), \dots, f(\lambda_n))$ for every $f \in H(\Lambda) = H(A)$. Choose an admissible cycle Γ for $\sigma(\Lambda)$ and U_f and write E_i for the diagonal matrix, where the i-th entry equals 1 and all others are zero. Then, use the fact that $\sigma(\Lambda) \cap [\Gamma] = \emptyset$, proposition 1.2.9 (*i*) and (*iii*) as well as the complex-valued version of Cauchy's integral formula to deduce

$$\begin{aligned}
f(\Lambda) &= \frac{1}{2\pi i}\int_\Gamma f(\zeta)(\zeta I-\Lambda)^{-1}\,d\zeta\\
&= \frac{1}{2\pi i}\int_\Gamma f(\zeta)\left(\sum_{i=1}^n(\zeta-\lambda_i)^{-1}E_i\right)d\zeta\\
&= \sum_{i=1}^n\left(\frac{1}{2\pi i}\int_\Gamma f(\zeta)(\zeta-\lambda_i)^{-1}\,d\zeta\right)E_i\\
&= \sum_{i=1}^n f(\lambda_i)E_i = \operatorname{diag}(f(\lambda_1), \dots, f(\lambda_n))\ .
\end{aligned}$$

Summarizing:

$$f(A) = g(A) \iff \operatorname{diag}(f(\lambda_1), \dots, f(\lambda_n)) = \operatorname{diag}(g(\lambda_1), \dots, g(\lambda_n)) \iff f(\lambda_i) = g(\lambda_i), \quad i = 1, \dots, n\ .$$

(*iii*): Let $B \in \mathbb{C}^{n\times n}$ commute with A. Write $\lambda_1, \dots, \lambda_n$ for the distinct eigenvalues of A and $\mu_1, \dots, \mu_n$ for the not necessarily distinct eigenvalues of B. We know from numerical mathematics that there exists a so-called *interpolation polynomial f for the distinct data points* $(\lambda_1, \mu_1), \dots, (\lambda_n, \mu_n)$, meaning that $f(\lambda_i) = \mu_i$, $i = 1, \dots, n$. We will prove that $B = f(A)$ for this $f \in H(A)$.
First, we show that, if A has distinct eigenvalues and B commutes with A, then B is diagonalizable. To see this, let $v \in \mathbb{C}^n \setminus \{0_{\mathbb{C}^n}\}$ be an eigenvector of A for the eigenvalue λ. Since A and B commute,

$$Av = \lambda v \implies BAv = B\lambda v \implies A(Bv) = \lambda(Bv)\ .$$

Thus, Bv is an element of the *eigenspace* $\ker(\lambda I - A)$. As the eigenvalues of A are distinct, $\ker(\lambda I - A)$ is a one-dimensional subspace of $\mathbb{C}^n$, so $Bv = \mu v$ for some $\mu \in \mathbb{C}$, meaning that v is an eigenvector for B.

Deduce from this that B is diagonalizable, using the classical characterization from linear algebra that an $n \times n$ matrix is diagonalizable if and only if $\mathbb{C}^n$ possesses a basis consisting only of eigenvectors of that matrix.

Another result from linear algebra now tells us that A and B are in fact *simultaneously diagonalizable*, meaning that there exists a common invertible $S \in \mathbb{C}^{n\times n}$, such that $A = S\Lambda S^{-1}$ and $B = SMS^{-1}$, where $\Lambda = \operatorname{diag}(\lambda_1, \dots, \lambda_n)$ and $M = \operatorname{diag}(\mu_1, \dots, \mu_n)$. Thus finally, by virtue of the proof of (*ii*) and the choice of f,

$$f(A) = S \operatorname{diag}(f(\lambda_1), \dots, f(\lambda_n))S^{-1} = SMS^{-1} = B \ .$$

(*iv*): Let $f : D_f \to \mathbb{C}$ be holomorphic on an open neighborhood U_f of $\sigma(A)$. Write U'_f for the set, where $f(\overline{z}) = \overline{f(z)}$ holds. Then, by assumption, $U''_f := U_f \cap U'_f$ is also an open neighborhood of $\sigma(A)$. Before we proceed, let us do a little symmetrization trick:

For every $D \subseteq \mathbb{C}$, write $\overline{D} := \{\overline{z} \mid z \in D\}$. The set $\overline{U''_f}$ is open and by construction $\overline{\sigma(A)} \subseteq \overline{U''_f}$. But we know from linear algebra that $\overline{\sigma(A)} = \sigma(A)$, so $\overline{U''_f}$ is also an open neighborhood of $\sigma(A)$. Because of this, $U'''_f := U''_f \cap \overline{U''_f}$ is an open neighborhood of $\sigma(A)$, on which the two properties of f mentioned above remain to hold and which is additionally symmetric in the sense that

$$\overline{U'''_f} = \overline{U''_f \cap \overline{U''_f}} = \overline{U''_f} \cap \overline{\overline{U''_f}} = \overline{U''_f} \cap U''_f = U'''_f \ . \tag{1.5}$$

Now, choose an admissible cycle Γ for $\sigma(A)$ and U'''_f. For every curve $\gamma : [a, b] \to D$ and every chain $\Gamma = \gamma_1 \oplus \dots \oplus \gamma_n$, set $\overline{\gamma} : [a, b] \to \overline{D}, t \mapsto \overline{\gamma(t)}$ and $\overline{\Gamma} := \overline{\gamma_1} \oplus \dots \oplus \overline{\gamma_n}$, respectively. $-\overline{\Gamma}$ is a cycle of rectifiable Jordan curves in $\overline{U'''_f} \setminus \overline{\sigma(A)} \overset{(1.5)}{=} U'''_f \setminus \sigma(A)$. Use the definition and proposition 1.2.7 (*iii*) to show that the winding numbers behave accordingly for $-\overline{\Gamma}$ to be admissible for $\sigma(A)$ and U'''_f. Verify elementarily that, if $\zeta I - A$ is invertible, then so is $\overline{\zeta} I - \overline{A}$ with inverse $\overline{(\zeta I - A)^{-1}}$. Now, use this result, the linearity of the *complex conjugation* as well as the properties $f(\overline{z}) = \overline{f(z)}$ and $A = \overline{A}$ to show that for every curve $\gamma_i : [a_i, b_i] \to \mathbb{C}$ appearing in Γ, every admissible sequence of partitions $(P_n)_{n\in\mathbb{N}}$ of $[a_i, b_i]$ and every corresponding tags $(T_n)_{n\in\mathbb{N}}$

$$-\overline{S((f \circ \Gamma) \cdot ((\Gamma \cdot I) - A)^{-1}, \Gamma, P_n, T_n)} = S((f \circ -\overline{\Gamma}) \cdot ((-\overline{\Gamma} \cdot I) - A)^{-1}, -\overline{\Gamma}, P_n, T_n) \ .$$

Let $n \to \infty$, use the arguments given above and multiply both sides with $1/(2\pi i)$ to find

$$\overline{f(A)} = \overline{\frac{1}{2\pi i} \int_\Gamma \frac{f(\zeta)}{\zeta I - A} \, d\zeta} = \frac{1}{2\pi i} \int_{-\overline{\Gamma}} \frac{f(\zeta)}{\zeta I - A} \, d\zeta = f(A) \ .$$

(*v*): Use the distributive law for matrix-vector multiplication to obtain a compatibility statement similar to proposition 1.2.9 (*iii*) with a vector $v \in \mathbb{C}^n$ playing the role of the element of the Banach algebra. Use the matrix-valued version of this result and verify that $(\zeta I - A)^{-1}v = (\zeta - \lambda)^{-1}v$ for every eigenvector $v \in \mathbb{C}^n \setminus \{0_{\mathbb{C}^n}\}$ of A for the eigenvalue λ and every $\zeta \in \rho(A)$. Finish by applying the complex-valued version of the argument.

$\square$

Chapter 2

Theoretical results

2.1 Preliminaries

Before we establish and discuss the actual results about *generators* for *continuous-time Markov chains*, let us quickly recapitulate some basic definitions. This part of the section mostly follows [Norr]. For the definition of a *discrete-time Markov chain*, let $(\Omega, \mathcal{F}, \mathbb{P})$ be a *probability space* and let I be a finite set of cardinality $d \in \mathbb{N}$. Additionally, assume $\lambda = (\lambda_i)_{1 \leq i \leq d}$ to be a *distribution on* I and $P = (p_{ij})_{1 \leq i,j \leq d}$ to be a *stochastic matrix*, meaning that every entry of P is a non-negative real number and P has row-sums one, i. e. $\sum_{j=1}^{d} p_{ij} = 1$ for every $i \in \{1, \ldots, d\}$.

Definition 2.1.1 (Discrete-time Markov chain)**.** A stochastic process $(X_n)_{n \in \mathbb{N}_0}$ from Ω to I is called a *discrete-time Markov chain on the state space I with initial distribution λ and transition matrix P*, if for every $n \in \mathbb{N}_0$ and every $i_0, \ldots, i_{n+1} \in I$

(i) $\mathbb{P}(X_0 = i_0) = \lambda_{i_0}$,

(ii) $\mathbb{P}(X_{n+1} = i_{n+1} \mid X_0 = i_0, \ldots, X_n = i_n) = p_{i_n i_{n+1}}$.

In this case, we write $(X_n)_{n \in \mathbb{N}_0} \sim Markov(\lambda, P)$.

To be able to define *continuous-time Markov chains*, we need to establish a few more concepts: for any complex square matrix A, the holomorphic functional calculus gives rise to the expression

$$\exp(A) = \sum_{n=0}^{\infty} \frac{A^n}{n!} ,$$

which we shall call the *matrix exponential of* A from now on. Assume Q to be a Q*-matrix*, meaning that every off-diagonal entry is a non-negative real number and Q has row-sums zero. To avoid measurability problems, we consider only *right-continuous* stochastic processes $(X_t)_{t \in [0,\infty)}$ from Ω to I, i. e. for every $\omega \in \Omega$ and every $t \geq 0$ there shall always exist $\varepsilon > 0$ such that $X_s(\omega) = X_t(\omega)$ for every $t \leq s \leq t + \varepsilon$.

We implicitly sum up a few results from Markov chain theory by making the following definition:

Definition 2.1.2 (Continuous-time Markov chain)**.** A right-continuous stochastic process $(X_t)_{t\in[0,\infty)}$ from Ω to I is called a *continuous-time Markov chain on the state space I with initial distribution λ and generator matrix Q*, if for every $n \in \mathbb{N}_0$, $0 \leq t_0 \leq t_1 \leq \ldots \leq t_{n+1}$ and every $i_0, \ldots, i_{n+1} \in I$

(i) $\mathbb{P}(X_0 = i_0) = \lambda_{i_0}$,

(ii) $\mathbb{P}(X_{t_{n+1}} = i_{n+1} \mid X_{t_0} = i_0, \ldots, X_{t_n} = i_n) = p_{i_n i_{n+1}}(t_{n+1} - t_n)$.

where $(p_{ij}(t))_{1\leq i,j\leq d} := P(t) := \exp(tQ)$. In this case, we write $(X_n)_{n\in\mathbb{N}_0} \sim Markov(\lambda, Q)$.

With these definitions, we can discuss the ideas given in the introduction more meaningfully: since in reality we rarely ever observe the course of events from the very start, the initial distributions are mostly negligible for us.

Now, if we want to understand and eventually forecast the behavior of a real process, then an intuitive approach, which is also followed by Israel et al., would be to determine or rather estimate a corresponding transition matrix P for a fixed time span like one year. This matrix will only give rise to a discrete-time model, which we subsequently wish to embed into a continuous-time Markov chain. Hence, we assume P to be $P(1)$ of a continuous-time Markov chain and try to find a generator Q for $(P(t))_{t\in[0,\infty)}$. This means, our central equation reads precisely as

$$\exp(Q) = P \ , \tag{2.1}$$

where P is a stochastic and Q is a Q-matrix. A possible alternative to the above procedure could be to limit our estimation to those transition matrices P that are $\exp(Q)$ for some Q-matrix Q. This technique, however, will not be discussed in more detail in this thesis.

As indicated in the introduction, we will use the holomorphic functional calculus to try and define a matrix logarithm to solve this equation. Since such an approach is based on complex functions, we lastly want to take a closer look at the difficulties that arise, when trying to generalize the logarithm function from the real line to the complex plane:

If $z \in \mathbb{C} \setminus \{0\}$ is any non-zero complex number, than there exist infinitely many *logarithms of z*, i. e. complex numbers $(z_k)_{k\in\mathbb{Z}}$ such that $e^{z_k} = z \ \forall\, k \in \mathbb{Z}$. Namely, let z be given in *polar coordinates* as $re^{i\theta}$. This representation is unique, if we require $r > 0$ and $\theta \in (-\pi, \pi]$, which we shall do throughout this thesis. The logarithms of z can then be described as $z_k := \ln(r) + i(\theta + 2\pi k)$, $k \in \mathbb{Z}$, where "ln" denotes the real logarithm function. Thus, if we tried to define a logarithm function on the complex plane, we would have infinitely many possible output values to choose from for every single input argument. To recover a proper functional structure, one makes the following definition:

Definition 2.1.3 (Branches of the complex logarithm)**.** Let $U \subseteq \mathbb{C}$ be an open, connected set. We call a function $f : U \to \mathbb{C}$ a *branch of the complex logarithm on* U, if f is continuous and $e^{f(z)} = z \; \forall \, z \in U$.

It is a standard result from complex analysis that for every open connected set $U \subseteq \mathbb{C} \setminus \{0\}$, there exists exactly one branch f of the complex logarithm such that $\mathrm{Im}(f(z)) \in (-\pi, \pi]$ for every $z \in U$. We call this function the *principal branch of the complex logarithm on* U and denote it by "Log". The other branches of the complex logarithm on U are then given by $f_k := \mathrm{Log} + 2\pi k i$ for $k \in \mathbb{Z}$.

2.2 Embedding problem

Before we start with actual existence results, we notice the following observation by Israel et al. [IRW, p. 9]: it suffices to judge embeddability of a discrete- into a continuous-time Markov chain based on the transition matrix P; no other powers of this matrix need to be considered. To see this, let P be a stochastic matrix and fix a positive integer n. Now, if P has a generator Q, then $\exp(nQ) = \exp(Q)^n$ by virtue of the composition identity, and so nQ is a generator for P^n. Conversely, if P^n has a generator Q, then analogously Q/n is a generator for P. Hence, P^n is embeddable, if and only if P is embeddable. Let us now take a look at a first existence result from [IRW, p. 4]:

Theorem 2.2.1 (Mercator series)**.** Let P be a stochastic matrix and let $\|.\|$ denote an arbitrary matrix norm. If $S := r(P - I) < 1$, then the *Mercator series*

$$\widetilde{Q} := \sum_{n=1}^{\infty} (-1)^{n+1} (P - I)^n / n \tag{2.2}$$

converges geometrically quickly, i. e. $\sup_{N \geq 1} |a_N / q^N| < \infty$ for some $0 \leq q < 1$, where $a_N := \left\| \sum_{n=N+1}^{\infty} (-1)^{n+1} (P - I)^n / n \right\|$. The limit $\widetilde{Q}$ satisfies $\exp(\widetilde{Q}) = P$ and has row-sums zero. If $S \geq 1$, then the series fails to converge.

Proof. Assume $S < 1$. Choose $q := S^{1/2} < 1$ and use the norm axioms for $\|.\|$ to show

$$\left| \frac{a_N}{q^N} \right| \leq \sum_{n=N+1}^{\infty} \frac{\|(P-I)^n\|}{nS^{N/2}} \leq \sum_{n=N+1}^{\infty} \frac{\|(P-I)^n\|}{nS^{n/2}} \leq \sum_{n=1}^{\infty} \frac{\|(P-I)^n\|}{nS^{n/2}} \ .$$

Now, the series on the far right converges by virtue of the *root test*:

$$\left| \frac{\|(P-I)^n\|}{nS^{n/2}} \right|^{1/n} = \frac{1}{n^{1/n}} \ \frac{\|(P-I)^n\|^{1/n}}{S} \ S^{1/2} \to S^{1/2} < 1 \ ,$$

where the asymptotic statement follows from the spectral radius formula. Hence, $\sup_{N \geq 1} |a_N / q^N| \leq \sum_{n=1}^{\infty} \|(P-I)^n\| / (nS^{n/2}) < \infty$ and we have established the geometric rate of convergence.

Next, use the complex version of the *fundamental theorem of calculus* and *term by term differentiation* to show that the complex series $\sum_{n=1}^{\infty} (-1)^{n+1} (z-1)^n / n$ converges to $\mathrm{Log}(z)$ for every $z \in B_1(1)$. The condition $S < 1$ implies that $\sigma(P) \subseteq B_1(1)$. Thus, $\widetilde{Q} = \mathrm{Log}(P)$ by virtue of proposition 1.3.4 (*ii*). The composition identity now gives

$$\exp(\widetilde{Q}) = \exp(\mathrm{Log}(P)) = (\exp \circ \mathrm{Log})(P) = P \ .$$

To prove that $\widetilde{Q}$ has row-sums zero, notice the following little auxiliary result:
If $d \times d$ matrices A and B have row-sums α and β, respectively, then

$$\sum_{j=1}^{d}\sum_{k=1}^{d} a_{ik}b_{kj} = \sum_{k=1}^{d} a_{ik}\left(\sum_{j=1}^{d} b_{kj}\right) = \sum_{k=1}^{d} a_{ik}(\beta) = \beta\left(\sum_{k=1}^{d} a_{ik}\right) = \beta(\alpha) \tag{2.3}$$

and so AB has row-sums $\alpha\beta$.
Hence, as $(P-I)$ has row-sums zero, so does $(P-I)^n$ for every $n \in \mathbb{N}$ and thus also $\widetilde{Q}$.
Now, suppose $S \geq 1$. It can be shown (see [deBo]) that a matrix satisfies $A^n \to 0$ as $n \to \infty$, if and only if $r(A) < 1$. Thus, in this case, some coordinate of the powers of $P - I$ does not converge to zero, which violates the necessary condition for series convergence in that coordinate.

□

The next lemma by Israel et al. [IRW, p. 5] gives a sufficient condition for the above theorem to hold, which is considerably easier to verify and, in fact, is fulfilled by almost all matrices that we encounter in the credit rating context:

Lemma 2.2.2 (Diagonal dominance)**.** Suppose that P is *strictly diagonal dominant*, i. e. $p_{ii} > 1/2$, $i = 1, \ldots, d$. Then, $S < 1$ and the series in (2.2) converges.

Proof. Write $m := \min_{1 \leq i \leq d} p_{ii}$. Without loss of generality, we can assume $m < 1$, because otherwise P was the identity matrix and the assertion would follow immediately. Now, $R := 1/(1-m)(P - mI)$ is a stochastic matrix satisfying

$$P - I = (1-m)(R - I) \ . \tag{2.4}$$

Let $\|.\| := \|.\|_\infty$ denote the ∞-norm given by $\|A\|_\infty := \max_{1\leq i\leq d} \sum_{j=1}^{d} |a_{ij}|$. It follows directly from the definition that every stochastic matrix has an ∞-norm of one, so $\|R\| = 1$. By the triangle inequality, $\|R - I\| \leq 2$ and reinserting from (2.4), $\|P - I\| \leq 2 - 2m$. Use submultiplicativity of the norm inductively to find $\|(P-I)^n\| \leq (2-2m)^n$, so that $\|(P-I)^n\|^{1/n} \leq 2 - 2m$. Now, the spectral radius formula gives

$$S = \lim_{n\to\infty} \|(P-I)^n\|^{1/n} \leq 2 - 2m$$

and, as $m > 1/2$, the right hand-side is less than one. □

It is shown in [Cuth] that under the condition of the last lemma, the limit matrix $\widetilde{Q}$ from theorem 2.2.1 is actually the only possible generator. Notice that $\widetilde{Q}$ is not automatically a Q-matrix, as negative off-diagonal entries might arise. We will take a look at two simple methods for fixing this problem due to Israel et al. in the next chapter. Let us now consider an example which shows that our search for a generator should not rely on theorem 2.2.1 only:

Example 2.2.3. (i) [SiSp, p. 8f.]: There is a stochastic matrix P, where the series in theorem 2.2.1 fails to converge, but a generator Q still exists:

$$P := \begin{pmatrix} .3654 & .3762 & .2584 \\ .3292 & .3567 & .3141 \\ .4040 & .3188 & .2772 \end{pmatrix}.$$

$\lambda = .053i$ is an eigenvalue of P and thus, $S \not< 1$. But a generator Q for P can be given by

$$Q := \begin{pmatrix} -1.805 & 1.718 & 0.087 \\ 0.044 & -1.784 & 1.740 \\ 2.262 & 0.017 & -2.279 \end{pmatrix}.$$

(ii) [IRW, p. 14]: There is a stochastic matrix P, where the series in theorem 2.2.1 converges to a limit which is not a generator, but a generator Q still exists:

$$P := \begin{pmatrix} .284779445 & .284035268 & .283826586 & .147358701 \\ .284191780 & .284779445 & .284035268 & .146993507 \\ .283487477 & .284191780 & .284779445 & .147541298 \\ .284543931 & .283487477 & .284191780 & .147776812 \end{pmatrix}.$$

Then, the series in theorem 2.2.1 converges to

$$\widetilde{Q} = \begin{pmatrix} -6.194496074 & 2.807322994 & 1.374197570 & 2.012975514 \\ 3.882167062 & -6.194496074 & 2.807322994 & -.494993979 \\ -.954631245 & 3.882167062 & -6.194496074 & 3.266960260 \\ 6.300566216 & -.954631245 & 3.882167062 & -9.228102034 \end{pmatrix},$$

which has negative off-diagonal entries. But a generator Q for P can be given by

$$Q := \begin{pmatrix} -5.642931358 & 5.642931358 & 0.000000000 & 0.000000000 \\ 0.000000000 & -5.642931358 & 5.642931358 & 0.000000000 \\ 0.000000000 & 0.000000000 & -5.642931358 & 5.642931358 \\ 61.410871840 & 0.000000000 & 0.000000000 & -61.410871840 \end{pmatrix} .$$

We will now take a look at a few necessary conditions for the existence of a generator from different authors. We shall call a state $j \in I$ accessible from $i \in I$, if there is a sequences of states $i = k_0, k_1, \ldots, k_n = j$ such that $p_{k_\ell k_{\ell+1}} > 0$ for every $\ell \in \{0, \ldots, n-1\}$.

Theorem 2.2.4 (Necessary conditions)**.** Let P be a stochastic matrix. For P to have a generator, it is necessary that

(i) [King, p. 14]: $\det(P) > 0$,

(ii) [Good]: $\det(P) \le \prod_{i=1}^{d} p_{ii}$,

(iii) [Chu]: There are no states i and j such that j is accessible from i, but $p_{ij} = 0$.

Proof. (i): We prove the well-known formula that for every $d \times d$ matrix Q

$$\det(\exp(Q)) = \exp(\operatorname{tr}(Q)) . \tag{2.5}$$

To do so, write Q in (possibly complex) *Jordan canonical form* $Q = SJS^{-1}$, where $S \in \mathbb{C}^{d\times d}$ is invertible and $J \in \mathbb{C}^{d\times d}$ is an upper triangular matrix having the (not necessarily distinct) eigenvalues $\mu_1, \ldots, \mu_d$ of Q on its diagonal. Now, as seen in the proof of proposition 1.3.7,

$$\exp(Q) = \exp(SJS^{-1}) = S\exp(J)S^{-1} .$$

Compute the matrix products appearing in the power series to find that $\exp(J)$ is again an upper-triangular matrix and that the diagonal of $\exp(J)$ reads $\exp(\mu_1), \ldots, \exp(\mu_d)$. Deduce

$$\det(\exp(Q)) = \det(S\exp(J)S^{-1}) = \det(S)\det(\exp(J))\det(S^{-1}) = \det(\exp(J)) = \exp(\mu_1 + \ldots + \mu_d) .$$

But we know from linear algebra that *similar matrices*, i. e. matrices in a relation like Q and J, have the same trace and so $\mu_1 + \ldots + \mu_d = \operatorname{tr}(J) = \operatorname{tr}(Q)$ and (2.5) follows.
Now, if a matrix P should be $\exp(Q)$ for some real matrix Q (whether a generator or not), then necessarily $\det(P) = \det(\exp(Q)) = \exp(\operatorname{tr}(Q)) > 0$.

(ii): Let P have a generator Q and write $P(t) := \exp(tQ)$ for $t \geq 0$. We consider the real-valued function $p_{ii}(t) := (P(t))_{ii}$. This mapping is represented everywhere by a convergent power series, since

$$p_{ii}(t) = \left(\sum_{n=0}^{\infty} \frac{(tQ)^n}{n!} \right)_{ii} = \sum_{n=0}^{\infty} (Q^n)_{ii} \frac{t^n}{n!} \ .$$

Thus, $p_{ii}(t)$ is differentiable and we can use *term by term differentiation* to find

$$\begin{aligned} p'_{ii}(t) &= \sum_{n=0}^{\infty} (Q^n)_{ii} \frac{nt^{n-1}}{n!} \\ &= \left(\sum_{n=1}^{\infty} Q^n \frac{t^{n-1}}{(n-1)!} \right)_{ii} \\ &= \left(Q \sum_{n=1}^{\infty} \frac{(tQ)^{n-1}}{(n-1)!} \right)_{ii} \\ &= (Q \exp(tQ))_{ii} \ . \end{aligned}$$

Since Q has non-negative off-diagonal entries and $P(t) = \exp(tQ)$ is a stochastic matrix, for every $t \geq 0$

$$p'_{ii}(t) = (Q \exp(tQ))_{ii} = \sum_{j=1}^{d} q_{ij} (P(t))_{ji} \geq q_{ii} (P(t))_{ii} = q_{ii} p_{ii}(t) \ .$$

We follow [Zhao] and deduce

$$\frac{d}{dt} \Big(p_{ii}(t) \exp(-tq_{ii}) \Big) = \Big(p'_{ii}(t) - q_{ii} p_{ii}(t) \Big) \exp(-tq_{ii}) \geq 0 \ .$$

This means that $p_{ii}(t) \exp(-tq_{ii}) \geq p_{ii}(0) \exp(-0\, q_{ii}) = 1$, which is equivalent to $p_{ii}(t) \geq \exp(tq_{ii})$ and especially implies $p_{ii} \geq \exp(q_{ii})$. Thus, using (2.5),

$$\prod_{i=1}^{d} p_{ii} \geq \prod_{i=1}^{d} \exp(q_{ii}) = \exp(\mathrm{tr}(Q)) = \det(P) \ .$$

(iii): We prove the *Lévy dichotomy*, which states the following: if $(p_{ij}(t))_{1 \leq i,j \leq d} := P(t) := \exp(tQ)$ with some generator Q, then for every two states $i, j \in I$ we find either $p_{ij}(t) > 0 \ \forall\, t > 0$ or $p_{ij}(t) = 0 \ \forall\, t > 0$. Assume $p_{ij}(t) = 0$ for some $t > 0$. Since $p_{jj}(s) \to 1$ as $s \to 0$, there has to be $s^* > 0$ such that $p_{jj}(s) > 0$ for every $s < s^*$. Thus, $p_{ij}(t/n) > 0$ for sufficiently large $n \in \mathbb{N}$. Using a generalization of the *Cauchy product theorem*, we can show that

$$\exp(A + B) = \exp(A) \exp(B) \tag{2.6}$$

for commuting matrices A, B. This gives $P(t) = P(t/n)(P(t/n))^{n-1}$ and especially $p_{ij}(t) \geq p_{ij}(t/n) \cdot (p_{jj}(t/n))^{n-1}$. Hence, $p_{ij}(t/n) = 0$ for all sufficiently large $n \in \mathbb{N}$. But a standard result from complex analysis says that the zeros of an entire function, like $p_{ij}(s)$, can only have an accumulation point (here 0), if the function is identical to zero. This means $p_{ij}(s) = 0 \; \forall \, s > 0$, which proves the assertion.

$\square$

Let us rephrase part (*iii*) of the last theorem in a more explicit version due to Israel et al. [IRW, p. 17]:

Lemma 2.2.5 (Quantitative Lévy dichotomy)**.** If P has a generator, then the entries of P must satisfy

$$p_{ij} \geq m^m r^r (m+r)^{-(m+r)} \sum_{k=1}^{d} (p_{ik} - b_m)(p_{kj} - b_r) \mathbb{1}_{p_{ik} > b_m, p_{kj} > b_r}$$

for all positive integers m and r. Here, $b_m := 1 - \sum_{\ell=0}^{m} e^{-\lambda} \lambda^\ell / \ell!$, which equals the probability that a Poisson random variable with mean $\lambda := \max 1 \leq i \leq d(-q_{ii})$ takes a value greater than m. Additionally, $\mathbb{1}_A$ denotes the indicator function of the event A.

Proof. See [IRW, p. 26]. $\square$

To finish this section, I want to quote a few more existence results by different authors. The proofs can be found in the respective sources:

Proposition 2.2.6 (Additional results)**.** Let P be a $d \times d$ transition matrix.

(*i*) [Elf]: If any eigenvalues $\lambda \neq 1$ of P satisfies $|\lambda| = 1$, then no generator for P exists.

(*ii*) [Elf]: If any negative eigenvalue of P has odd algebraic multiplicity, then no generator for P exists.

(*iii*) [King, p. 16]: The set of all embeddable matrices is relatively closed inside the set of all matrices with positive determinant.

(*iv*) [King, p. 22]: A *regular* matrix P is embeddable, if and only if for every $n \in \mathbb{N}$ there exists a stochastic matrix $P_{1/n}$ such that $(P_{1/n})^n = P$.

(*v*) [Runn]: For P to be embeddable, all eigenvalues of P have to lie in $D \cup \overline{D}$, where D is the region of the complex plane bounded by the curve

$$\left\{ \Big(\exp\big(-t + t\cos(2\pi/d)\big) \cos\big(t \sin(2\pi/d)\big), \exp\big(-t + t\cos(2\pi/d)\big) \sin\big(t \sin(2\pi/d)\big) \Big) \mid 0 \leq t \leq \pi / \sin(2\pi/d) \right\}.$$

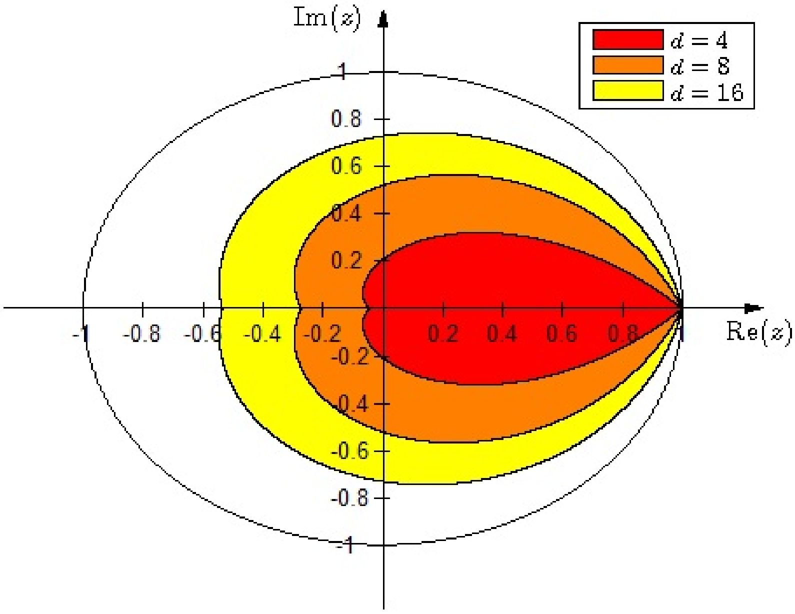

Figure 2.1: The unit circle and the region mentioned in part (v) of the last proposition for different values of the matrix dimension d.

Summarizing, the task of finding necessary conditions for the existence of a generator (or equivalently, sufficient conditions for non-embeddability) seems to be solved quite comprehensively. In contrast, however, there are only few results about sufficient requirements for a transition matrix to possess a generator, which is somewhat unsatisfactory for practical applications.

2.3 Identification problem

Let us first consider an example from [IRW, p. 15] which motivates our endeavor in this section:

Example 2.3.1 (Ambiguity)**.** There is a stochastic matrix P that has two generators Q_1 and Q_2:

$$P := \begin{pmatrix} (2+3b)/5 & (2-2b)/5 & (1-b)/5 \\ (2-2b)/5 & (2+3b)/5 & (1-b)/5 \\ (2-2b)/5 & (2-2b)/5 & (1+4b)/5 \end{pmatrix},$$

where $b := e^{-4\pi}$. Two generators Q_1 and Q_2 can be given by

$$Q_1 := \begin{pmatrix} -2\pi & 2\pi & 0 \\ 0 & -2\pi & 2\pi \\ 4\pi & 0 & -4\pi \end{pmatrix}$$

and

$$Q_2 := \begin{pmatrix} -12\pi/5 & 8\pi/5 & 4\pi/5 \\ 8\pi/5 & -12\pi/5 & 4\pi/5 \\ 8\pi/5 & 8\pi/5 & -16\pi/5 \end{pmatrix}.$$

Next, I cite some sufficient conditions for a stochastic matrix P to have at most one generator, which were observed by Israel et al. [IRW, p. 16]:

Theorem 2.3.2 (Sufficient conditions)**.** Let P be a stochastic matrix.

(i): If $\det(P) > 1/2$, then P has at most one generator.

(ii): If $\det(P) > 1/2$ and $\|P-I\| < 1/2$ (in an arbitrary operator norm), then the only possible generator for P is $\mathrm{Log}(P)$.

(iii): If P has distinct eigenvalues and $\det(P) > e^{-\pi}$, then the only possible generator for P is $\mathrm{Log}(P)$.

Proof. (i): Assume P had a generator Q. Since $\det(P) > 1/2$ and $\det(P) = \exp(\mathrm{tr}(Q))$ by (2.5), it follows that $\mathrm{tr}(Q) > -\ln(2)$. Write $\|.\|$ for the operator norm induced by the L^1 norm for vectors, i. e. $\|Q\| = \max_{1\le j\le d}\sum_{i=1}^{d}|q_{ij}|$. But for a Q-matrix $|q_{ij}| \le |q_{ii}| = -q_{ii}$, so $\|Q\| \le -\mathrm{tr}(Q) < \ln(2)$. We finish the proof by showing that the matrix exponential is one-to-one on $\{Q \in \mathbb{C}^{d\times d} \mid \|Q\| < \ln(2)\}$: Therefore, let Q_1, Q_2 be two complex $d \times d$ matrices such that $\|Q_1\|, \|Q_2\| < r$ for some $0 \leq r < \ln(2)$ and $Q_1 \neq Q_2$, which is equivalent to $\|Q_1 - Q_2\| > 0$. Then, using the triangle inequality and the formula $A^n - B^n = \sum_{k=0}^{n-1} A^k(A-B)B^{n-1-k}$ for matrices A, B, which is a generalization of the well-known identity $x^n - y^n = \sum_{k=0}^{n-1} x^k(x-y)y^{n-1-k}$ for numbers x, y, deduce

$$\begin{aligned}
\|\exp(Q_1) - \exp(Q_2) - (Q_1 - Q_2)\| &\leq \sum_{n=2}^{\infty} \|Q_1^n - Q_2^n\|/n! \\
&\leq \sum_{n=2}^{\infty}\sum_{k=0}^{n-1} \|Q_1^k(Q_1 - Q_2)Q_2^{n-1-k}\|/n! \\
&\leq \|Q_1 - Q_2\| \sum_{n=2}^{\infty} nr^{n-1}/n! \\
&= \|Q_1 - Q_2\|(e^r - 1) < \|Q_1 - Q_2\| \ .
\end{aligned}$$

This means that $\exp(Q_1) - \exp(Q_2) \neq 0$, which is equivalent to $\exp(Q_1) \neq \exp(Q_1)$.

(ii): Let $\|.\|$ denote any operator norm such that $\|P - I\| =: r < 1/2$. As a corollary of the spectral radius formula, $S = r(P - I) \leq \|P - I\| = r < 1/2$, so the series in theorem 2.2.1 converges. Use the triangle inequality and the power series expansion of the real logarithm ln in $(-1, 1]$ to deduce

$$\|\operatorname{Log}(P)\| \leq \sum_{n=0}^{\infty} r^n/n = -\sum_{n=0}^{\infty} (-1)^{n+1}(-r)^n/n = -\ln(1-r) < \ln(2) \ .$$

The assertion now follows from the proof of (i).

(iii): Again, assume P had a generator Q. Since P has distinct eigenvalues and Q commutes with $\exp(Q) = P$, we know from proposition 1.3.7 (iii) that Q has to be $g(P)$ for some $g \in H(P)$. Use part (ii) of the same proposition and the composition identity to deduce that g has to satisfy $e^{g(\lambda)} = \lambda$ for every eigenvalue λ of P. This means $g(\lambda)$ is one of the logarithms of λ for every eigenvalue λ of P. Since $\det(P) > e^{-\pi}$, the complex number 0 can not be an eigenvalue of P and so we know from complex analysis that there always exists a branch f of the complex logarithm which is holomorphic in an open neighborhood of $\sigma(P)$ and satisfies $f(\lambda) = g(\lambda)$ for every λ. This means that $Q = g(P) = f(P)$. We will show that necessarily $f = \operatorname{Log}$: it follows from $\det(P) > e^{-\pi}$ that $0 \geq \operatorname{tr}(Q) > -\pi$. Write $r := -\operatorname{tr}(Q)$. Now, $Q + rI$ is a matrix that has only non-negative entries. By virtue of proposition 1.3.4 and the spectral mapping theorem, $\sigma(Q + rI) = \sigma(Q) + r$. Since Q is a generator, $\sigma(Q) \cap (0, \infty) = \emptyset$ and so the maximum non-negative eigenvalue of $Q + rI$ is r. The well-known *Perron-Frobenius theorem* now says that the absolute value of every element of $\sigma(Q + rI)$ can not be greater than r. In particular, the imaginary part of every eigenvalue of Q has to lie in $[-r, r]$. But Log is the only branch of the complex logarithm, whose imaginary part lies in $(-\pi, \pi)$, so f can only be Log. $\square$

Existence and uniqueness of a generator for a stochastic matrix with only real distinct eigenvalues can be evaluated easily using the following criterion from [IRW, p. 16]:

Proposition 2.3.3 (Real distinct eigenvalues)**.** Let P be a stochastic matrix with only real distinct eigenvalues.

(i) If all eigenvalues of P are positive, then $\mathrm{Log}(P)$ is the only possible generator for P.

(ii) If P has any negative eigenvalues, then there is no generator for P.

Proof. (i): Suppose Q is a generator for P. First we show that, in this situation, the eigenvalues of Q have to be real. Since $P = \exp(Q)$ has distinct eigenvalues and $\sigma(\exp(Q)) = \exp(\sigma(Q))$ by virtue of the spectral mapping theorem, exp is one-to-one on $\sigma(Q)$. Now, if we assumed an eigenvalue $\lambda \in \mathbb{C} \setminus \mathbb{R}$ of Q, then $\overline{\lambda} \neq \lambda$ would also be in $\sigma(Q)$, because Q is a real matrix. But as the eigenvalues of $P = \exp(Q)$ are real, it would follow that $\exp(\overline{\lambda}) = \overline{\exp(\lambda)} = \exp(\lambda)$, which contradicts injectivity of exp on $\sigma(Q)$.
Since the matrix P has distinct eigenvalues, we know from the proof of the last theorem that $Q = f(P)$, where f is some branch of the complex logarithm. Now, the above arguments force $f(\lambda)$ to be real for every $\lambda \in \sigma(Q)$. Thus, the principal branch Log is the only possible candidate for f, so $Q = \mathrm{Log}(P)$.
(ii): If P has any negative eigenvalues, then Log can not be defined in an open neighborhood of $\sigma(Q)$ and thus, by virtue of the proof of (i), no generator exists.

□

The last result due to Israel et al. [IRW, p. 18] is very useful for making an algorithmic approach to finding generators for continuous-time Markov chains:

Theorem 2.3.4 (Limiting the possible branches)**.** Let P be a stochastic matrix with distinct eigenvalues. If Q is a generator for P, then every eigenvalue λ of Q satisfies $|\,\mathrm{Im}(\lambda)| \leq -\ln(\det(P))$ (necessarily $\det(P) > 0$ by virtue of theorem 2.2.4 (i)). In particular, only finitely many branches of the complex logarithm could possibly give rise to a generator for P.

Proof. We follow the proof of theorem 2.3.2 (i) closely. If P has a generator Q, then again $\exp(\mathrm{tr}(Q)) = \det(P)$, so $\mathrm{tr}(Q) = \ln(\det(P))$. Write $r := -\,\mathrm{tr}(Q)$. Then, as argued before, $Q + rI$ is a matrix with non-negative entries and maximum non-negative eigenvalue r. Deduce with the *Perron-Frobenius theorem* that $|\,\mathrm{Im}(\lambda)| \leq |\lambda| \leq r = -\ln(\det(P))$ for every eigenvalue λ of Q.

□

We draw the conclusion that the identification problem has been solved a bit more gratifyingly in the literature than the embedding problem of the last section. A lot of easy-to-check sufficient conditions involving only the determinant of the matrix were found. Necessary requirements for uniqueness of a generator would have been of no great value, anyway.

Chapter 3

Algorithms and numerical examples

3.1 Algorithms to search for a generator

Let us first consider the case where the transition matrix P has distinct eigenvalues. In [IRW, p. 18f.], the following algorithm to search for a generator in this situation was proposed:

We have argued in the last section that a possible generator for a stochastic matrix P with distinct eigenvalues can only have the form $Q = f(P)$, where f is some branch of the complex logarithm which is analytic in a neighborhood of the spectrum of P. Furthermore, we have discussed that such a function f has to satisfy $f(\lambda_j) = \ln(r_j) + i(\theta_j + 2\pi k_j)$, where $\lambda_1 = r_1 e^{i\theta_1}, \ldots, \lambda_d = r_d e^{i\theta_d}$ are the eigenvalues of P in unique polar coordinate representation with $r_j > 0$, $\theta_j \in (-\pi, \pi]$, $j = 1, \ldots, d$ and $k_1, \ldots, k_d$ are integers. Theorem 2.3.4 gives an additional bound on the imaginary part of the values of f, which ensures that only finitely many branches need to be considered. Thus, we could simply search for a generator for P by computing $f(P)$ for each branch f in question and check, if $f(P)$ is a generator. But proposition 1.3.7 (ii) tells us that $f(P) = g(P)$ for any other function g satisfying $f(\lambda) = g(\lambda)$ for every eigenvalue λ of P. Hence, the possible candidates can be computed more easily using an *interpolation polynomial* g from numerical mathematics. The complete algorithm reads as follows:

Algorithm 3.1.1 (Distinct eigenvalues)**.** Let P be a stochastic matrix.

(i) Compute the eigenvalues $\lambda_1, \ldots, \lambda_d$ of P and validate that they are distinct.

(ii) For each eigenvalue $\lambda_j = r_j e^{i\theta_j}$, $r_j > 0, \theta_j \in (-\pi, \pi]$ choose an integer k_j, such that $f(\lambda_j) := \ln(r_j) + i(\theta_j + 2\pi k_j)$ satisfies $|\operatorname{Im}(f(\lambda_j))| = |\theta_j + 2\pi k_j| \leq -\ln(\det(P))$.

(iii) Write g for the *Lagrange interpolation polynomial for the data points* $(\lambda_1, f(\lambda_1)), \ldots, (\lambda_d, f(\lambda_d))$, which is given for $\lambda \in \mathbb{C}$ by

$$g(\lambda) := \sum_{j=1}^{d} \left(\prod_{\substack{k=1 \\ k \neq j}}^{d} \frac{\lambda - \lambda_k}{\lambda_j - \lambda_k} \right) f(\lambda_j) \; .$$

(*iv*) Compute the matrix $g(P)$ for this g and check, if it is a generator for P.

(*v*) Repeat from (*ii*) with a different choice for the integers $k_1, \ldots, k_d$. Continue until all admissible constellations have been considered.

Remark 3.1.2 (Modifications of the above algorithm)**.** A very similar algorithm to the one presented above already appeared earlier in [SiSp, pp. 16-19, 25-26]. The only real difference was the bounding result for the eigenvalues of the generator which was invoked to limit the number of possible branches of the complex logarithm. As seen above, Israel et al. used theorem 2.3.4, whereas Singer and Spilerman relied on proposition 2.2.6 (*v*). [SiSp, pp. 19-25, 26-29] also contains an adaption of the above algorithm to the case, where the eigenvalues of P are not distinct. The essential idea is to require certain derivatives of the interpolation polynomial and the branches of the complex logarithm to agree at repeated eigenvalues. However, since in practice most transition matrices seem to have distinct eigenvalues, this method is not discussed in more detail here.

If an exact generator could not be found, the following modification techniques from [IRW, p. 6] at least allows us to determine reasonable approximations based on theorem 2.2.1:

Algorithm 3.1.3 (Non-negativity modifications)**.** Empirically, if negative off-diagonal entries arise in the matrix $\widetilde{Q}$ from theorem 2.2.1, then they will usually be quite small. A naive approach would be to change all negative off-diagonal entries to zero and adjust the diagonal entries accordingly to ensure that the new matrix still has row-sums zero. This means, we would modify $\widetilde{Q}$ to Q by setting

$$q_{ij} := \max(\widetilde{q}_{ij}, 0), i \neq j; \quad q_{ii} := \widetilde{q}_{ii} + \sum_{\substack{j=1 \\ j \neq i}}^{d} \min(\widetilde{q}_{ij}, 0) \ . \tag{3.1}$$

A more refined method is to distribute the negative excess over the diagonal entry and all non-negative entries of a row, proportional to their size. We can leave rows that are identical to zero as they are, since they need no modification. For each other row i we could write

$$G_i := |\widetilde{q}_{ii}| + \sum_{\substack{j=1 \\ j \neq i}}^{d} \max(\widetilde{q}_{ij}, 0); \quad B_i := \sum_{\substack{j=1 \\ j \neq i}}^{d} \min(\widetilde{q}_{ij}, 0) \ .$$

for the "good" and "bad" totals and find $G_i > 0$.

We would then set

$$q_{ij} := \begin{cases} 0, & i \neq j \text{ and } \widetilde{q}_{ij} < 0 \\ \widetilde{q}_{ij} - B_i |\widetilde{q}_{ij}| / G_i, & \text{otherwise} \end{cases} . \tag{3.2}$$

When the diagonal entries of $\widetilde{Q}$ have a large absolute and the negative off-diagonal entries remain comparatively small, the two modifications will lead to very similar results.

The next algorithm is taken from [JLT, p. 504f.]. It was deduced in an economic model of greater scope with an underlying assumption that the probability of a debtor's credit rating changing more than once in a year is negligible.

Algorithm 3.1.4 (Explicit construction)**.** Let P be a $d \times d$transition matrix. For $i = 1, \dots, d-1$ and $j = 1, \dots, d$ set

$$q_{ii} := \ln(p_{ii}); \quad q_{ij} := p_{ij} \ln(p_{ii})/(p_{ii} - 1), \ i \neq j \ . \tag{3.3}$$

Treat the case $i = d$ by $q_{dj} := 0$ for $j = 1, \dots, d$.

To finish this section, we take a look at a simple rule from [IRW, p. 15], which allows substantiated choice of a generator, if there are multiple options:

Algorithm 3.1.5 (Selection rule)**.** If a stochastic matrix P possesses multiple generators, then choose the one which minimizes the quantity

$$J = \sum_{i,j=1}^{d} |i - j| \, q_{ij}$$

This way, you ensure that jumping to a remote rating is not too likely, which is adequate for credit transition matrices.

3.2 Numerical examples

Let us now try to apply the developed theory to some real examples. It has to be said in advance that, up to today, this unfortunately does not work as expected for credit rating transition matrices. This is due to the fact that, so far, such matrices never pass the test of the Lévy dichotomy. Hence, an exact generator will never exist and most of the results become obsolete. If we consider the developments of rating transition matrices until today, we can at least hope that with an increasing number of observations, this problem could disappear. Also, the situation might already be more favorable these days in other fields of research, where Markov chain models are used.
In the following, let us "at least" try to find approximate generators for some real credit rating matrices. We will have to make up more or less heuristic rules and hypotheses to guide our search. Despite the adverse circumstances, we shall try to investigate, if one of the methods from the previous section is superior to the others and, if there are factors in the underlying data which influence the outcome notably.

Example 3.2.1 (Standard & Poor's 1981-2013)**.** Our starting point shall be the average annual transition matrix for the years $1981-2013$ provided by the credit rating agency "Standard & Poor's" in their 2013 corporate default and rating transition study [SnP1]. We reallocate the share of the transitions to the status "Not rated" using the intuitive equation (31) from [JLT, p. 506]. A representation of the resulting matrix P with four decimal places reads

$$P \approx \begin{pmatrix} 0.9004 & 0.0918 & 0.0055 & 0.0005 & 0.0008 & 0.0003 & 0.0005 & 0.0000 \\ 0.0057 & 0.9006 & 0.0861 & 0.0058 & 0.0006 & 0.0007 & 0.0002 & 0.0002 \\ 0.0003 & 0.0196 & 0.9164 & 0.0575 & 0.0037 & 0.0015 & 0.0002 & 0.0007 \\ 0.0001 & 0.0013 & 0.0383 & 0.9096 & 0.0408 & 0.0063 & 0.0014 & 0.0022 \\ 0.0002 & 0.0004 & 0.0017 & 0.0576 & 0.8451 & 0.0786 & 0.0076 & 0.0089 \\ 0.0000 & 0.0003 & 0.0012 & 0.0025 & 0.0621 & 0.8368 & 0.0505 & 0.0465 \\ 0.0000 & 0.0000 & 0.0018 & 0.0027 & 0.0081 & 0.1583 & 0.5140 & 0.3152 \\ 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 1.0000 \end{pmatrix}.$$

The eight rows and columns represent, in order, the credit ratings AAA, AA, A, BBB, BB, B, CCC/C and Default. As already announced, P does not satisfy the Lévy Dichotomy so an exact generator does not exist. We observe that the eigenvalues of P are distinct and positive, so proposition 2.3.3 (i) would have ruled out all branches of the complex logarithm but the principal one for our search anyway.

Our heuristic conclusion is to simply compute $\mathrm{Log}(P)$ and modify the result to become a Q-matrix using (3.2). During this process, we come across four negative off-diagonal entries in the matrix $\widetilde{Q}$, which underpins that an exact generator does not exist. We note that none of these entries has an absolute value greater than $1.5 \cdot 10^{-4}$. The modified matrix Q_{IRW} looks as follows:

$$Q_{IRW} \approx \begin{pmatrix} -0.1051 & 0.1020 & 0.0012 & 0.0002 & 0.0009 & 0.0002 & 0.0007 & 0.0000 \\ 0.0064 & -0.1061 & 0.0948 & 0.0034 & 0.0004 & 0.0007 & 0.0003 & 0.0001 \\ 0.0003 & 0.0216 & -0.0897 & 0.0629 & 0.0026 & 0.0013 & 0.0002 & 0.0006 \\ 0.0001 & 0.0009 & 0.0420 & -0.0976 & 0.0463 & 0.0049 & 0.0016 & 0.0018 \\ 0.0002 & 0.0004 & 0.0004 & 0.0658 & -0.1733 & 0.0928 & 0.0076 & 0.0061 \\ 0.0000 & 0.0004 & 0.0013 & 0.0003 & 0.0740 & -0.1895 & 0.0765 & 0.0371 \\ 0.0000 & 0.0000 & 0.0023 & 0.0035 & 0.0025 & 0.2404 & -0.6765 & 0.4279 \\ 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 \end{pmatrix} .$$

Let us measure the precision of our approximation by the average absolute error in each entry. It makes sense to consider this characteristic, because it is easy to interpret and invariant to a possible change in the size of the encountered matrices. Thus, let $\|A\|$ denote the sum of the absolute values of all entries of A divided by the square of the dimension of A. For the given matrix Q_{IRW} we compute

$$\|P - \exp(Q_{IRW})\| \approx 6.60 \cdot 10^{-6} \ ,$$

so this approximation seems pretty decent.

Let us now use the method of algorithm 3.1.4 suggested by Jarrow et al. and compare the results: based on the assumption that there is rarely ever more than one transition in a year we compute a different approximate generator Q_{JLT}. When rounded to four decimal places, the result reads

$$Q_{JLT} \approx \begin{pmatrix} -0.1049 & 0.0967 & 0.0058 & 0.0005 & 0.0009 & 0.0003 & 0.0005 & 0.0000 \\ 0.0060 & -0.1047 & 0.0907 & 0.0061 & 0.0007 & 0.0008 & 0.0002 & 0.0002 \\ 0.0003 & 0.0205 & -0.0873 & 0.0601 & 0.0038 & 0.0015 & 0.0002 & 0.0008 \\ 0.0001 & 0.0013 & 0.0402 & -0.0948 & 0.0427 & 0.0066 & 0.0015 & 0.0023 \\ 0.0002 & 0.0005 & 0.0018 & 0.0626 & -0.1683 & 0.0853 & 0.0083 & 0.0096 \\ 0.0000 & 0.0004 & 0.0014 & 0.0027 & 0.0678 & -0.1782 & 0.0551 & 0.0508 \\ 0.0000 & 0.0000 & 0.0024 & 0.0037 & 0.0111 & 0.2167 & -0.6656 & 0.4317 \\ 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 \end{pmatrix} .$$

We observe that this approximation is substantially less accurate, as

$$\|P - \exp(Q_{JLT})\| \approx 1.80 \cdot 10^{-3} \ .$$

The error committed when using the method of Israel et al. was about 270 times smaller. Hence, this example suggests that the first approach is superior.

Example 3.2.2 (Standard & Poor's 2014)**.** Let us now examine a transition matrix of the same credit rating agency based on data of only the year 2014 [SnP2]. After distributing the "Not rated" weights to the other entries as above, the matrix now looks much more sparse than in the previous case:

$$P \approx \begin{pmatrix} 0.7895 & 0.2105 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 \\ 0.0000 & 0.9782 & 0.0218 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 \\ 0.0000 & 0.0087 & 0.9626 & 0.0263 & 0.0024 & 0.0000 & 0.0000 & 0.0000 \\ 0.0000 & 0.0000 & 0.0305 & 0.9465 & 0.0231 & 0.0000 & 0.0000 & 0.0000 \\ 0.0000 & 0.0000 & 0.0000 & 0.0423 & 0.9125 & 0.0442 & 0.0010 & 0.0000 \\ 0.0000 & 0.0000 & 0.0000 & 0.0007 & 0.0455 & 0.9128 & 0.0322 & 0.0088 \\ 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0949 & 0.6789 & 0.2263 \\ 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 1.0000 \end{pmatrix} .$$

Like in the first example, P fails the test of the Lévy Dichotomy and all eigenvalues of P are real and positive, so again we build on $\mathrm{Log}(P)$ to come up with an approximate generator. This time, however, $\widetilde{Q}$ has 15 negative off-diagonal entries and the smallest of them has an absolute value of about $3.2 \cdot 10^{-3}$. Based on this observation, we formulate the hypothesis that a greater number of zeros in off-diagonal entries of a transition matrix P apart from the last row increases the multitude and absolute value of negative off-diagonal entries in the corresponding logarithm $\widetilde{Q}$. After modifying $\widetilde{Q}$ with (3.2), we obtain the approximate generator

$$Q_{IRW} \approx \begin{pmatrix} -0.2350 & 0.2405 & 0.0000 & 0.0001 & 0.0000 & 0.0000 & 0.0000 & 0.0000 \\ 0.0000 & -0.0220 & 0.0227 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 \\ 0.0000 & 0.0090 & -0.0386 & 0.0275 & 0.0022 & 0.0000 & 0.0000 & 0.0000 \\ 0.0000 & 0.0000 & 0.0321 & -0.0557 & 0.0250 & 0.0000 & 0.0000 & 0.0000 \\ 0.0000 & 0.0000 & 0.0000 & 0.0458 & -0.0928 & 0.0487 & 0.0002 & 0.0000 \\ 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0501 & -0.0945 & 0.0409 & 0.0043 \\ 0.0000 & 0.0000 & 0.0000 & 0.0001 & 0.0000 & 0.1208 & -0.3884 & 0.2738 \\ 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 \end{pmatrix} .$$

This time, the precision of Q_{IRW} is distinctly lower:

$$\|P - \exp(Q_{IRW})\| \approx 2.39 \cdot 10^{-4} .$$

The approximate generator from the method of Jarrow et al. computes to

$$Q_{JLT} \approx \begin{pmatrix} -0.2364 & 0.2364 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 \\ 0.0000 & -0.0221 & 0.0221 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 \\ 0.0000 & 0.0089 & -0.0381 & 0.0268 & 0.0024 & 0.0000 & 0.0000 & 0.0000 \\ 0.0000 & 0.0000 & 0.0313 & -0.0550 & 0.0237 & 0.0000 & 0.0000 & 0.0000 \\ 0.0000 & 0.0000 & 0.0000 & 0.0443 & -0.0915 & 0.0462 & 0.0010 & 0.0000 \\ 0.0000 & 0.0000 & 0.0000 & 0.0007 & 0.0476 & -0.0913 & 0.0337 & 0.0092 \\ 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.1144 & -0.3873 & 0.2729 \\ 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 \end{pmatrix} .$$

The precision of this approximation has increased, as

$$\|P - \exp(Q_{JLT})\| \approx 6.25 \cdot 10^{-4} .$$

We conclude that the method of Israel et al. still leads to slightly more accurate results in this case, but the errors of the two approximations show contrary tendencies and the ratio of them has reduced drastically to about 2.6.

Example 3.2.3 (Moody's 1970-2007)**.** Next, we change to a different credit rating agency, but go back to the time horizon of the first example. We consider the average annual transition matrix from "Moody's" special comment on corporate rating transition rates in the years $1970-2007$ [Moo1]. Again, we re-assign the "Withdrawn rating" share to the other entries and obtain the following matrix:

$$P \approx \begin{pmatrix} 0.9163 & 0.0770 & 0.0066 & 0.0000 & 0.0002 & 0.0000 & 0.0000 & 0.0000 \\ 0.0113 & 0.9132 & 0.0721 & 0.0027 & 0.0006 & 0.0002 & 0.0000 & 0.0001 \\ 0.0007 & 0.0284 & 0.9130 & 0.0514 & 0.0051 & 0.0009 & 0.0002 & 0.0002 \\ 0.0005 & 0.0020 & 0.0515 & 0.8883 & 0.0454 & 0.0082 & 0.0024 & 0.0018 \\ 0.0001 & 0.0006 & 0.0042 & 0.0625 & 0.8294 & 0.0848 & 0.0063 & 0.0120 \\ 0.0001 & 0.0005 & 0.0018 & 0.0039 & 0.0621 & 0.8193 & 0.0624 & 0.0500 \\ 0.0000 & 0.0003 & 0.0003 & 0.0019 & 0.0073 & 0.1122 & 0.6856 & 0.1923 \\ 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 1.0000 \end{pmatrix} .$$

This time, the columns stand for the ratings Aaa, Aa, A Baa, Ba, B, Caa-C and Default. Once more, P does not satisfy the Lévy Dichotomy and there is a similar number of zeros accounting for this as in the first example. Also, P has real and positive eigenvalues again. We compute that $\mathrm{Log}(P)$ has five negative off-diagonal entries; this time their absolute value does not exceed $2.1 \cdot 10^{-4}$. This supports our hypothesis from the previous case. We modify $\widetilde{Q}$ using (3.2) and receive the approximate generator

$$Q_{IRW} \approx \begin{pmatrix} -0.0879 & 0.0842 & 0.0039 & 0.0000 & 0.0001 & 0.0000 & 0.0000 & 0.0000 \\ 0.0123 & -0.0926 & 0.0790 & 0.0007 & 0.0004 & 0.0001 & 0.0000 & 0.0001 \\ 0.0005 & 0.0311 & -0.0938 & 0.0570 & 0.0044 & 0.0006 & 0.0002 & 0.0001 \\ 0.0005 & 0.0013 & 0.0571 & -0.1220 & 0.0526 & 0.0067 & 0.0027 & 0.0011 \\ 0.0001 & 0.0006 & 0.0026 & 0.0728 & -0.1929 & 0.1028 & 0.0039 & 0.0102 \\ 0.0001 & 0.0005 & 0.0018 & 0.0017 & 0.0754 & -0.2091 & 0.0833 & 0.0462 \\ 0.0000 & 0.0004 & 0.0002 & 0.0021 & 0.0038 & 0.1501 & -0.3840 & 0.2276 \\ 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 \end{pmatrix} .$$

The error of approximation is very similar to the first example:

$$\|P - \exp(Q_{IRW})\| \approx 7.45 \cdot 10^{-6} \ .$$

Using the method of Jarrow et al., we get the matrix

$$Q_{JLT} \approx \begin{pmatrix} -0.0875 & 0.0804 & 0.0069 & 0.0000 & 0.0002 & 0.0000 & 0.0000 & 0.0000 \\ 0.0118 & -0.0908 & 0.0754 & 0.0028 & 0.0006 & 0.0002 & 0.0000 & 0.0001 \\ 0.0007 & 0.0297 & -0.0910 & 0.0537 & 0.0054 & 0.0010 & 0.0002 & 0.0002 \\ 0.0005 & 0.0022 & 0.0546 & -0.1185 & 0.0481 & 0.0086 & 0.0026 & 0.0019 \\ 0.0001 & 0.0007 & 0.0046 & 0.0686 & -0.1870 & 0.0930 & 0.0069 & 0.0132 \\ 0.0001 & 0.0005 & 0.0019 & 0.0043 & 0.0685 & -0.1993 & 0.0688 & 0.0552 \\ 0.0000 & 0.0004 & 0.0004 & 0.0023 & 0.0087 & 0.1348 & -0.3774 & 0.2308 \\ 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 \end{pmatrix} .$$

Again, the precision is similar to the first example:

$$\|P - \exp(Q_{JLT})\| \approx 1.70 \cdot 10^{-3} .$$

The same applies to the quotient of the errors, which calculates as approximately 222. From these observations we suspect that changing the credit rating agency while keeping the other factors unaltered does not influence the outcome of our search for a generator notably.

For the next two examples, I no longer present the complete transition matrix and approximate generators, but only the interesting results and their consequences:

Example 3.2.4 (Moody's 1920-2010)**.** We wish to test the hypotheses about the influence of the length of the observation period and the off-diagonal zeros in a transition matrix from the first two examples. Since choosing a different credit rating agency does not seem to disturb the results of our approximations, we can do so by studying the average one-year transition matrix from "Moody's" special comment on corporate default and recovery rates for the years $1920 - 2010$ [Moo2]. After making the usual adjustments, we obtain a 9×9 transition matrix P with four "problematic" off-diagonal zeros and an approximate generator $\widetilde{Q}$ with five negative off-diagonal entries, bounded by $4.1 \cdot 10^{-5}$. This supports the second assumption mentioned above. The approximation error of the Israel method decreases to $\|P - \exp(Q_{IRW})\| \approx 1.99 \cdot 10^{-6}$, while the one of the Jarrow algorithm increases to $\|P - \exp(Q_{JLT})\| \approx 1.90 \cdot 10^{-3}$. Hence, their ratio amplifies to about 938, which underpins the first hypothesis.

Example 3.2.5 (Fitch 1994-2014)**.** In this last example, we consider the average annual transition matrix for the years $1994 - 2014$ from the 2014 global corporate finance transition and default study by the credit rating agency "Fitch" [Fit]. Here, the ratings are finely subdivided into 18 categories, giving a much larger transition matrix than in all the previous examples. Additionally, for the first time, eigenvalues in $\mathbb{C} \setminus \mathbb{R}$ occur and $\pi < -\ln(\det(P)) \approx 7.80 < 3\pi$, so hypothetically we can no longer rely on the principal branch Log of the complex logarithm for our search only. Nevertheless, the results of our investigation are not very different from the previous examples. We consider both adding and subtracting $2\pi i$ in the logarithms of the six eigenvalues in $\mathbb{C} \setminus \mathbb{R}$, in all 3^6 possible constellations. Out of all approximate generators arising from these combinations, $\operatorname{Log}(P)$ is still the most precise one. Due to the extensive computation time needed, I was not able to test all 3^{18} variations of allowing a modification in the logarithms of every eigenvalue, also the real ones. I can only suspect that this would not have changed the outcome. The approximation errors amount to $\|P - \exp(Q_{IRW})\| \approx 9.83 \cdot 10^{-5}$ and $\|P - \exp(Q_{JLT})\| \approx 3.30 \cdot 10^{-3}$ and their ratio is about 34, which matches our hypothesis about the influence of the observation period length.

Conclusion 3.2.6. Since the five examples in this section all contribute to a considerably harmonic overall picture, I dare to make some tentative conclusions: in practice, it seems to work best to simply compute $\operatorname{Log}(P)$ for a transition matrix and then modify the result to become a Q-matrix. This can be realized efficiently using interpolation polynomials as suggested by Israel. Applying the logarithm directly to the diagonal matrix in a *Jordan decomposition* takes much longer run time and is said to be numerically unstable. The negative off-diagonal entries in $\widetilde{Q}$ always stayed comparatively small and I would regard the average approximation error in each entry as tolerable. We could observe the tendency that the two methods seem to react contrarily to an increase in the length of the observation period. This could be due to Jarrow's assumption that there rarely ever is more than one transition per year, which perhaps is not applicable to the real credit rating process and causes even stronger discrepancies in the long run. Furthermore, the minimum approximation error was larger for more sparse matrices, which could be interpreted as meaning that the represented processes are less typically Markovian in the sense of the Lévy dichotomy. I also remark that I consider the idea of modifying a transition matrix beforehand to avoid this problem mentioned in [IRW, p. 20] unfavorable. One would have to make arbitrary judgments concerning which value the zeros on the off-diagonal should be raised to. I have tested that making only slight modifications in these entries (of the order 10^{-4}), still did not make any of the transition matrices embeddable (recall that the Lévy dichotomy is only a necessary condition for the existence of a generator). But altering the values of the problematic entries any greater, would be all the more discretionary. Lastly, I comment that it might be possible to find a rule for reallocating the "Not rated" weights which does more justice to the real credit rating process than Jarrow's guideline, if one analyzed, what are typically the reasons for such a transition to happen.

Bibliography

[Bals] Balser, W. (2010): *Vorlesungsmanuskript zur Funktionentheorie*; retrieved September 28, 2015 from
http://www.mathematik.uni-ulm.de/m5/balser/Skripten/Funktionentheorie-Master.pdf

[Berg] Berg, C. (2012): *Complex Analysis*, p. 40f.; retrieved September 28, 2015 from
http://www.math.ku.dk/noter/filer/koman-12.pdf

[Born] Bornemann, F. (2013): *Funktionentheorie*, Springer Basel

[Chu] Chung, K. L. (1967): *Markov chains with stationary transition probabilities*, Springer New York

[Cuth] Cuthbert, J. R. (1972): *On uniqueness of the logarithm for Markov semi-groups.* J. London Mathematical Society 4, p. 629

[deBo] de Boor, C. (2013): *Linear Algebra*, p. 171f.; retrieved September 28, 2015 from
http://pages.cs.wisc.edu/~deboor/443/book.pdf

[Down] Down, A. (2005): *Winding Number*; retrived September 28, 2015 from
http://people.duke.edu/~ad159/files/m185/11.pdf

[Elf] Elfving, G. (1937): *Zur Theorie der Markoffschen Ketten*, Acta Social Science Fennicae n., series A.2, no. 8, 1-17

[Fit] Fitch Ratings (2015): *Global Corporate Finance 2014 Transition and Default Study*, p. 23; retrieved September 28, 2015 from
https://www.fitchratings.com/web_content/nrsro/nav/NRSRO_Exhibit-1.pdf

[Garr] Garrett, P. (2014): *Examples of operators and spectra*, p. 1f.; retrieved September 28, 2015 from
http://www.math.umn.edu/~garrett/m/fun/notes_2012-13/06b_examples_spectra.pdf

[Good] Goodman, G. S. (1970): *An intrinsic time for non-stationary finite Markov chains*, Zeitschrift für Wahrscheinlichkeitstheorie 16, pp. 165-180

[IRW] Israel, R. B.; Rosenthal, J. S.; Wei, J. Z. (2000): *Finding Generators for Markov Chains via Empirical Transition Matrices, with Applications to Credit Ratings*

[JLT] Jarrow, R. A.; Lando, D.; Turnbull, S. M. (1997): *A Markov model for the term structure of credit risk spreads*, Review of Financial Studies 10, pp. 481-523. ; retrieved September 28, 2015 from http://forum.johnson.cornell.edu/faculty/jarrow/055%20Markov%20Model%20Credit%20Risk%20RFS%201997.pdf

[King] Kingman, J. F. C. (1962): *The imbedding problem for finite Markov chains*, Zeitschrift für Wahrscheinlichkeitstheorie 1

[Moo1] Moody's Investor Services (2008): *Corporate One-to-Five-Year Rating Transition Rates*, p. 2; retrieved September 28, 2015 from https://www.moodys.com/sites/products/DefaultResearch/2007100000486361.pdf

[Moo2] Moody's Inverstor Services (2011): *Corporate Default and Recovery Rates, 1920-2010*, p. 23; retrieved September 28, 2015 from http://efinance.org.cn/cn/FEben/Corporate%20Default%20and%20Recovery%20Rates,1920-2010.pdf

[Neun] Neuner, C. (2010): *Der Riesz-Dunford'sche Funktionalkalkül*; retrieved September 28, 2015 from http://www.asc.tuwien.ac.at/~funkana/downloads_general/sem_neuner.pdf

[Norr] Norris, J. R. (1998): *Markov chains*, Cambridge University Press, pp. 1-105

[Rud] Rudin, W. (1991): *Functional Analysis*, McGraw-Hill Singapore, second edition, p. 253ff.

[Runn] Runnenberg, J. Th. (1962): *On Elfving's problem of imbedding a time-discrete Markov chain in a continuous time one for finitely many states*, Proceedings, Koninklijke Nederlandse Akademie van Wetenschappen 28 series A, Mathematical Sciences 65 (5), pp. 536-541

[Schn] Schnaubelt, R. (2012): *Lecture notes spectral theory*, p. 77ff.; retrieved September 28, 2015 from http://www.math.kit.edu/iana3/~schnaubelt/media/st-skript.pdf

[SiSp] Singer, B.; Spilerman, S. (1976): *The Representation of Social Processes by Markov Models*; retrieved September 28, 2015 from http://www.stat.cmu.edu/~fienberg/Stat36-835/SingerSpilerman-AJS-1976.pdf

[SnP1] Standard & Poor's Financial Services (2014): *Default, Transition, and Recovery: 2013 Annual Global Corporate Default Study And Rating Transitions*, p. 52; retrieved September 28, 2015 from http://www.maalot.co.il/publications/fts20140324161422.pdf

[SnP2] Standard & Poor's Financial Services (2015): *Default, Transition, and Recovery: 2014 Annual Global Corporate Default Study And Rating Transitions*, p. 51; retrieved September 28, 2015 from http://www.nact.org/resources/2014_SP_Global_Corporate_Default_Study.pdf

[Swan] Swanson, H. (2015): *Why isn't every element of the spectrum an eigenvalue? (Where is the error in my proof?)*; retrieved September 28, 2015 from
http://math.stackexchange.com/questions/1373884/why-isnt-every-element-of-the-\
spectrum-an-eigenvalue-where-is-the-error-in-my

[Tayl] Taylor, M. (year unknown): *Banach algebras - Bare Bones Basics*, p. 24; retrieved September 28, 2015 from
http://www.unc.edu/math/Faculty/met/banalg.pdf

[Zhao] Zhao, M. (2011): *The Gronwall's Inequality*, p. 1; retrieved September 28, 2015 from
http://www.math.ubc.ca/~mingfeng/Talks%20and%20Notes/The%20Gronwall's%
20inequality.pdf

List of Figures

Declaration

I hereby declare that I have written this bachelor thesis completely on my own and that I have not used any other sources than the ones listed in the bibliography to do so. This bachelor thesis was not used in the same or a similar version to achieve an academic grading or is being published elsewhere.

Marburg, September 28, 2015

(Leon Roschig, author)