Universiteit Leiden

From discrete-time Markov chains to coalgebras: A general theory of limit points

Sam Testers (s3710688)

Supervisors:
Henning Basold & Moritz Otto & Oisín Flynn-Connolly

Bachelor Thesis– Wiskunde & Informatica

Leiden Institute of Advanced Computer Science (LIACS)
liacs.leidenuniv.nl June 30, 2026

Abstract

A coalgebra is a structure in the context of category theory, which can be used to model various observation-based dynamical systems. A few examples are automata, reactive systems, continuous-time Markov chains and discrete-time Markov chains. The high level of abstraction makes a coalgebra powerful: A theorem stated on coalgebras can be used on all systems which can be modeled as such. For discrete-time Markov chains, limit points are extensively researched. In the case of discrete-time Markov chains, these limit points are called stationary distributions. In this thesis, we translate one theorem for finding stationary distributions of discrete-time Markov chains to a theorem for finding limit points for some general coalgebra using the Banach fixed point theorem. To demonstrate that the theorem is applicable to other systems than discrete-time Markov chains, we apply the theorem to a continuous-time Markov chain.

Contents

1 Introduction

Many systems in mathematics and computer science evolve over time, possibly in a non-deterministic manner. A common assumption of such systems is that their future behavior can be calculated using the current state of the system. in other words, they are observation based. Therefore, a natural problem in the analysis of such systems is understanding their long-term behavior.

Discrete-time Markov chains are an example of such systems which has been extensively researched [13, 8]. The discrete-time Markov chain consists of a state space and a transition function. The transition function can take a distribution over the state space and evolve it for one time step. For discrete-time Markov chains, long-term behavior is generally studied by finding a stationary distribution: A distribution over the state space that remains invariant under the transition function. Under suitable conditions one can prove that the transition function is a contraction under the total variation metric. Key to this theorem is that the total variation metric is a complete metric on the space of distributions over the state space. Therefore, by the Banach Fixed point theorem [11, Corollary 5.1.3], repeatedly applying the transition function on an arbitrary distribution over the state space will converge to a unique stationary distribution [8, Theorem 4.29]. This stationary distribution can be seen as the limit point of the discrete-time Markov chain.

Many systems share a similar observation-based structure as discrete-time Markov chains. Coalgebras provide a categorical framework for describing such systems. A coalgebra consists of a carrier object, together with a morphism which describes how each state of the system evolves [2]. By choosing appropriate morphisms, different types of systems can be represented as coalgebras. A few examples are the Markov decision processes, reactive systems, and automata [15], as well as ordinary differential equations [3].

The main advantage of the abstraction of coalgebras is that properties proven for coalgebras can be applied to all systems that can be modeled as such. With this motivation, we created the following research question: How can [8, Theorem 4.29] for finding stationary distributions of discrete-time Markov chains be generalized to find limit points in a coalgebraic setting?

To achieve this goal, we start by modeling the discrete-time Markov chain as a coalgebra, using the coalgebraic model described from [15] together with other concepts of category theory from [9, 2]. We rewrite the proof of [8, Theorem 4.29] using the language of coalgebra and the coalgebraic model of the Markov chain. By analyzing this proof we isolate assumptions required in order for the theorem to hold. These assumptions become the conditions under which the general limit point theorem for coalgebras hold. Since we use the Banach fixed point theorem, these conditions can be divided into those required to define a complete metric and those required to prove that the transition function is a contraction under this metric.

Lastly, we create a coalgebraic model of a continuous-time Markov chain using a composition of the evolution functor [10, Definition 2.12], together with the distribution functor [15]. By applying the general limit point theorem on this coalgebraic model of the continues-time Markov chain, we demonstrate that the theorem extends to coalgebras other than the discrete-time Markov chain.

1.1 Thesis overview

In Section 2 we will introduce the necessary preliminaries of the discrete-time Markov chain, including the limit point theorem we will translate. In Subsection 3.1, we will introduce the preliminary definitions of category theory. In Section 4, we generalize the theorem for finding the limit point of the discrete-time Markov chain to other coalgebras. In Subsection 4.1, we will model a discrete-time Markov chain as a coalgebra, and translate the theorem and proof for finding its stationary distribution to the coalgebraic setting. In Subsection 4.2, we analyze the statement and the proof of the coalgebraic theorem for the discrete-time Markov chain, and determine which restrictions we need on our coalgebra and underlying category and functor in order to apply the theorem for finding limit points in for general coalgebras. In Subsection 4.3, we apply the limit point theorem for coalgebras to the continuous-time Markov chain. Lastly, in Section 5, we reflect on the results and supply possible exploits for future research.

1.2 Related work

The stationary distributions of discrete-time Markov chains have been extensively researched. For finite state space, a common theorem for finding the stationary distribution is Theorem 4.1 of [14] which states that the stationary distribution is an eigenvalue of the transition matrix P. For non-finite state space, a useful theorem is Theorem 9.8 of [13], which states that if the Markov chain is irreducible, aperiodic and positive recurrent, then $(\nu P^n)_{n\in\mathbb{N}}$ converges to a unique stationary distribution π. Both theorems heavily rely on structural properties of the discrete-time Markov chain, which would prove to be difficult to verify for general coalgebras. Therefore, we opted for Theorem 4.29 from [8], which requires a transition function on which we can apply the Banach fixed point theorem.

The Banach fixed point theorem has not been studied extensively within category theory. This is partially due to the fact that the collection of metric spaces with contractive maps as morphisms does not form a category due to the non-existence of the identity morphisms. One exception is [1] where the Banach fixed point is used on an endofunctor on the category of data types enriched over complete metric spaces. Another exception is [5] where the Banach fixed point is generalized to categories enriched over a quantale.

In Subsection 4.3, we apply the general fixed-point theorem to the continuous-time Markov chain. In [12], a continuous-time Markov chain is modeled via a graded monad coalgebra, where a monoid represents the passage of time and the monad grading captures the systems behavior at different time scales.

2 Discrete-time Markov chain

In this section we will introduce some necessary background on discrete-time Markov chains. The content of this section is based on the book *Markov Chains* by James Norris [13]. We will start by giving the definition of a discrete-time Markov chain. Afterwards we will discuss limit behavior. Although we provide the necessarily background in probability theory, a basic understanding is recommended. A more comprehensive introduction to probability theory can be found in [7].

2.1 Definitions

In this section we will only consider discrete probability distributions.

Let $(\Omega, \mathcal{F}, \mathbb{P})$ and let Y be a set. A random variable on Y is a mapping $X : \Omega \to Y$ such that the probability of X taking on a certain element $y \in Y$ can be measured using our probability function $\mathbb{P} : \mathcal{F} \to [0, 1]$:

$$\mathbb{P}(X = y) = \mathbb{P}(\{\omega \in \Omega \mid X(\omega) = y\}).$$

We can simplify this equation by coupling X to a probability measure $\nu : Y \to \mathbb{R}_{\geq 0}$ such that $\nu(y) = \mathbb{P}(X = y)$. In this case, we write $X \sim \nu$. Since probabilities can not be negative, we require that $\nu(y) \geq 0$ for all $y \in Y$. Moreover, we require $\int\limits_{y \in Y} \nu(y) dy = 1$, since X always takes on a value in Y.

We define the support of a distribution ν as follows:

$$\text{supp}(\nu) = \{y \in Y \mid \nu(y) > 0\}.$$

When this support is countable, we can define a sum over this support: Consider injection $\phi : \text{supp}(\nu) \to \mathbb{N}$. By restricting the codomain to the image of ϕ, we get an isomorphism $\phi : \text{supp}(\nu) \to \text{im}(\nu)$, with inverse $\psi : \text{im}(\nu) \to \text{supp}(\nu)$. From this, we can define a sequence $(a_n)_{n \in \mathbb{N}}$ as

$$a_n = \begin{cases} \nu(\psi(n)) & \text{if } n \in im(\nu) \\ 0 & \text{otherwise} \end{cases}.$$

By defining $\sum\limits_{x \in \text{supp}(\nu)} \nu(x) := \sum\limits_{n=0}^{\infty} a_n$, we get a well defined sum over the countable support of ν. Using this sum, we can rewrite condition $\int\limits_{y \in Y} \nu(y) dy = 1$ to $\sum\limits_{y \in \text{supp}(\nu)} \nu(y) = 1$.

Definition 2.1. (stochastic process). Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and let Y be a set. A sequence of random variables $\mathcal{X} = (X_n)_{n \in \mathbb{N}}$, $X_n : \Omega \to Y$ is called a discrete stochastic process on Y. Let $(\nu_n)_{n \in \mathbb{N}}$ be the sequence of probability distributions on Y with $X_n \sim \nu_n$ for all $n \in \mathbb{N}$. we define the support of $\mathcal{X}$ as

$$\text{supp}(\mathcal{X}) = \bigcup_{n \in \mathbb{N}} \text{supp}(\nu_n).$$

Note that, if $\text{supp}(\nu_n)$ is countable for all $n \in \mathbb{N}$, the support of $\mathcal{X}$ is a countable union of countable sets. Therefore $\text{supp}(\mathcal{X})$ is countable.

Definition 2.2. (Discrete-time Markov chain). Let $\mathcal{X} = (X_n)_{n\in\mathbb{N}}$ be a stochastic process on S with a countable support. We call $(X_n)_{n\in\mathbb{N}}$ a discrete-time Markov chain if it obeys the following axiom:

$$\mathbb{P}(X_{n+1} = j | X_n = i, X_{n-1} = x_{n-1}, ..., X_0 = x_0) = \mathbb{P}(X_{n+1} = j | X_n = i)$$

$\forall i, j \in S$, $\forall n \in \mathbb{N}$ and $\forall x_{n-1}, ..., x_0 \in S$ with $\mathbb{P}(X_{n-1} = x_{n-1}, ..., X_0 = x_0) \neq 0$. This property is referred as the Markov property. Additionally, the Markov chain is called time-homogeneous if it obeys the following additional axiom

$$\mathbb{P}(X_{n+1} = j \mid X_n = i) = \mathbb{P}(X_1 = j \mid X_0 = i).$$

$\forall i, j \in S$ and $\forall n \in \mathbb{N}$. In this case, we write p_{ij} for $\mathbb{P}(X_1 = j \mid X_0 = i)$.

In this thesis, we only consider time homogeneous Markov chains. Since time-homogeneous Markov chains have the property that the transition probabilities do not depend on time or the previously visited states, we can intuitively model them as a graph, as we see in the next example.

Example 2.3. (Discrete-time Markov chain). Consider the discrete-time Markov chain $(X_n)_{n\in\mathbb{N}}$ with the following transitions:

$$\begin{aligned} p_{11} &= 0 & p_{12} &= \frac{1}{3} & p_{13} &= \frac{2}{3} \\ p_{21} &= \frac{1}{2} & p_{22} &= 0 & p_{23} &= \frac{1}{2} \\ p_{31} &= 0 & p_{32} &= 1 & p_{33} &= 0. \end{aligned}$$

This Markov chain can be modeled using a graph with nodes in S and edges labeled with the transition probabilities (see figure 1).

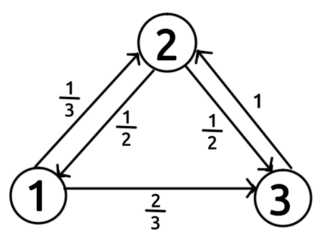

Figure 1: Example of a simple discrete-time Markov chain with $S = \{1, 2, 3\}$.

Using transitions p_{ij} and the current state of a discrete-time Markov chain, we can determine the probabilities for the Markov chain to transition to each possible state after one time step. Generally, we do not know the current state of the Markov chain. Instead, we know a distribution ν on S, describing the chances for the Markov chain to be in a certain state. In this case, We can advance the Markov chain by one time step by applying the transition function T, which is stated below. By multiplying the probabilities of ν with the transition probabilities, ν is transforms into the distribution on S which describes the chances for the Markov chain to be in a certain state after one time step:

$$T(\nu)(j) = \sum_{i\in S^*} \nu(i) \cdot p_{ij} \quad \text{for all } j \in S.$$

Since T transforms a distribution on S into another distribution on S, we can iteratively apply T. This way, we can calculate where the Markov chain will be after an arbitrary amount of time steps.

2.2 Stationary Distribution

In this section we will define a stationary distribution of a discrete-time Markov chain. Afterwards we will state the theorem for finding stationary distribution of discrete-time Markov chains which we will be generalizing to other coalgebras in Section 4.

Definition 2.4. (Stationary distribution). Let $\mathcal{X} = (X_n)_{n\in\mathbb{N}}$ be a discrete-time Markov chain with transition function T. A distribution π on S is called a stationary distribution of $\mathcal{X}$ if

$$T(\pi) = \pi.$$

In other words, π is a stationary distribution if it does not change when we take a time step. A natural question is, when we have a distribution ν over S, and we keep applying T, will it eventually converge to a stationary distribution. A useful theorem for finding stationary points of T is the Banach fixed point theorem.

Theorem 2.5. (Banach fixed point theorem). Consider a metric space (X, d) with $X \neq \emptyset$. Suppose that (X, d) is complete and let $T : X \to X$ be a contraction on X. Then there exists a unique fixed point x^* of T. Moreover, for any $x_0 \in X$, we have that $T^n(x_0)$ converges to x^*.

Proof. See 5.1.2 and 5.1.3 of [11]. □

In this article we will consider the limit point theorem stated below. In the proof of this theorem, it is shown that, under certain conditions, the transition function T is a contraction under the total variation distance d_{TV}, which is a complete metric on the set of distributions on S. Therefore, by applying the Banach fixed point theorem, we find a unique limit point of T by iteratively applying T to an initial distribution δ over S.

Theorem 2.6. (Theorem 4.29 of [8]) Assume there exists a scalar $\alpha \in (0, 1)$ and a distribution ν on S such that $p_{ij} \geq \alpha \cdot \nu(j)$ for all $i, j \in S$. Then $T^n(\delta)$ converges to a unique stationary distribution π for all distributions δ on S.

In Section 4, we will generalize this theorem to general coalgebras satisfying some conditions. In the proof we will work towards using the Banach Fixed point theorem.

3 Category Theory

Category theory is an abstract mathematical language which, instead of focusing on the internal details of a mathematical structures (objects), focuses on the relationships (morphisms) between the structures. This high level of abstraction makes it a powerful tool which allows us to model several mathematical structures under a single framework. A commonly used framework is the coalgebra.

Coalgebras offer an abstract framework for modeling observation-based dynamical systems. A few examples are the discrete-time Markov chains, Markov decision processes, reactive systems, and automata (see [15]). The strength of the coalgebra lies in its high level of generality: Since many systems can be modeled as a coalgebra, theorems that are stated in the coalgebraic framework can be applied to all systems that can be modeled as such. With this motivation, in

Section 4, we will translate theorems for finding fixed points for discrete-time Markov chains to this coalgebraic framwork. In order to do this, we will first need a basic understanding of category theory.

In Section 3.1, we will give some necessary definitions in category theory. We will be using [2] and [9] as our sources.

3.1 Categories, coalgebras and monads

Definition 3.1. (Category). A category $\mathcal{C}$ consists of a class of objects, denoted by $ob(\mathcal{C})$ and a class of morphisms. Between every pair of objects A, B we have a collection $\mathcal{C}(A, B)$ which are the morphisms $f : A \to B$ between A and B such that

- For every object A there exists an identity morphism $\mathrm{id}_A : A \to A$.
- For $f : A \to B$ and $g : B \to C$ there exists a unique composite morphism $g \circ f : A \to C$.

These satisfy the following axioms:

- For every morphism $f : A \to B$, we have $\mathrm{id}_B \circ f = f \circ \mathrm{id}_A = f$.
- For any three morphisms $f : A \to B$, $g : B \to C$, $h : C \to D$, we have $(h \circ g) \circ f = h \circ (g \circ f)$.

Definition 3.2. (Terminal object). Let $\mathcal{C}$ be a category. An object $1 \in \mathrm{ob}(\mathcal{C})$ is called a terminal object if for every $A \in \mathrm{ob}(\mathcal{C})$ there exists a unique morphism

$$!_A : A \to 1.$$

Definition 3.3. (Global element). Let $\mathcal{C}$ be a category with terminal object 1. A global element of an object $A \in \mathrm{ob}(\mathcal{C})$ is a morphism

$$x : 1 \to A.$$

Definition 3.4. (Generator). An object $G \in \mathrm{ob}(\mathcal{C})$ is a generator of $\mathcal{C}$ if for all $A, B \in \mathrm{ob}(\mathcal{C})$ and all morphisms $f, g \in \mathcal{C}(A, B)$ we have

$$(f \circ x = g \circ x \ \forall x \in \mathcal{C}(G, A)) \Rightarrow f = g.$$

Notation 3.5. Consider morphism $f : A \to B$ and global element $x : 1 \to A$. We define the application of f to x as $f \circ x : 1 \to B$. We denote this global element as $f(x)$.

A simple example of a category is **Set**, where the objects consists of sets and the morphisms are the functions between those sets. In the category **Set**, every singleton $\{x\} \in \mathrm{ob}(\mathbf{Set})$ is a terminal object, and every object $X \in \mathrm{ob}(\mathbf{Set})$ with $X \neq \emptyset$ is a generator.

3.1.1 Coalgebras

In order to define a coalgebra, we first must define a functor. Just like we have morphisms between objects, a functor is a morphism between categories. This functor must preserves structure between the objects and morphisms of the two categories. On a (endo)functor, we can define a coalgebra, with which we can model state-based systems.

Definition 3.6. (Functor). Consider $\mathcal{C}$ and $\mathcal{D}$ two categories. A functor $F : \mathcal{C} \to \mathcal{D}$ consists of the following two mappings:

- A mapping $F : \mathrm{ob}(\mathcal{C}) \to \mathrm{ob}(\mathcal{D})$.
- For every $A, B \in \mathrm{ob}(\mathcal{C})$, a mapping $F : \mathcal{C}(A, B) \to \mathcal{D}(F(A), F(B))$.

These mappings must obey the following axioms:

- For every $A \in \mathrm{ob}(\mathcal{C})$, we have $F(\mathrm{id}_A) = \mathrm{id}_{F(A)}$.
- For $f : A \to B$ and $g : B \to C$, we have
$$F(g \circ f) = F(g) \circ F(f).$$

Notation 3.7. To avoid the excessive use of parentheses, we sometimes write Ff and FA instead of $F(f)$ and $F(A)$.

Definition 3.8. (Endofunctor). A *functor* $F : \mathcal{C} \to \mathcal{C}$ with the same codomain as its domain is called an *endofunctor*.

A simple example of such an endofunctor is the identity functor $1_{\mathcal{C}} : \mathcal{C} \to \mathcal{C}$ which maps any object X and any morphism f to itself.

Definition 3.9. (Faithful functor). We call functor $F : \mathcal{C} \to \mathcal{D}$ *faithful* if, for all objects $A, B \in \mathrm{ob}(\mathcal{C})$, and all morphisms $f, g \in \mathcal{C}(A, B)$ with $f \neq g$, we have
$$F(f) \neq F(g).$$

Definition 3.10. (Concrete Category). We call category $\mathcal{C}$ concrete if there exists a faithful functor $U : \mathcal{C} \to \mathbf{Set}$. Functor U is often called the forgetful functor.

A simple example of a concrete category is **Set** itself, where the identity functor acts as the forgetful functor. Concrete categories prove to be fruitful in category theory since they allow categorical structures to be studied pointwise on the underlying set.

Definition 3.11. (Algebra). An F-*algebra*, or in short an *algebra*, consists of an object $X \in \mathcal{C}$ and a morphism $\alpha : FX \to X$. We call X the carrier, and α the algebra.

Definition 3.12. (Category of Algebras). Consider an endofunctor F on category $\mathcal{C}$. We call $\mathbf{Alg}(F)$ the category of algebras on F with all F-algebras $\alpha : FX \to X$, $X \in \mathrm{ob}(\mathcal{C})$ as objects and morphisms $f : X \to Y$ between algebra (α, X) with (β, Y) such that the diagram below commutes.

$$\begin{tikzcd} FX \arrow[r, "\alpha"] \arrow[d, "Ff"'] & X \arrow[d, "f"] \\ FY \arrow[r, "\beta"'] & Y \end{tikzcd}$$

Definition 3.13. (Coalgebra). An F-*coalgebra*, or in short a *coalgebra*, is the dual of an F-*algebra*. It consists of an object $X \in \mathcal{C}$ and a morphism $c : X \to FX$. We call X the carrier, and c the coalgebra structure.

Definition 3.14. (Category of Coalgebras). Consider an endofunctor F on category $\mathcal{C}$. We call **Coalg**(F) the category of coalgebras on F with all F-coalgebras $\alpha : X \to FX$, $X \in \mathrm{ob}(\mathcal{C})$ as objects and morphisms $f : X \to Y$ between coalgebra (α, X) with (β, Y) such that the following diagram below commutes.

$$\begin{tikzcd} X \arrow[r, "\alpha"] \arrow[d, "f"'] & FX \arrow[d, "Ff"] \\ Y \arrow[r, "\beta"'] & FY \end{tikzcd}$$

In order to prove that **Alg**(F) and **Coalg**(F) are categories, we need to verify the conditions in Definition 3.1. In appendix D and E respectively, we verify these conditions.

3.1.2 Natural Transformations and Monads

In category theory, natural transformations are the morphisms of functors. Given an endofunctor, a monad consists of two natural transformations: A *multiplication* and a *unit* transformation. These two transformations allow values to be lifted into the endofunctor, and nested structures to be flattened. By composing these transformations, we can move between iterated applications of the endofunctor, which allows us to chain operations. For example, in Subsection 3.2 we use the *multiplication* to define the 1-*step* transition probabilities for discrete-time Markov chains, which we can iterate.

Definition 3.15. (Natural Transformation). Let $\mathcal{C}$ and $\mathcal{D}$ be two categories, with two functors $F, H : \mathcal{C} \to \mathcal{D}$ between them. A *natural transformation* α from F to H consists of a collection of morphisms $\alpha_X : F(X) \to H(X)$ such that the following diagram commutes for all morphisms $f : X \to Y$ of $\mathcal{C}$.

$$\begin{tikzcd} FX \arrow[r, "\alpha_X"] \arrow[d, "Ff"'] & HX \arrow[d, "Hf"] \\ FY \arrow[r, "\alpha_Y"'] & HY \end{tikzcd}$$

In this case, we use double arrows and write $\alpha : F \Rightarrow H$ to distinguish it from morphisms between objects.

Definition 3.16. (monad) . A monad (T, μ, η) on category $\mathcal{C}$ consists of an endofunctor $T : \mathcal{C} \to \mathcal{C}$ and two natural transformations: A *unit* $\eta : \mathrm{id}_{\mathcal{C}} \Rightarrow T$, and *multiplication* $\mu : T^2 \Rightarrow T$. These are required to make the following diagrams commute for $X \in \mathcal{C}$.

$$\begin{tikzcd} TX \arrow[r, "T\eta_X"] \arrow[d, "\eta_{TX}"'] \arrow[dr, "\mathrm{id}_{TX}"] & T^2X \arrow[d, "\mu_X"] \\ T^2X \arrow[r, "\mu_X"'] & TX \end{tikzcd} \qquad \begin{tikzcd} T^3X \arrow[r, "\mu_{TX}"] \arrow[d, "T\mu_X"'] & T^2X \arrow[d, "\mu_X"] \\ T^2X \arrow[r, "\mu_X"'] & TX \end{tikzcd}$$

Definition 3.17. (comonad) . A comonad (T, δ, ϵ) on category $\mathcal{C}$ is the dual notion of a monad of T. It consists of an endofunctor $T : \mathcal{C} \to \mathcal{C}$ and two natural transformations: A *counit* $\epsilon : T \Rightarrow \mathrm{id}_{\mathcal{C}}$, and *comultiplication* $\mu : T \Rightarrow T^2$. These are required to make the following diagrams commute for $X \in \mathcal{C}$.

$$
\begin{tikzcd}
TX & T^2X \arrow[l, "T\epsilon_X"'] \\
T^2X \arrow[u, "\epsilon_T X"] & TX \arrow[l, "T\delta_X"] \arrow[u, "\delta_X"'] \arrow[ul, "\mathrm{id}_{TX}"']
\end{tikzcd}
\qquad
\begin{tikzcd}
T^3X & T^2X \arrow[l, "\delta_{TX}"'] \\
T^2X \arrow[u, "T\delta_X"] & TX \arrow[l, "\delta_X"] \arrow[u, "T\delta_X"']
\end{tikzcd}
$$

In Definition 3.11, we have defined an T-*algebra* and in Definition 3.16 we have introduced some additional structure on T in the form of a monad. An Eilenberg-Moore algebra is an algebra that preserves the monad structure.

Definition 3.18. (EM-Algebra). Consider the monad (T, μ, ν) on $\mathcal{C}$ and consider algebra $\alpha : TX \to X$. α is called an Eilenberg-Moore algebra of (T, μ, ν) if the following diagrams commute.

$$
\begin{tikzcd}
X \arrow[r, "\eta_X"] \arrow[dr, "\mathrm{id}_X"] & TX \arrow[d, "\alpha"] \\
& X
\end{tikzcd}
\qquad
\begin{tikzcd}
T^2X \arrow[r, "T\alpha"] \arrow[d, "\mu_X"'] & TX \arrow[d, "\alpha"] \\
TX \arrow[r, "\alpha"'] & X
\end{tikzcd}
$$

Similar to Definition 3.18, given a comonad (T, δ, ϵ), we can define an Eilenberg-Moore coalgebra, which respects the structure added by the comonad.

Definition 3.19. (EM-Coalgebra). Consider the comonad (T, δ, ϵ) on $\mathcal{C}$ and consider coalgebra $c : X \to TX$. c is called an Eilenberg-Moore coalgebra of (T, δ, ϵ) if the following diagrams commute.

$$
\begin{tikzcd}
X & TX \arrow[l, "\epsilon_X"'] \\
& X \arrow[u, "c"'] \arrow[ul, "\mathrm{id}_X"']
\end{tikzcd}
\qquad
\begin{tikzcd}
T^2X & TX \arrow[l, "Tc"'] \\
TX \arrow[u, "\delta_X"] & X \arrow[l, "c"] \arrow[u, "c"']
\end{tikzcd}
$$

When two monads are composed, the functors can generally not be interchanged without preserving the monad structure. A distributive law is a natural transformation needed in order to interchange composed monads, while respecting the monad structure [4].

Definition 3.20. (Distributive Law). Let (T, μ^T, ν^T) and (S, μ^S, ν^S) be monads on $\mathcal{C}$. We call $\lambda : TS \Rightarrow ST$ a distributive law if the following diagrams commute.

$$
\begin{tikzcd}
T(X) \arrow[d, "T(\eta^S_X)"'] \arrow[dr, "\eta^S_{TX}"] & \\
TS(X) \arrow[r, "\lambda_X"'] & ST(X)
\end{tikzcd}
\qquad
\begin{tikzcd}
TSS(X) \arrow[r, "\lambda_{SX}"] \arrow[d, "T(\mu^S_X)"'] & STS(X) \arrow[r, "S(\lambda_X)"] & SST(X) \arrow[d, "\mu^S_{T(X)}"] \\
TS(X) \arrow[rr, "\lambda_X"'] & & ST(X)
\end{tikzcd}
$$

$$
\begin{tikzcd}
S(X) \arrow[d, "\eta^T_{S(X)}"'] \arrow[dr, "S(\eta^T_X)"] & \\
TS(X) \arrow[r, "\lambda_X"'] & ST(X)
\end{tikzcd}
\qquad
\begin{tikzcd}
TTS(X) \arrow[r, "T\lambda_X"] \arrow[d, "\mu^T_{SX}"'] & TST(X) \arrow[r, "\lambda_{TX}"] & STT(X) \arrow[d, "S\mu^T_X"] \\
TS(X) \arrow[rr, "\lambda_X"'] & & ST(X)
\end{tikzcd}
$$

Given a distributive law $\lambda : TS \Rightarrow ST$ on two monads (T, μ^T, ν^T) and (S, μ^S, ν^S), proposition on page 97 of [4] states that we can equip ST with a monad $(ST, \mu_\lambda, \eta_\lambda)$ where $\mu_\lambda : (ST)^2 \Rightarrow ST$ and $\eta_\lambda : \mathrm{id}_\mathcal{C} \Rightarrow ST$ are formed by the following compositions:

$$STST(X) \xrightarrow{S\lambda_{TX}} SSTT(X) \xrightarrow{S^2(\mu^T_X)} SST(X) \xrightarrow{\mu^S_{T(X)}} ST(X)$$
$$X \xrightarrow{\eta^T_X} T(X) \xrightarrow{\eta^S_{T(X)}} ST(X) \ .$$

Lastly, we will define the Kleisli composition and extension. This composition will be used in Definition 3.27 to define an endomorphism which we can iterate to find the limit point.

Definition 3.21. (Kleisli composition). Consider a monad (T, μ, η) on $\mathcal{C}$, morphisms $f : A \to T(B)$ and $g : B \to T(C)$. We define the Kleisli composition $\#_T$ on (T, μ, η) of g and g as follows

$$g \#_T f := \mu \circ T(g) \circ f : A \to T(C).$$

Definition 3.22. (Kleisli-extension). Consider a monad (T, μ, η) on $\mathcal{C}$ and a morphisms $f : A \to T(B)$. We define the Kleisli-extension of (T, μ, η) on f as

$$f^{\#_T} := \mu \circ T(f) : T(S) \to T(S).$$

3.2 Discrete-time Markov chain as a Coalgebra

In this section we will model the discrete-time Markov chains described in Section 2 as a coalgebra. This coalgebra will send every state x to the transition distribution of state x. Therefore, we will start by define a functor that sends an element X to the discrete distributions on X. Afterwards, we will use this functor to create a coalgebra which models a discrete-time Markov chain. This construction can be found in [15], section 3. Afterwards, we will introduce a monad on the functor and a transition function and stationary point of a discrete-time Markov chain.

Definition 3.23. (Distribution Functor). The probability distribution functor $\mathbb{D} : \mathbf{Set} \to \mathbf{Set}$ maps an object X to the set of all probability distributions with countable support:

$$\mathbb{D}X = \left\{ \nu : X \to \mathbb{R}_{\geq 0} \mid \sum_{x \in \mathrm{supp}(\nu)} \nu(x) = 1 \right\}.$$

A morphism $f : X \to Y$ is mapped to $\mathbb{D}f : \mathbb{D}X \to \mathbb{D}Y$ such that for all $\nu \in \mathbb{D}X$ and all $y \in Y$ we have

$$\mathbb{D}f(\nu)(y) = \sum_{x \in \mathrm{supp}(\nu) : f(x) = y} \nu(x).$$

Since $\mathbb{D}$ is a functor, it is required to preserve identity and composition. In appendix A, we prove that $\mathbb{D}$ satisfies these required axioms.

In order to model a discrete-time Markov chain, we need to model the transitions p_{ij} for $i, j \in S$. Since we require the support of a Markov chain to be countable, we have that $\{j \in S : p_{ij} > 0\}$ is countable for all $i \in S$. Therefore we can model a discrete-time Markov chain as a coalgebra $c : S \to \mathbb{D}S$ which maps each state i to the transition distribution of i.

Definition 3.24. (Markov chain as a coalgebra). Take $\mathcal{X} = (X_n)_{n\in\mathbb{N}}$ a discrete-time Markov chain on state space S and transition probabilities p_{ij} for $i, j \in S$. We define the $\mathbb{D}$-coalgebra $c : S \to \mathbb{D}S$ which models this discrete-time Markov chain as

$$c(i)(j) = p_{ij}.$$

Example 3.25. (discrete-time Markov chain as a coalgebra) Consider the Markov chain of Example 2.3. This Markov chain can be modeled as a coalgebra $c : S \to \mathbb{D}S$ with

$$\begin{aligned} c(1) &= \{1 \mapsto 0, 2 \mapsto \tfrac{1}{3}, 3 \mapsto \tfrac{2}{3}\} \\ c(2) &= \{1 \mapsto \tfrac{1}{2}, 2 \mapsto 0, 3 \mapsto \tfrac{1}{2}\} \\ c(3) &= \{1 \mapsto 0, 2 \mapsto 1, 3 \mapsto 0\}. \end{aligned}$$

In this thesis, we want to examine limit behavior of discrete-time Markov chains. Therefore, we want to be able to repeatedly apply c to some initial distribution ν. In order to do so, we will define a transition function, which is equivalent to the transition function defined in Section 2 using coalgebra c. The tool that we will be using to define this transition function is the distribution monad. This monad is introduced in lemma 5.1.5 of [9].

Theorem 3.26. (Distribution Monad). Endofunctor $\mathbb{D}$ can be equipped with a monad structure $(\mathbb{D}, \eta, \mu)$, such that for all sets X and all $x, y \in X$ and $\lambda \in \mathbb{D}^2X$ we have

$$\eta_X(x)(y) = \delta_x(y) = \begin{cases} 1 & \text{if } y = x \\ 0 & \text{if } y \neq x \end{cases} \qquad \mu_X(\lambda)(x) \quad = \sum_{\gamma\in\mathrm{supp}(\lambda)} \lambda(\gamma)\cdot\gamma(x)$$

Proof. The proof consists of some routine calculations where we confirm that the diagrams in Definition 3.16 commute and can be found in appendix B. □

Definition 3.27. (Transition Function T_c). Consider the discrete-time Markov chain $c : S \to \mathbb{D}S$ and let $(\mathbb{D}, \eta, \mu)$ be the distribution monad described above. We define the transition function $T_c : \mathbb{D}S \to \mathbb{D}S$ of c as:

$$T_c := c^{\#_{\mathbb{D}}} = \mu_S \circ \mathbb{D}c.$$

Notation 3.28. In some cases we wish to work on morphisms to $\mathbb{D}S$ instead of $\mathbb{D}S$ itself. We define $\hat{T}_c^A : \mathbf{Set}(A, \mathbb{D}S) \to \mathbf{Set}(A, \mathbb{D}S)$ as $\hat{T}_c^A(f) = T_c \circ f$ for all $f \in \mathbf{Set}(A, \mathbb{D}S)$.

Proposition 3.29. (equivalence between T_c and T from Section 2). Let c be a Markov chain with state space S and transition function $T_c : \mathbb{D}S \to \mathbb{D}S$. Then the following equality holds for all $\nu \in \mathbb{D}S$ and all $j \in S$:

$$T_c(\nu)(j) = \sum_{i\in\mathrm{supp}(\nu)} \nu(i)\cdot c(i)(j).$$

Proof. Consider $\delta \in \mathbb{D}S$ and $j \in S$. By Definition 3.27, we have that the equality $T_c(\nu)(j) = \mu_S \circ \mathbb{D}c(\nu)(j) = \mu_S(\mathbb{D}c(\nu))(j)$ holds. Moreover, we have

$$\mathbb{D}c(\nu) = \left\{ \rho \mapsto \sum_{x\in\mathrm{supp}(\nu):c(x)=\rho} \nu(x) \right\}.$$

Therefore, by definition of μ, we have

$$T_c(\nu)(j) = \mu_S(\mathbb{D}c(\nu))(j) = \sum_{i \in \mathrm{supp}(\nu)} \nu(i) \cdot c(i)(j).$$

$\square$

Since this proposition proves that T_c models transition function T of section 2, we can use T_c as the 1-*step* transition function of c.

Definition 3.30. (stationary distribution) Consider the discrete-time Markov chain $c : S \to \mathbb{D}S$. We call $\pi : 1 \to \mathbb{D}S$ a stationary distribution of c if $\hat{T}_c^1(\pi) = \pi$.

We have defined and proven that T_c can be used to calculate the 1-*step* transition probabilities beginning in some initial distribution $\nu : 1 \to \mathbb{D}S$. Since, $\hat{T}_c$ is an endomorphism, we can iterate and thus calculate the n-*step* transition probabilities of c for arbitrary $n \in \mathbb{N}$. In this thesis we want to examine the limit behavior the iteration $(\hat{T}_c^1)^n(\nu)$. In particular, we want to determine whether $(\hat{T}_c^1)^n(\nu)$ converges to a stationary distribution.

4 Fixed point theorem for arbitrary coalgebra

This thesis aims to translate fixed point theorems for discrete-time Markov chains to arbitrary coalgebras. In this section, we will generalize Theorem 2.6 to a general coalgebra. In that theorem, we use the Banach Fixed point theorem to prove the existence of a stationary distribution and prove that we can reach this distribution by repeatedly applying transition function T_c. In Subsection 4.1, we will construct the proof of this theorem for a discrete-time Markov chain modeled as a coalgebra (see Subsection 3.2). In Subsection 4.2 we determine which conditions we used for the theorem to hold for discrete-time Markov chains. These conditions form the basis under which the theorem for general coalgebras will hold. Lastly, in Subsection 4.3, we will apply the theorem to a continuous-time Markov chain, which can be modeled using a coalgebra.

4.1 Fixed point theorem for Discrete-time Markov chains

In this subsection, we will rewrite and prove Theorem 2.6 where the Markov chain is modeled as a coalgebra. In order to do so, we will have to define concepts which can be generalized for other coalgebras.

In order to apply the Banach Fixed point theorem, we will need to define a complete metric $\overline{d}$ on $\mathbb{D}S$. In the section below, we will construct metric $\overline{d}$ which is based on the Wasserstein-1 distance on the discrete metric of S (see remark 6.4 in [16]). We will start by defining an expected value function on probability distributions.

We require the expected value function $E : \mathbb{D}\mathbb{R}^{\infty}_{\geq 0} \to \mathbb{R}^{\infty}_{\geq 0}$ of $\mathbb{D}\mathbb{R}^{\infty}_{\geq 0}$ on $\mathbb{R}^{\infty}_{\geq 0} = \mathbb{R}_{\geq 0} \cup \{\infty\}$ to be an Eilenberg-Moore algebra. By formulating this expected value as an Eilenberg-Moore algebra, we can generalize this concept to other functors F while still respecting the underlying monad structure. In the case of functor $\mathbb{D}$, this algebra can be defined using the standard expectation:

$$E(\nu) = \sum_{x \in \operatorname{supp}(\nu)} x \cdot \nu(x).$$

Note that there exists probability distributions $\nu \in \mathbb{D}\mathbb{R}^{\infty}_{\geq 0}$ with $\infty \in \operatorname{supp}(\nu)$. We define multiplication and summation with respect to ∞ as:

$$\infty + x = x + \infty = \infty \text{ for all } x \in \mathbb{R}^{\infty}_{\geq 0}$$

$$x \cdot \infty = \infty \cdot x = \begin{cases} \infty & \text{if } x \neq 0 \\ 0 & \text{if } x = 0 \end{cases}.$$

To prove that E is a EM-algebra, we have to prove preservation of η and μ.

Preservation of η

For $x \in \mathbb{R}^{\infty}_{\geq 0}$ we have $E \circ \eta_{\mathbb{R}^{\infty}_{\geq 0}}(x) = E(\delta_x) = \sum_{x \in \operatorname{supp}(\delta_x)} x \cdot \delta_x(x) = x = \operatorname{id}_{\mathbb{R}^{\infty}_{\geq 0}}(x)$, which proves the preservation of η.

Preservation of μ

For $\nu \in \mathbb{D}^2\mathbb{R}^{\infty}_{\geq 0}$ and $r \in \mathbb{R}^{\infty}_{\geq 0}$ we have

$$\mu_{\mathbb{R}^{\infty}_{\geq 0}}(\nu)(r) = \sum_{\delta \in \operatorname{supp}(\nu)} \nu(\delta) \cdot \delta(r).$$

And thus

$$\begin{aligned} E \circ \mu_{\mathbb{R}^{\infty}_{\geq 0}}(\nu) &= \sum_{r \in \operatorname{supp}(\mathbb{D}E(\nu))} r \cdot \mu_{\mathbb{R}^{\infty}_{\geq 0}}(\nu) = \sum_{r \in \operatorname{supp}(\mathbb{D}E(\nu))} r \cdot \left(\sum_{\delta \in \operatorname{supp}(\nu)} \nu(\delta) \cdot \delta(r) \right) \\ &= \sum_{\delta \in \operatorname{supp}(\nu)} \nu(\delta) \cdot \left(\sum_{r \in \mathbb{R}} r \cdot \delta(r) \right) = \sum_{\delta \in \operatorname{supp}(\nu)} \nu(\delta) \cdot E(\delta). \end{aligned}$$

Here we can interchange the summations by applying Tonelli's theorem for interchanging summations with non-negative terms. For $r \in \mathbb{R}^{\infty}_{\geq 0}$ we have

$$\mathbb{D}E(\nu)(r) = \sum_{\delta \in \operatorname{supp}(\nu) : E(\delta) = r} \nu(\delta).$$

Therefore, we have

$$\begin{aligned} E \circ \mathbb{D}E(\nu) &= \sum_{r \in \operatorname{supp}(\mathbb{D}E(\nu))} r \cdot \left(\sum_{\delta \in \operatorname{supp}(\nu) : E(\delta) = r} \nu(\delta) \right) = \sum_{r \in \operatorname{supp}(\mathbb{D}E(\nu))} \sum_{\delta \in \operatorname{supp}(\nu) : E(\delta) = r} r \cdot \nu(\delta) \\ &= \sum_{\delta \in \operatorname{supp}(\nu)} \nu(\delta) \cdot E(\delta) = E \circ \mu_{\mathbb{R}}(\nu). \end{aligned}$$

This concludes the prove for preservation of μ.

Using E, we will construct an expected value on $\mathbb{D}S$. First consider the partial order $\sqsubseteq$ on $\mathbf{Set}(A, \mathbb{R}^{\infty}_{\geq 0})$ for all $A \in \mathrm{ob}(\mathbf{Set})$ as the point-wise order under the standard order $(\leq)$ of $\mathbb{R}_{\geq 0} \cup \infty$:

$$f \sqsubseteq g \;\Leftrightarrow f(x) \leq g(x) \text{ for all } x \in A.$$

Next, let $e : S \to \mathbb{R}^{\infty}_{\geq 0}$ with $e(x) = 1$ for all $x \in S$. Using these two concepts, we can define our expected value function on $\mathbb{D}S$. For $f : S \to \mathbb{R}^{\infty}_{\geq 0}$ with $f \sqsubseteq e$, we define $E_f := E \circ \mathbb{D}f : \mathbb{D}S \to \mathbb{R}^{\infty}_{\geq 0}$. For $\delta \in \mathbb{D}S$ we have

$$E_f(\delta) = E \circ \mathbb{D}f(\delta) = \sum_{x \in \mathrm{supp}(\mathbb{D}f(\delta))} x \cdot \sum_{y \in \mathrm{supp}(\delta): f(y) = x} \delta(y) =$$

$$\sum_{x \in \mathrm{supp}(\mathbb{D}f(\delta))} \; \sum_{y \in \mathrm{supp}(\delta): f(y) = x} f(y) \cdot \delta(y) = \sum_{y \in \mathrm{supp}(\delta)} f(y) \cdot \delta(y).$$

Here the last equality follows from $x \in \mathrm{supp}(\mathbb{D}f(\delta)) \Leftrightarrow \sum_{y \in \mathrm{supp}(\delta): f(y) = x} \delta(x) \neq 0$, and thus $\mathrm{supp}(\mathbb{D}f(\delta)) = f(\mathrm{supp}(\delta)) = \{f(y) \mid y \in \mathrm{supp}(\delta)\}$. Note that $E_f(\delta)$ is equivalent to the standard expected value $\mathbb{E}_\delta[f]$ on the set of distributions of S.

Also note that, since $f(s) \leq 1$ for all $s \in S$, we have

$$E_f(\delta) = \sum_{y \in \mathrm{supp}(\delta)} f(y) \cdot \delta(y) \leq \sum_{y \in \mathrm{supp}(\delta)} \delta(y) = 1.$$

Proposition 4.1. (Metric on $\mathbb{D}S$). The mapping $\overline{d} : \mathbb{D}S \times \mathbb{D}S \to \mathbb{R}$ with

$$\overline{d}(\nu, \delta) = \sup_{f \sqsubseteq e} |E_f(\nu) - E_f(\delta)| \;\text{ for } \nu, \delta \in \mathbb{D}S$$

is a metric on $\mathbb{D}S$.

Proof. First note that $E_f(\nu), E_f(\delta) \leq 1$ for all $\nu, \delta \in \mathbb{D}S$ and all $f \sqsubseteq e$. Therefore, $|E_f(\nu) - E_f(\delta)|$ is well defined and bounded by 1. As a result, $\overline{d}(\nu, \delta)$ is also bounded by 1.

In order to prove that $\overline{d}$ is a metric on $\mathbb{D}S$, we need to prove reflexivity, symmetry and the triangle inequality.

Reflexivity
For $\nu \in \mathbb{D}S$ we have $\overline{d}(\nu, \nu) = \sup_{f \sqsubseteq e} |E_f(\nu) - E_f(\nu)| = 0$. Moreover, for $\nu, \rho \in \mathbb{D}S$ with $\nu \neq \rho$ there exists at least one $s \in S$ with $\nu(s) \neq \rho(s)$. Consider the function $\delta_s : S \to \mathbb{R}$ with

$$\delta_s(x) = \begin{cases} 1 & \text{if } x = s \\ 0 & \text{otherwise} \end{cases}.$$

Since $\delta_s \sqsubseteq e$, we have

$$\overline{d}(\nu, \rho) = \sup_{f \sqsubseteq e} |E_f(\nu) - E_f(\rho)| \geq |E_{\delta_s}(\nu) - E_{\delta_s}(\rho)| = |\nu(s) - \rho(s)| > 0.$$

Symmetry
For $\rho, \nu \in \mathbb{D}S$ we have

$$\overline{d}(\nu, \rho) = \sup_{f \sqsubseteq e} |E_f(\nu) - E_f(\rho)| = \sup_{f \sqsubseteq e} |E_f(\rho) - E_f(\nu)| = \overline{d}(\rho, \nu).$$

Triangle inequality
For $\psi, \nu, \rho \in \mathbb{D}S$ we have

$$\overline{d}(\psi, \rho) = \sup_{f \sqsubseteq e} |E_f(\psi) - E_f(\rho)| = \sup_{f \sqsubseteq e} |(E_f(\psi) - E_f(\nu)) + (E_f(\nu) - E_f(\rho))|$$

$$\leq \sup_{f \sqsubseteq e}(|E_f(\psi) - E_f(\nu)| + |E_f(\nu) - E_f(\rho)|) \leq (\sup_{f \sqsubseteq e} |E_f(\psi) - E_f(\nu)|) + (\sup_{g \sqsubseteq e} |E_g(\nu) - E_g(\rho)|)$$

$$= \overline{d}(\psi, \nu) + \overline{d}(\nu, \rho).$$

□

Proposition 4.2. (completeness of metric $\overline{d}$). The metric space $(\mathbb{D}S, \overline{d})$ with $\overline{d} : (\mathbb{D}S)^2 \to \mathbb{R}_{\geq 0}$ the metric from Proposition 4.1 is complete.

Proof. Let $(\lambda_n)_{n \in \mathbb{N}}$ be a Cauchy sequence in $\mathbb{D}S$.
Let $z \in S$, and let $\delta_z : S \to \mathbb{R}^{\infty}_{\geq 0}$ be the function with

$$\delta_z(x) = \begin{cases} 1 & \text{if } x = z \\ 0 & \text{if } x \neq z \end{cases}.$$

By definition of a Cauchy sequence, for all $z \in S$ and for all $\epsilon \in \mathbb{R}_{>0}$ there exists an $N \in \mathbb{N}$ such that for all $m, n > N$, we have

$$\epsilon > \overline{d}(\lambda_n, \lambda_m) = \sup_{f \sqsubseteq e} |E_f(\lambda_n) - E_f(\lambda_m)| \geq |E_{\delta_z}(\lambda_n) - E_{\delta_z}(\lambda_m)|$$

$$= \left| \sum_{x \in \text{supp}(\lambda_n)} \delta_z(x)\lambda_n(x) - \sum_{x \in \text{supp}(\lambda_m)} \delta_z(x)\lambda_m(x) \right| = |\lambda_n(z) - \lambda_m(z)|.$$

Therefore, for all $z \in S$, the sequence $(\lambda_n(z))_{n \in \mathbb{N}}$ is a Cauchy sequence in $\mathbb{R}_{\geq 0}$ with the Euclidean metric. Since $\mathbb{R}_{\geq 0}$ is complete, $\lambda_n(z)$ converges to a unique value $\lambda(z) \in \mathbb{R}_{\geq 0}$ for all $z \in S$.

Now that we have proven that λ_n point-wise converges to a function $\lambda : S \to [0, \infty)$, we will prove that λ_n converges under $\overline{d}$ to λ. We start by proving that $\lambda \in \mathbb{D}S$.

Consider $\overline{S} = \bigcup_{n \in \mathbb{N}} \text{supp}(\lambda_n)$. Since $\overline{S}$ is a countable union of countable sets, $\overline{S}$ is countable. Moreover, for $x \in S \setminus \overline{S}$ we have $\lambda(x) = \lim_{n \to \infty} \lambda_n(x) = \lim_{n \to \infty} 0 = 0$. Therefore, $\text{supp}(\lambda) \subseteq \overline{S}$.

Since $\overline{S}$ is countable, we can write it as $\overline{S} = \{a_1, a_2, ...\}$. Consider $F_m = \{a_1, a_2, ..., a_m\}$. Since λ_n pointwise converges to λ, the finite sum $\sum_{x \in F_m} \lambda_n(x)$ converges to $\sum_{x \in F_m} \lambda(x)$ for all $m \in \mathbb{N}_{\geq 1}$. Therefore, for all $m \in \mathbb{N}_{\geq 1}$ and for all $\epsilon > 0$ there exists an $N \in \mathbb{N}$ such that for all $n \geq N$ we have

$$-\epsilon < \sum_{x \in F_m} \lambda_n(x) - \sum_{x \in F_m} \lambda(x) < \epsilon.$$

By taking $m \to \infty$, we get

$$-\epsilon < \sum_{x \in \overline{S}} \lambda_n(x) - \sum_{x \in \overline{S}} \lambda(x) < \epsilon.$$

Since λ_n is a distribution over S, the sum over the probabilities is equal to 1. Therefore, we have

$$-\epsilon < 1 - \sum_{x \in \overline{S}} \lambda(x) < \epsilon.$$

Since this holds for arbitrary small $\epsilon > 0$, we have $\sum_{x \in \overline{S}} \lambda(x) = 1$. Moreover, since $\lambda_n(x) \geq 0$ for all $n \in \mathbb{N}$ and λ_n pointwise converges to λ, we have $\lambda(x) \geq 0$ for all $x \in S$. Therefore, $\lambda \in \mathbb{D}S$.

Let $f \sqsubseteq e$ be arbitrary chosen. We have

$$|E_f(\lambda_n) - E_f(\lambda)| = \left| \sum_{x \in \overline{S}} f(x)\lambda_n(x) - \sum_{x \in \overline{S}} f(x)\lambda(x) \right|$$

$$= \left| \sum_{x \in \overline{S}} f(x)(\lambda_n(x) - \lambda(x)) \right| \leq \sum_{x \in \overline{S}} f(x) \left| \lambda_n(x) - \lambda(x) \right| \leq \sum_{x \in \overline{S}} \left| \lambda_n(x) - \lambda(x) \right|.$$

In the first inequality, we use the triangle inequality. We are allowed to use the triangle inequality if $\sum_{x \in \overline{S}} f(x) |\lambda_n(x) - \lambda(x)|$ converges. Since $\sum_{x \in \overline{S}} f(x) |\lambda_n(x) - \lambda(x)| \leq \sum_{x \in \overline{S}} |\lambda_n(x) - \lambda(x)| \leq 2$, and since each term is non-negative, we have convergence.

Let $\epsilon > 0$ be arbitrarily chosen. Since $F_m \to \overline{S}$ when $m \to \infty$, we have that $\sum_{x \in F_m} \lambda(x) \to 1$ when $m \to \infty$. Therefore, there exist an $N \in \mathbb{N}$ such that for all $m \geq N$ we have

$$\left| \sum_{x \in F_m} \lambda(x) - 1 \right| < \epsilon.$$

Since $\sum_{x \in F_m} \lambda(x) \leq 1$, we can rewrite the inequality to

$$\sum_{x \in F_m} \lambda(x) > 1 - \epsilon.$$

As a result, we have

$$\sum_{x \notin F_m} \lambda(x) < \epsilon.$$

For all $n \in \mathbb{N}$ we have

$$\sum_{x \in \overline{S}} |\lambda_n(x) - \lambda(x)| = \sum_{x \in F_m} |\lambda_n(x) - \lambda(x)| + \sum_{x \notin F_m} |\lambda_n(x) - \lambda(x)| .$$

Since λ_n pointwise converges to λ, and since F_m is a finite set, for sufficiently large n we have $\sum_{x \in F_m} |\lambda_n(x) - \lambda(x)| < \epsilon$.

Moreover, using the triangle inequality on $|\lambda_n(x) - \lambda(x)|$ for all $n \in F_m$, we get

$$\sum_{x \notin F_m} |\lambda_n(x) - \lambda(x)| \leq \sum_{x \notin F_m} \lambda_n(x) + \lambda(x) = \sum_{x \notin F_m} \lambda_n(x) + \sum_{x \notin F_m} \lambda(x) < \sum_{x \notin F_m} \lambda_n(x) + \epsilon.$$

Since λ_n is a probability distribution, and $\sum_{x \in F_m} \lambda_n(x)$ converges to $\sum_{x \in F_m} \lambda(x)$, for sufficiently large n we have

$$\sum_{x \notin F_m} \lambda_n(x) = 1 - \sum_{x \in F_m} \lambda_n(x) < 1 - \sum_{x \in F_m} \lambda(x) + \epsilon = \sum_{x \notin F_m} \lambda(x) + \epsilon < 2\epsilon.$$

By combining the two results above, we get

$$\sum_{x \notin F_m} |\lambda_n(x) - \lambda(x)| < \sum_{x \notin F_m} \lambda_n(x) + \epsilon < 3\epsilon.$$

Therefore, for sufficiently large n, we have

$$|E_f(\lambda_n) - E_f(\lambda)| \leq \sum_{x \in \overline{S}} |\lambda_n(x) - \lambda(x)|$$

$$= \sum_{x \in F_m} |\lambda_n(x) - \lambda(x)| + \sum_{x \notin F_m} |\lambda_n(x) - \lambda(x)| < \epsilon + 3\epsilon = 4\epsilon.$$

Since this inequality holds for all $\epsilon > 0$ and arbitrary $f \sqsubseteq e$, this proves that, for all $f \sqsubseteq e$, $|E_f(\lambda_n) - E_f(\lambda)| \to 0$ when $n \to \infty$. Consequently, we have

$$d(\lambda_n, \lambda) = \sup_{f \sqsubseteq e} |E_f(\lambda_n) - E_f(\lambda)| \to 0 \text{ when } n \to \infty.$$

Since the Cauchy sequence $(\lambda_n)_{n \in \mathbb{N}}$ converges in $(\mathbb{D}S, \overline{d})$, we have proven that $(\mathbb{D}S, \overline{d})$ is complete. $\square$

Another property of $\overline{d}$ is non-expansiveness under multiplication by Markov kernels. This property is described, and proven in Proposition 4.3. This property will prove useful in the proof of the main theorem of this subsection.

Proposition 4.3. (non-expansiveness under transformation by a coalgebra). Let $\overline{d}$ be the metric defined in Proposition 4.1 and let $k : S \to \mathbb{D}S$ be a coalgebra with corresponding transition function $T_k : \mathbb{D}(S) \to \mathbb{D}(S)$ given in Definition 3.27. Then, for all $\nu, \delta \in \mathbb{D}S$, we have

$$\overline{d}(T_k(\nu), T_k(\delta)) \leq \overline{d}(\nu, \delta).$$

Proof. Let $\nu, \delta \in \mathbb{D}S$ and let $f \sqsubseteq e$ be arbitrary chosen. Let $k : S \to \mathbb{D}S$ be an arbitrary coalgebra, with transition function T_k. We have

$$E_f(T_k(\nu)) = \sum_{j \in \mathrm{supp}(T_k(\nu))} T_k(\nu)(j) \cdot f(j) = \sum_{j \in \mathrm{supp}(T_k(\nu))} \left(\sum_{i \in \mathrm{supp}(\nu)} k(i)(j) \cdot \nu(i) \right) f(j)$$

$$= \sum_{i \in \mathrm{supp}(\nu)} \nu(i) \left(\sum_{j \in \mathrm{supp}(k(i))} k(i)(j) \cdot f(j) \right) = \sum_{i \in \mathrm{supp}(\nu)} \nu(i) E_f(k(i)).$$

Here we interchange the summation using Tonelli's theorem for interchanging summations with non-negative terms. Consider $g_f : S \to \mathbb{R}^{\infty}_{\geq 0}$ with $g_f(i) = E_f(k(i))$. Since $0 \leq f(x) \leq 1$ for all $x \in S$, we have

$$g_f(i) = E_f(k(i)) = \sum_{j \in \mathrm{supp}(k(i))} k(i)(j) \cdot f(j) \leq \sum_{j \in \mathrm{supp}(k(i))} k(i)(j) = 1.$$

Therefore, we have $g_f \sqsubseteq e$. Moreover, for any distribution $\nu \in \mathbb{D}S$, we have

$$E_f(T_k(\nu)) = \sum_{i \in \mathrm{supp}(\nu)} \nu(i) E_f(k(i)) = E_{g_f}(\nu).$$

Therefore, we have

$$\overline{d}(T_k(\nu), T_k(\delta)) = \sup_{f \sqsubseteq e} |E_f(T_k(\nu)) - E_f(T_k(\delta))|$$

$$= \sup_{f \sqsubseteq e} |E_{g_f}(\nu) - E_{g_f}(\delta)| \leq \sup_{h \sqsubseteq e} |E_h(\nu) - E_h(\delta)| = \overline{d}(\nu, \delta)$$

$\square$

The last tool we need in order to formulate the fixed point theorem is a partial order on $\mathbb{D}S$. However, instead of defining this order on $\mathbb{D}S$, we will provide an order on the set of partial distributions $\mathbb{D}_{\leq 1}S$ on S.

Definition 4.4. (partial probability functor). The partial probability distribution functor $\mathbb{D}_{\leq 1} : \mathbf{Set} \to \mathbf{Set}$ maps an object X to the set of all partial probability distributions with countable support:

$$\mathbb{D}_{\leq 1}X = \left\{ \nu : X \to \mathbb{R}_{\geq 0} \mid \sum_{x \in \mathrm{supp}(\nu)} \nu(x) \leq 1 \right\}.$$

A morphism $f : X \to Y$ is mapped to $\mathbb{D}_{\leq 1}f : \mathbb{D}_{\leq 1}X \to \mathbb{D}_{\leq 1}Y$ such that for all $\nu \in \mathbb{D}_{\leq 1}X$ and all $y \in Y$ we have

$$\mathbb{D}_{\leq 1}f(\nu)(y) = \sum_{x \in \mathrm{supp}(\nu) : f(x) = y} \nu(x).$$

In order to prove that $\mathbb{D}_{\leq 1}$ is a functor, we need to show that it preserves the identity morphisms and it preserves composition. This proof is identical of the proof for functor $\mathbb{D}$ and can be found in appendix A.

Since the probabilities of a partial probability distribution does not have add up to one, we can introduce the following non-trivial partial order to $(A, \mathbb{D}_{\leq 1}X)$.

Lemma 4.5. Consider $f, g \in \mathbf{Set}(Y, \mathbb{D}_{\leq 1}X)$. We define the relation $\leq_{\mathbb{D}}$ on $\mathbf{Set}(Y, \mathbb{D}_{\leq 1}X)$ as

$$f \leq_{\mathbb{D}} g \Leftrightarrow f(y)(x) \leq g(y)(x) \text{ for all } y \in Y \text{ and all } x \in X$$

with $\leq$ is the standard order on $\mathbb{R}$. Then, $\leq_{\mathbb{D}}$ is a partial order on $\mathbf{Set}(Y, \mathbb{D}_{\leq 1}X)$.

Since $\leq$ is a partial order on $\mathbb{R}$, it is easy to show that reflexivity, antisymmetry and transitivity can be transferred to $\leq_{\mathbb{D}}$.

Since the total mass of a partial probability distribution does not have to add up to 1, we can define scalar multiplication with scalars in $[0, 1]$ on $\mathbb{D}_{\leq 1}S$.

Definition 4.6. (scalar multiplication on $\mathbb{D}S$). Consider $\alpha \in [0, 1]$. We define the scalar multiplication function $\alpha_{-} : \mathbf{Set}(Y, \mathbb{D}_{\leq 1}X) \to \mathbf{Set}(Y, \mathbb{D}_{\leq 1}X)$ as $(\alpha f)(y)(x) = \alpha \cdot f(y)(x)$ for all $y \in Y$ and all $x \in X$. In other words, α_{-} scales the function values of f with factor α.

Consider a partial probability distribution $\beta : 1 \to \mathbb{D}_{\leq 1}S$, the unique mapping $! : S \to 1$ of terminal object 1, and $c : S \to \mathbb{D}S$. We say that $\alpha\beta \leq_{\mathbb{D}} c$ if $\alpha\beta \circ ! \leq_{\mathbb{D}} c$. In other words, for all $x, y \in S$ and for $* \in 1$ we have $\alpha\beta \circ !(x)(y) = \alpha\beta(*)(y) \leq c(x)(y)$. In this case, we have the inequality commuting diagram below.

$$\begin{tikzcd}
S \arrow[r, "c"] \arrow[d, "!"'] \arrow[dr, dashed, "\leq_{\mathbb{D}}"] & \mathbb{D}S \arrow[d, hook] \\
1 \arrow[r, "\alpha\beta"] & \mathbb{D}_{\leq 1}S
\end{tikzcd}$$

Proposition 4.7. (Transition functions are closed under convex combinations). Let $c : S \to \mathbb{D}S$ be a coalgebra on S and let $T_c : \mathbb{D}S \to \mathbb{D}S$ be the corresponding transition function. Suppose there exists an $\alpha \in (0, 1)$ and $\beta \in \mathbf{Set}(1, \mathbb{D}S)$ with $c \geq_{\mathbb{D}} \alpha\beta$. Then there exists a coalgebra $k : S \to \mathbb{D}S$ such that for all $\nu \in \mathbb{D}S$ and all $y \in S$ we have

$$T_c(\nu)(y) = \alpha \cdot \beta(*)(y) + (1 - \alpha) \cdot T_k(\nu)(y).$$

Proof. For readability, we will define $\dot{\beta} := \beta(*) \in \mathbb{D}S$. Consider $k : S \to (S \to \mathbb{R})$ defined by

$$k(x)(y) = \frac{1}{1 - \alpha} \cdot c(x)(y) - \frac{\alpha}{1 - \alpha} \cdot \dot{\beta}(y).$$

We will start by proving k is a coalgebra. Since $c \geq_{\mathbb{D}} \alpha\beta$, for all $x, y \in S$ we have $c(x)(y) \geq \alpha \cdot \dot{\beta}(y)$, and thus

$$k(x)(y) \geq \frac{\alpha}{1 - \alpha} \cdot \dot{\beta}(y) - \frac{\alpha}{1 - \alpha} \cdot \dot{\beta}(y) = 0.$$

Moreover, when we take a constant $x \in S$ and sum y over S, we get

$$\sum_{y \in \operatorname{supp}(k(x))} k(x)(y) = \sum_{y \in \operatorname{supp}(k(x))} \frac{1}{1 - \alpha} \cdot c(x)(y) - \frac{\alpha}{1 - \alpha} \cdot \dot{\beta}(y)$$

$$= \frac{1}{1-\alpha}\left(\sum_{y \in \mathrm{supp}(c(x))} c(x)(y)\right) - \frac{\alpha}{1-\alpha}\left(\sum_{y \in \mathrm{supp}(\dot{\beta})} \dot{\beta}(y)\right) = \frac{1}{1-\alpha} - \frac{\alpha}{1-\alpha} = 1.$$

The two results above prove that $k(x) \in \mathbb{D}S$ for all $x \in S$, and thus $k : S \to \mathbb{D}S$ is a coalgebra.

For all $\nu \in \mathbb{D}S$ and all $x \in S$ we have

$$T_k(\nu)(y) = \sum_{x \in \mathrm{supp}(\nu)} \nu(x) \cdot k(x)(y) = \sum_{x \in \mathrm{supp}(\nu)} \nu(x)\left(\frac{1}{1-\alpha} \cdot c(x)(y) - \frac{\alpha}{1-\alpha} \cdot \dot{\beta}(y)\right)$$

$$= \left(\sum_{x \in \mathrm{supp}(\nu)} \nu(x) \cdot \frac{1}{1-\alpha} \cdot c(x)(y)\right) - \left(\sum_{x \in \mathrm{supp}(\nu)} \nu(x)\right) \cdot \frac{\alpha}{1-\alpha} \cdot \dot{\beta}(y) = \frac{1}{1-\alpha} \cdot T_c(\nu)(y) - \frac{\alpha}{1-\alpha} \cdot \dot{\beta}(y).$$

By rewriting this expression, we get our result: For all $\nu \in \mathbb{D}S$ and all $y \in S$ we have

$$T_c^1(\nu)(y) = \alpha \cdot \dot{\beta}(y) + (1-\alpha) \cdot T_k^1(\nu)(y).$$

□

Equipped with the tools listed above, we can formulate and proof the main theorem of this subsection.

Theorem 4.8. (Limit point of $(T_c^1)^n$).
Let $c : S \to \mathbb{D}S$ be a coalgebra on state space S. Assume there exists an $\alpha \in (0,1)$ and a probability distribution $\beta : 1 \to \mathbb{D}S$ with $c \geq_{\mathbb{D}} \alpha\beta$. Then, for any distribution $\nu : 1 \to \mathbb{D}S$, $((\hat{T}_c^1)^n(\nu))_{n \in \mathbb{N}}$ converges to a unique stationary distribution $\pi : 1 \to \mathbb{D}S$. Here $\hat{T}_c^1$ is the transition function of coalgebra c defined in Definition 3.27.

Proof. For the sake of readability, we define $\dot{\beta} := \beta(*) \in \mathbb{D}S$. Take an arbitrary probability distribution $\nu \in \mathbb{D}S$. Since $c \geq_{\mathbb{D}} \alpha\beta$, by Proposition 4.7, there exist a coalgebra $k : S \to \mathbb{D}S$ such that for all $j \in S$ we have

$$T_c(\nu)(j) = \alpha \cdot \dot{\beta}(j) + (1-\alpha) \cdot T_k(\nu)(j).$$

Take arbitrary probability $\nu, \delta \in \mathbb{D}S$. For every $f \sqsubseteq e$, we have

$$E_f(T_c(\nu)) = \sum_{i \in \mathrm{supp}(T_c(\nu))} f(i) \cdot T_c(\nu)(i) = \sum_{i \in \mathrm{supp}(\dot{\beta}) \cup \mathrm{supp}(T_k(\nu))} f(i)(\alpha \cdot \dot{\beta}(i) + (1-\alpha) \cdot T_k(\nu)(i))$$

$$= \alpha E_f(\dot{\beta}) + (1-\alpha) E_f(T_k(\nu)).$$

Therefore

$$|E_f(T_c(\nu)) - E_f(T_c(\delta))| = \left|\alpha E_f(\dot{\beta}) + (1-\alpha)E_f(T_k(\nu)) - (\alpha E_f(\dot{\beta}) + (1-\alpha)E_f(T_k(\delta)))\right|$$

$$= (1-\alpha)\left|E_f(T_k(\nu)) - E_f(T_k(\delta))\right|.$$

By taking the supremum over this expression, we get

$$\begin{aligned}
\overline{d}(T_c(\nu), T_c(\delta)) &= \sup_{f \sqsubseteq e} |E_f(T_c(\nu)) - E_f(T_c(\delta))| \\
&= \sup_{f \sqsubseteq e} ((1-\alpha) \, |E_f(T_k(\nu)) - E_f(T_k(\delta))|) \\
&= (1-\alpha) \cdot \sup_{f \sqsubseteq e} |E_f(T_k(\nu)) - E_f(T_k(\delta))| \\
&= (1-\alpha)\overline{d}(T_k(\nu), T_k(\delta)) \\
&\leq (1-\alpha)\overline{d}(\nu, \delta).
\end{aligned}$$

The last step is by Proposition 4.3. By the Banach's fixed point theorem, we get that there exists a unique fixed point $\dot{\pi} \in \mathbb{D}S$ of T_c with $\lim_{n\to\infty} (T_c)^n(\nu) = \dot{\pi}$. Therefore, $\pi : 1 \to \mathbb{D}S$ with $\pi(*) = \dot{\pi}$ is a unique stationary point of $\hat{T}_c^1$ with $(\hat{T}_c^1)^n(\nu) \to \pi$ when $n \to \infty$.

□

4.2 Limit points of the General Coalgebra

In this section we generalize Theorem 4.8 for finding a limit point for a discrete-time Markov chain to other coalgebras $c : S \to FS$. We achieve this goal by determining which assumptions we need to take on the coalgebra, functor and the underlying category in order for the theorem to hold. These assumptions become the conditions for the general limit theorem. To prove that the theorem is applicable to some other coalgebra than the discrete-time Markov chain, in Subsection 4.3 we apply the theorem to the continuous-time Markov chain.

Before we will list the assumptions, we need to define two categories which we will be using.

Definition 4.9. (Category of Posets). We define category **PoSet** with posets $(X, \leq_X)$ as objects and functions $f : (X, \leq_X) \to (Y, \leq_Y)$ with $x \leq_X y \Rightarrow f(x) \leq_Y f(y)$ as morphisms. In other words, $\mathbf{PoSet}((X, \leq_X), (Y, \leq_Y))$ consists of all monotone functions between $(X, \leq_X)$ and $(Y, \leq_Y)$.

Definition 4.10. (Category of Metric spaces). We define the category **Met** with metric spaces (X, d_X) as objects and functions $f : (X, d_X) \to (Y, d_Y)$ with $d_Y(f(x), f(y)) \leq d_X(x, y)$ as morphisms. In other words $\mathbf{Met}((X, d_X), (Y, d_Y))$ consists of all non-expansive functions between (X, d_X) and (Y, d_Y).

In the section below, we will enumerate the conditions needed for the theorem. The goal of this thesis is to construct a categorical method to find limit-points on coalgebras. Therefore we state conditions on arbitrary categories. Under each assumption, we discuss where this is used in the theorem for the discrete-time Markov chain. Afterwards, we will state, and proof the general limit theorem for coalgebras.

0. **Let $\mathcal{C}$ be a locally small category with a terminal object 1.**

   In the case of the discrete-time Markov chain, we have $\mathcal{C} = \mathbf{Set}$.

1. **We require functors $\mathcal{F} : \mathcal{C} \to \mathcal{C}$ and $\mathcal{K} : \mathcal{C} \to \mathcal{C}$, and a coalgebra $c : S \to \mathcal{F}(S)$.**

   In the general case $c : S \to \mathcal{F}(S)$ would be the coalgebra for which we will find the limit point. In the case of the discrete-time Markov chain, we have $\mathcal{F} = \mathbb{D}$ and $\mathcal{K} = \mathbb{D}_{\leq 1}$. Since the limit-point of the discrete-time Markov chain is stated for regular distribution functions, we require the functor $\mathcal{F}$, and since some required structure, such as the partial order $\leq_{\mathbb{D}}$, is defined on partial distributions, we require an additional functor $\mathcal{K}$. In the general limit theorem, we preform all calculations on $\mathcal{K}(S)$ instead of $\mathcal{F}(S)$. Therefore, we will mainly make assumptions involving $\mathcal{K}(S)$.

2. **We require an inclusion transformation $i : \mathcal{F} \Rightarrow \mathcal{K}$ and a normalization transformation $r : \mathcal{K} \Rightarrow \mathcal{F}$ such that $r \circ i = \mathrm{id}_{\mathcal{F}}$.**

   Since our theorem is stated on functor $\mathcal{F}$, and since, in the general limit theorem, we perform all calculations on $\mathcal{K}$, we need a way to move between these functors. In the case of the discrete-time Markov chain, we have $i_S(\nu)(x) = \nu(x)$, and $r_S(\delta)(x) = \frac{\delta(x)}{\sum_{y \in \mathrm{supp}(\delta)} \delta(y)}$.

   For $\nu \in \mathbb{D}(S)$, the sum over the probabilities is equal to 1. Therefore, when we apply $i_S \circ r_S$ to $\nu \in \mathbb{D}(S)$ and $x \in S$, we get $r_S(i_S(\nu))(x) = \nu(x)$. For this reason, we require $r \circ i = \mathrm{id}_{\mathcal{F}}$.

   Using i and r we can define a normalized morphism $\nu \in \mathcal{C}(A, \mathcal{K}(S))$ and we define what it means for a mapping $f : \mathcal{C}(A, \mathcal{K}(S)) \to \mathcal{C}(A, \mathcal{K}(S))$ to be norm-persevering (see Definition 4.11 and 4.12 underneath the enumeration), which is used in assumption **8.**

3. **For all $A, B \in \mathrm{ob}(\mathcal{C})$ we have that $\mathcal{C}(A, \mathcal{K}(B)) \in \mathbf{PoSet}$. We denote the partial order by $\geq_{\mathcal{K}}$.**

   In the case of discrete-time Markov chain $\geq_{\mathcal{K}}$ is the partial order $\geq_{\mathbb{D}}$ used in the statement of Theorem 4.8 for finding limit-points of discrete-time Markov chains.

4. **We choose an object $R \in \mathrm{ob}(\mathcal{C})$, an Eilenberg-Moore algebra $E : R \to \mathcal{K}(R)$ and a morphism $e : S \to R$ such that the following conditions hold.**

   A) For all $A \in \mathrm{ob}(\mathcal{C})$, we have that $\mathcal{C}(A, R) \in \mathbf{PoSet}$ with partial order $\sqsubseteq$.

   B) For morphism $f : A \to \mathcal{K}(S)$ with $f \sqsubseteq e$ we define $E_f := E \circ \mathcal{K}f : \mathcal{K}(S) \to R$.

   In the case of discrete-time Markov chains, we used a similar construction to define the expectation value. We have $R = \mathbb{R}^{\infty}_{\geq 0}$, $e : S \to \mathbb{R}^{\infty}_{\geq 0}$ with $e(x) = 1$ for all $x \in S$ and an Eilenberg-Moore algebra $E : \mathbb{D}(\mathbb{R}^{\infty}_{\geq 0}) \to \mathbb{R}^{\infty}_{\geq 0}$ such that for all $f \sqsubseteq e$ there exists an expected value function $E_f = E \circ \mathbb{D}f : \mathbb{D}S \to \mathbb{R}^{\infty}_{\geq 0}$.

5. **For all $A \in \mathrm{ob}(\mathcal{C})$ we have**

   A) $\mathcal{C}(A, R) \in \mathbf{Met}$ with metric d^A_R.

   B) For all $f \sqsubseteq e$ we have a function $d^A_f : \mathcal{C}(A, \mathcal{K}(S))^2 \to R$ defined by

   $$d^A_f(\nu, \delta) = d^A_R(E_f \circ \nu, E_f \circ \delta).$$

In Lemma 4.14, we prove that d_f^A forms a pseudometric on $\mathcal{C}(A, \mathcal{K}(S))$. Note that, in the case of a concrete category, it is sufficient to have a metric d_R on R and the pseudometric $\dot{d}_f^A$ on $\mathcal{K}(S)$ defined by

$$\dot{d}_f^A(\phi, \psi) = d_R(E_f(\phi), E_f(\psi)),$$

since these metrics can be transformed into the metrics described above. By taking the supremum metric induced by d_R, we gain metric d_R^A on $\mathcal{C}(A, \mathcal{K}(S))$. Moreover, by taking

$$d_f^A(\nu, \delta) = d_R^A(E_f \circ \nu, E_f \circ \delta) = \sup_{a \in A} d_R(E_f(\nu(a)), E_f(\delta(a))) = \sup_{a \in A} \dot{d}_f^A(\nu(a), \delta(a)),$$

we have a pseudometric d_f^A on $\mathcal{C}(A, \mathcal{K}(S))$.

6. **We define $\overline{d}_S^A : \mathcal{C}(A, \mathcal{K}(S))^2 \to \mathbb{R}_{\geq 0}$ as $\overline{d}_S^A(\nu, \delta) = \sup_{f \sqsubseteq e} d_f^A(\nu, \delta)$ and require the following conditions:**

   A) For all $\delta, \nu : A \to \mathcal{K}S$ and all $f \sqsubseteq e$ there exist $\alpha \in \mathbb{R}_{\geq 0}$ such that $d_f^A(\delta, \nu) \leq \alpha$. In other words, the supremum over all $f \sqsubseteq e$ is finite, and thus $\overline{d}_S^A$ is well-defined.

   B) For all $\delta, \nu : A \to \mathcal{K}S$ with $\delta \neq \nu$ there exist an $f \in \mathcal{C}(S, R)$ such that $E_f \circ \delta \neq E_f \circ \nu$.

   In Lemma 4.15, we show that, under these two conditions, $\overline{d}_S^f$ is a metric on $\mathcal{C}(A, \mathcal{K}(S))$. Note that, similarly to d_f^A , on concrete categories, it is sufficient to have the metric $\overline{d}_S$ on $\mathcal{K}(S)$ with

$$\overline{d}_S(\phi, \psi) = \sup_{f \sqsubseteq e} \dot{d}_S^f(\phi, \psi),$$

   since we can transform it to a metric $\overline{d}_S^A$ on $\mathcal{C}(A, \mathcal{K}(S))$ by taking the supremum:

$$\overline{d}_S^A(\nu, \delta) = \sup_{a \in A} \overline{d}_S(\nu(a), \delta(a)) = \sup_{a \in A} \sup_{f \sqsubseteq e} \dot{d}_S^f(\nu(a), \delta(a))$$

$$= \sup_{f \sqsubseteq e} \sup_{a \in A} \dot{d}_S^f(\nu(a), \delta(a)) = \sup_{f \sqsubseteq e} \overline{d}_S^f(\nu, \delta).$$

   In the case of the discrete-time Markov chain, we defined d_R to be the euclidean distance on $\mathbb{R}_{\geq 0}^{\infty}$. Therefore, we have

$$\overline{d}_S(\nu, \delta) = \sup_{f \sqsubseteq e} \dot{d}_S^f(\nu, \delta) = \sup_{f \sqsubseteq e} d_R(E_f(\phi), E_f(\psi)) = \sup_{f \sqsubseteq e} |E_f(\phi) - E_f(\psi)|.$$

7. **We require $\overline{d}_S^A$ to be complete.**

   Since we want to find the limit point of the coalgebra by applying the Banach fixed point theorem, we require $\overline{d}_S^A$ to be complete.

   Note that, in the case that $\mathcal{C}$ is a concrete category, it is sufficient to require that $\overline{d}_S$ defined in condition 6 to be complete. Since $\overline{d}_S^A(\nu, \delta) = \sup_{a \in A} \overline{d}_S(\nu(a), \delta(a))$, Proposition 4.13 stated below the enumeration gives that $\overline{d}_S^A$ is also complete.

8. **We define $T_c := (i \circ c)^{\#_{\mathcal{K}}} = \mu \circ \mathcal{K}(i \circ c) : \mathcal{K}(S) \to \mathcal{K}(S)$ and we require $\hat{T}_c^A : \mathcal{C}(A, \mathcal{K}(S)) \to \mathcal{C}(A, \mathcal{K}(S))$ defined as $\hat{T}_c^A(\nu) := T_c \circ \nu$ to be norm preserving.**

   Using the Kleisli-extension, we have created an endomorphism T_c which calculates one step of the coalgebra. By using the composition operator, we transform T_c to $\hat{T}_c^A$ which works on the collection of morphisms with domain A. In the limit theorem for the general case, we will find a limit point for $\hat{T}_c^1$ by iteratively applying $\hat{T}_c^1$. Since we require the limit-point to be an element of $\mathcal{F}(S)$, we require $\hat{T}_c^1$ to be norm-preserving. For discrete-time Markov chains, we used a similar endomorphism T_c defined as $T_c := c^{\#_{\mathbb{D}}} : \mathbb{D}(S) \to \mathbb{D}(S)$ on which we apply the theorem for finding limit points of the discrete-time Markov chain.

In the conditions below, we state the requirements which correspond with assumptions needed in the statement and proof of Theorem 4.8 for finding limit points of discrete-time Markov chains. Below the requirements, we shortly explain where it is used in 4.8.

9. **There exist a rescaling operator $\otimes_{\mathcal{K}}^A : [0,1] \times \mathcal{C}(A, \mathcal{K}S) \to \mathcal{C}(A, \mathcal{K}S)$ such that there exists an $\alpha \in (0,1)$ and normalized morphism $\delta \in \mathcal{C}(1, \mathcal{K}(S))$ such that $\hat{T}_c^1(\nu) \geq_{\mathcal{K}} \alpha \otimes_{\mathcal{K}}^1 \delta \circ !$ for all normalized morphisms $\nu \in \mathcal{C}(1, \mathcal{K}(S))$.**
   In the rest of this thesis, we will write $\hat{T}_c^1 \geq_{\mathcal{K}} \alpha \otimes_{\mathcal{K}}^1 \delta$, and thus omit !.

   In the case of the discrete-time Markov chain, this corresponds to the main assumption of the theorem: There exists an $\alpha \in (0,1)$ and $\beta : 1 \to \mathbb{D}(S)$ such that $c \geq_{\mathbb{D}} \alpha\beta$. In the proof of Theorem 4.8, we show that $T_c^1 \geq_{\mathbb{D}} \alpha\beta$. Therefore, in the general case, we directly assume $\hat{T}_c^1 \geq_{\mathcal{K}} \alpha \otimes_{\mathcal{K}}^A \delta$.

10. **For all $A \in \text{ob}(\mathcal{C})$, there exist a convex combination operator $\odot_{\mathcal{K}}^A : [0,1] \times \mathcal{C}(A, \mathcal{K}S)^2 \to \mathcal{C}(A, \mathcal{K}S)$. Moreover, if there exists a normalized morphism $\delta \in \mathcal{C}(A, \mathcal{K}(S))$ there exists an $\alpha \in (0,1)$ such that for all normalized morphisms $\nu \in \mathcal{C}(A, \mathcal{K}(S))$ we have $\hat{T}_c^1(\nu) \geq_{\mathcal{K}} \alpha \otimes_{\mathcal{K}}^A \delta$. Then there exists a coalgebra $k : S \to \mathcal{K}S$ with transition function T_c^1 such that**

$$\boldsymbol{\hat{T}_c^1(\nu) = \odot_{\mathcal{K}}^A(\alpha, \delta, \hat{T}_k^1(\nu)).}$$

   In the proof Theorem 4.8, we construct a transition function $T_k : \mathbb{D}(S) \to \mathbb{D}(S)$ of a coalgebra $k : S \to \mathbb{D}S$ for which $T_c = \alpha \cdot \beta(*) + (1-\alpha) \cdot T_k$. By stating this requirement, we enforce we have a transition function $\overline{T}_k$, which serves a similar purpose in the general case.

11. **For all objects A of category $\mathcal{C}$, there exists a convex combination operator $\odot_R^A : [0,1] \times \mathcal{C}(A, R)^2 \to \mathcal{C}(A, R)$ such that for $\alpha \in [0,1]$, $f \sqsubseteq e$, $\nu, \delta \in \mathcal{C}(A, \mathcal{K}(S))$ we have**

$$\boldsymbol{E_f \circ \odot_{\mathcal{K}}^A(\alpha, \nu, \delta) = \odot_R^A(\alpha, E_f \circ \nu, E_f \circ \delta).}$$

   In the proof of Theorem 4.8, we use this assumption in order to show that

$$E_f(T_c(\nu)) = \alpha E_f(\beta(*)) + (1-\alpha) E_f(T_k(\nu)),$$

   where transition function T_c is a convex combination of β and T_k.

12. **For any $g, h, k \in \mathcal{C}(A, R)$ and for $\alpha \in [0,1]$ we have**

$$\boldsymbol{d_R^A(\odot_R^A(\alpha, g, h), \odot_R^A(\alpha, g, k)) = (1-\alpha) d_R^A(h, k).}$$

 In the proof of Theorem 4.8, we use this assumption when showing that

$$\overline{d}(T_c(\nu), T_c(\delta)) = (1-\alpha)\overline{d}(T_k(\nu), T_k(\delta)).$$

Here T_c is the transition function of discrete-time Markov chain c, and T_k is the transition function given by condition **10.**

13. **For any coalgebra $k : S \to \mathcal{K}(S)$, and the corresponding transition function $T_k := \mu \circ \mathcal{K}(k)$, we require $\hat{T}_k : \mathcal{C}(1, \mathcal{K}(S)) \to \mathcal{C}(1, \mathcal{K}(S))$ to be a non-expansive under $\overline{d}_S^1$.**

 Note that non-expansiveness is a rather strong assumption. Since it must hold for all coalgebras k, this condition restricts the scope on which systems the theorem can be applied greatly.

 In the case of the discrete-time Markov chain, this condition is used in order to prove that

$$(1-\alpha) \cdot \overline{d}(T_k(\nu), T_k(\delta)) \leq (1-\alpha) \cdot \overline{d}(\nu, \delta),$$

 which is the last inequality needed for a contraction.

Below, we state additional definitions and proposition used in the enumeration.

Definition 4.11. (Normalized). We call $\nu \in \mathcal{C}(A, \mathcal{K}(S))$ normalized if $i_S \circ r_S \circ \nu = \nu$.

Definition 4.12. (Norm Preserving). We call $f : \mathcal{C}(A, \mathcal{K}(S)) \to \mathcal{C}(A, \mathcal{K}(S))$ norm preserving if for all normalized objects $\nu \in \mathcal{C}(A, \mathcal{K}(S))$ we have that $f(\nu)$ is normalized.

Proposition 4.13. Let $\mathcal{C}$ be a concrete category and let $d : B^2 \to \mathbb{R}_{\geq 0}$ be a complete metric on $B \in \text{ob}(\mathcal{C})$. Then for any object $A \in \text{ob}(\mathcal{C})$, we have $\overline{d} : \mathcal{C}(A, B)^2 \to \mathbb{R}_{\geq 0}$ defined by

$$\overline{d}(f, g) = \sup_{a \in A} d(f(a), g(a))$$

is complete on $\mathcal{C}(A, B)$.

Proof. Take $(f_n)_{n \in \mathbb{N}}$ a Cauchy-sequence on $\mathcal{C}(A, B)$. Since d is complete, we have that f_n pointwise converges to a morphism $f \in \mathcal{C}(A, B)$ under d.

Since $(f_n)_{n \in \mathbb{N}}$ a Cauchy-sequence on $\mathcal{C}(A, B)$, for all $\epsilon > 0$ there exists an $N \in \mathbb{N}$ such that for all $n, m \in \mathbb{N}$ with $n, m > N$ we have

$$\overline{d}(f_n, f_m) = \sup_{a \in A} d(f_n(a), f_m(a)) < \frac{1}{2} \cdot \epsilon.$$

Therefore, for all $a \in A$, we have

$$d(f_n(a), f_m(a)) < \frac{1}{2} \cdot \epsilon.$$

Now let $m \to \infty$. Then we get $f_m(a) \to f(a)$, and therefore

$$d(f_n(a), f(a)) < \epsilon.$$

By taking the supremum over this expression, we get

$$\overline{d}(f_n, f) = \sup_{a \in A} d(f_n(a), f(a)) \leq \frac{1}{2} \cdot \epsilon < \epsilon.$$

Therefore, $(f_n)_{n \in \mathbb{N}}$ converges to f under $\overline{d}$. □

Now that we have stated all required assumptions, we will state and prove the limit theorem for general coalgebra. We begin by proving some preliminary lemmas.

Lemma 4.14. For all $f \in \mathcal{C}(A, R)$ with $f \sqsubseteq e$ we have that d_f^A is a pseudometric on $\mathcal{C}(A, GS)$. Meaning it satisfies the triangle inequality, symmetry and the null-distance property.

Proof. **null-distance property** For $\delta \in \mathcal{C}(A, \mathcal{K}S)$ we have

$$d_f^A(\delta, \delta) = d_R^A(E_f \circ \delta, E_f \circ \delta) = 0.$$

symmetry
For $\delta, \nu \in \mathcal{C}(A, \mathcal{K}S)$ we have

$$d_f^A(\delta, \nu) = d_R^A(E_f \circ \delta, E_f \circ \nu) = d_R^A(E_f \circ \nu, E_f \circ \delta) = d_f^A(\nu, \delta).$$

triangle inequality
For $\psi, \nu, \delta \in \mathcal{C}(A, \mathcal{K}S)$ we have

$$d_f^A(\psi, \delta) = d_R^A(E_f \circ \psi, E_f \circ \delta).$$

$$\leq d_R^A(E_f \circ \psi, E_f \circ \nu) + d_R^A(E_f \circ \nu, E_f \circ \delta) = d_f^A(\psi, \nu) + d_f^A(\nu, \delta).$$

□

Lemma 4.15. Assume for all $\delta, \nu \in \mathcal{C}(A, \mathcal{K}S)$ there exist some $f \in \mathcal{C}(A, R)$ with $f \sqsubseteq e$ such that $E_f \circ \delta \neq E_f \circ \nu$. Then $\overline{d}_S^A$ is a metric on $\mathcal{C}(A, \mathcal{K}S)$.

Proof. In order to prove this statement, we will need to prove 1) reflexivity 2) symmetry and 3) the triangle inequality. This proof relies heavily on the fact that d_f^A is a pseudo-metric for all $f \sqsubseteq e$.

reflexivity
For $\delta, \nu \in \mathcal{C}(A, \mathcal{K}S)$ with $\delta \neq \nu$ there exists an $f \in \mathcal{C}(A, R)$ with $f \sqsubseteq e$ such that $E_f \circ \delta \neq E_f \circ \nu$. Therefore we have

$$\overline{d}_S^A(\delta, \nu) = \sup_{g \sqsubseteq e} d_g^A(\delta, \nu) \geq d_f^A(\delta, \nu) = d_R^A(E_f \circ \delta, E_f \circ \nu) > 0.$$

Moreover, for all $\delta \in \mathcal{C}(A, \mathcal{K}S)$ we have

$$\overline{d}_S^A(\delta, \delta) = \sup_{g \sqsubseteq e} d_g^A(\delta, \delta) = 0,$$

where the last equality follows from the fact that $d_g^A(\delta,\delta) = 0$ for all $g \sqsubseteq e$.

symmetry
For $\delta, \nu \in \mathcal{C}(A, \mathcal{K}S)$, we have

$$\overline{d}_S^A(\delta,\nu) = \sup_{f \sqsubseteq e} d_f^S(\delta,\nu) = \sup_{f \sqsubseteq e} d_f^S(\nu,\delta) = \overline{d}_S^A(\nu,\delta).$$

triangle inequality
For $\psi, \nu, \delta \in \mathcal{C}(A, \mathcal{K}S)$, we have

$$\begin{aligned}\overline{d}_S^A(\psi,\delta) &= \sup_{f \sqsubseteq e} \overline{d}_f^A(\psi,\delta) \le \sup_{f \sqsubseteq e} \left(d_f^A(\psi,\nu) + d_f^A(\nu,\delta)\right)\\ &\le \left(\sup_{f \sqsubseteq e} d_f^A(\psi,\nu)\right) + \left(\sup_{g \sqsubseteq e} d_g^A(\nu,\delta)\right) = d_S^A(\psi,\nu) + d_S^A(\nu,\delta).\end{aligned}$$

□

Theorem 4.16. (Convergence theorem for coalgebras). If the conditions stated above are satisfied, the sequence $((\hat{T}_c^1)^n(i \circ \nu))_{n\in\mathbb{N}}$ will converge to the unique normalized stationary point $\pi : 1 \to \mathcal{K}S$ for all $\nu : 1 \to FS$. Moreover, we call $r \circ \pi : 1 \to \mathcal{F}S$ the corresponding stationary point in $\mathcal{F}S$.

Proof. Suppose that $\nu, \zeta : 1 \to \mathcal{F}S$ are arbitrarily given and let $\overline{\nu}, \overline{\zeta} : 1 \to \mathcal{K}S$ be the corresponded included mappings defined by $\overline{\zeta} := i_S \circ \zeta$ and $\overline{\nu} := i_S \circ \nu$. Since $r \circ i = \mathrm{id}_{\mathcal{F}}$, we have that $\overline{\zeta}$ and $\overline{\nu}$ are normalized.

Due to condition **9.** and **10.**, there exist a coalgebra $k : S \to \mathcal{K}(S)$ with transition function $\hat{T}_k^1 : \mathcal{C}(1, \mathcal{K}(S)) \to \mathcal{C}(1, \mathcal{K}(S))$ such that

$$\hat{T}_c^1(\overline{\nu}) = \odot_{\mathcal{K}}^1(\alpha, \delta, \hat{T}_k^1(\overline{\nu})) \qquad \hat{T}_c^1(\overline{\zeta}) = \odot_{\mathcal{K}}^1(\alpha, \delta, \hat{T}_k^1(\overline{\zeta})).$$

Moreover, using the remaining conditions, we can show that $\hat{T}_c^1$ is a contraction of $\overline{d}_S^1$:

$$\begin{aligned}
\overline{d}_S^1(\hat{T}(\overline{\nu}), \hat{T}(\overline{\zeta})) &= \sup_{f \sqsubseteq e} d_f^1(\hat{T}_c^1(\overline{\nu}), \hat{T}_c^1(\overline{\zeta})) && (\text{Definition of } \overline{d}_S^1)\\
&= \sup_{f \sqsubseteq e} d_R^1(E_f \circ (\hat{T}_c^1(\overline{\nu})), E_f \circ (\hat{T}_c^1(\overline{\zeta}))) && (\text{Definition of } \overline{d}_R^1)\\
&= \sup_{f \sqsubseteq e} d_R^1(E_f \circ \odot(\alpha, \delta, \hat{T}_k^1(\overline{\nu})), E_f \circ \odot(\alpha, \delta, \hat{T}_k^1(\overline{\zeta}))) && (\text{Condition } \mathbf{9.} \text{ and } \mathbf{10.})\\
&= \sup_{f \sqsubseteq e} d_R^1(\odot_R^1(\alpha, E_f \circ \delta, E_f \circ (\hat{T}_k^1(\overline{\nu}))), \odot_R^1(\alpha, E_f \circ \delta, E_f \circ (\hat{T}_k^1(\overline{\zeta})))) && (\text{Condition } \mathbf{11.})\\
&= \sup_{f \sqsubseteq e} (1-\alpha) \cdot d_R^1(E_f \circ (\hat{T}_k^1(\overline{\nu})), E_f \circ (\hat{T}_k^1(\overline{\zeta}))) && (\text{Condition } \mathbf{12.})\\
&= (1-\alpha) \cdot \sup_{f \sqsubseteq e} d_R^1(E_f \circ (\hat{T}_k^1(\overline{\nu})), E_f \circ (\hat{T}_k^1(\overline{\zeta}))) && (\text{Property of supremum})\\
&= (1-\alpha) \cdot \overline{d}_S^1(\hat{T}_k^1(\overline{\nu}), \hat{T}_k^1(\overline{\zeta})) && (\text{Definition of } \overline{d}_S^1)\\
&\le (1-\alpha) \cdot \overline{d}_S^1(\overline{\nu}, \overline{\zeta}). && (\text{Condition } \mathbf{13.})
\end{aligned}$$

Since we assume $\overline{d}_S^1$ is complete, we can apply the Banach Fixed point theorem. Therefore, $((\hat{T}_c^1)^n(i \circ \nu))_{n\in\mathbb{N}}$ converges to a unique fixed point $\pi \in \mathcal{C}(1, \mathcal{K}S)$ for any distribution $\nu \in \mathcal{C}(1, \mathcal{K}S)$.

Moreover, since $r_S \circ i_S \circ \nu = \nu$, we have that $i_S \circ \nu$ is normalized. Since $\hat{T}_c^1$ is norm-preserving, $(\hat{T}_c^1)^n(i_S \circ \nu)$ is normalized for all $n \in \mathbb{N}$. Therefore π is also normalized.

□

4.3 Continuous-time Markov chain

A continuous-time Markov chain $(X_t)_{t\in\mathbb{R}_{\geq 0}}$, similarly to the discrete-time Markov chain is a system that consists of a state space S and probabilistic transitions between these states. The transitions satisfy the Markov property, meaning that the transition probabilities for the next states are invariant of the previously visited states. In this thesis, we will only be looking at time-homogeneous Markov chains, meaning that the transitions also do not depend on the specific time they occur on. The continuous-time aspect means that the transitions also take into account the time the chain spends in a certain state. For these transitions, we use the semi-group $(P_t)_{t\in\mathbb{R}_{\geq 0}}$, with $P_t(x, y) = \mathbb{P}(X_t = y \mid X_0 = x)$, and where the multiplication is defined as $P_t \cdot P_s = P_{t+s}$. In this article, we restrict ourself to transitions with countable support. Meaning, for $x \in S$ and $t \in \mathbb{R}_{\geq 0}$, we have $\{y \in S \mid P_t(x, y) > 0\}$ is countable.

For continuous-time Markov chains, the algebraic multiplication of the transition functions correspond to the Chapman-Kolmogorov equations. For all $x, y \in S$ and $t \in \mathbb{R}_{\geq 0}$, we have

$$P_{t+s}(x, y) = \sum_{k:P_t(x,k)>0} P_t(x, k) \cdot P_s(k, y). \tag{1}$$

Most articles require the Markov chain to have time-homogeneous leaving times. In this article, we consider a more general model, where time-homogeneous leaving times are not required.

To model the continoues-time Markov chain as a coalgebra, we will be using a modified version of the evolution functor ([10], definition 2.12), equipped with the reader monad. We combine the reader monad with the partial distribution monad $(\mathbb{D}_{\leq 1}, \mu^{\mathbb{D}}, \eta^{\mathbb{D}})$ using a distributive law ([9], chapter 5).

Using the coalgebraic modeling of a continuous-time Markov chain, we will go prove that all conditions in Subsection 4.2 hold, after which we will apply Theorem 4.16.

Proposition 4.17. (evolution functor)
Consider $V : \mathbf{Set} \to \mathbf{Set}$ with

$$V(X) = \{f : \mathbb{R}_{\geq 0} \to X\} \quad \text{for } X \in \mathbf{Set}.$$

$$V(f)(e) = f \circ e \quad \text{for } f : A \to B \text{ and } e \in V(A).$$

Then V is an endofunctor on the category **Set**.

Proof. **Conservation of identity**
For $X \in \mathbf{Set}$, $e \in V(X)$ and $t \in \mathbb{R}_{\geq 0}$ we have

$$V(\mathrm{id}_X)(e)(t) = \mathrm{id}_X(e(t)) = e(t).$$

Conservation of composition
For $f : A \to B$, $g : B \to C$ and for $e \in V(A)$ we have

$$V(g) \circ V(f)(e) = V(g)(f \circ e) = g \circ f \circ e = V(g \circ f)(e).$$

$\square$

Definition 4.18. (Reader Monad). Consider the monad (V, μ^R, η^R) with $\mu^R : V^2 \Rightarrow V$ and $\eta^R : \mathrm{id}_{\mathbf{Set}} \Rightarrow V$ with

$$\mu^R(e)(t) = e(t)(t).$$
$$\eta^R(x)(t) = x.$$

This monad is called a reader monad.

We create a distributive law $\lambda : \mathbb{D}_{\leq 1} V \Rightarrow V\mathbb{D}_{\leq 1}$ between the partial distribution monad $(\mathbb{D}_{\leq 1}, \mu^{\mathbb{D}}, \eta^{\mathbb{D}})$, which is identical to the distribution monad from Theorem 3.26, and the reader monad (V, μ^R, η^R):

$$\lambda_X(\nu)(t)(x) = \sum_{e \in \mathrm{supp}(\nu)\ :\ e(t)=x} \nu(e).$$

In appendix C, we prove that λ is a distributive law between $\mathbb{D}_{\leq 1}$ and V.

Using distributive law λ, we can define monad $(V\mathbb{D}_{\leq 1}, \mu_\lambda, \eta_\lambda)$ as follows:

$$\eta^\lambda_X = \eta^R_{\mathbb{D}_{\leq 1}(X)} \circ \eta^{\mathbb{D}}_X \qquad \mu^\lambda_X = \mu^R_{\mathbb{D}_{\leq 1}X} \circ V^2(\mu^{\mathbb{D}}_X) \circ V(\lambda_{\mathbb{D}_{\leq 1}(X)}).$$

By writing out these compositions, we get

$$\eta^\lambda_X(x)(t)(y) = \eta^R_{\mathbb{D}_{\leq 1}(X)}(\eta^{\mathbb{D}}_X(x))(t)(y) = \begin{cases} 1 & \text{if } y = x \\ 0 & \text{otherwise} \end{cases}.$$

$$\begin{aligned}
\mu^\lambda_X(e)(t)(x) &= \mu^R_{\mathbb{D}_{\leq 1}X} \circ V^2(\mu^{\mathbb{D}}_X) \circ V(\lambda_{\mathbb{D}_{\leq 1}(X)})(e)(t)(x) \\
&= \mu^R_{\mathbb{D}_{\leq 1}X}(V^2(\mu^{\mathbb{D}}_X)(V(\lambda_{\mathbb{D}_{\leq 1}(X)})(e)))(t)(x) = V^2(\mu^{\mathbb{D}}_X)(V(\lambda_{\mathbb{D}_{\leq 1}(X)})(e))(t)(t)(x) \\
&= \mu^{\mathbb{D}}_X(V(\lambda_{\mathbb{D}_{\leq 1}(X)})(e)(t)(t))(x) = \mu^{\mathbb{D}}_X(\lambda_{\mathbb{D}_{\leq 1}(X)}(e(t))(t))(x) \\
&= \sum_{\phi \in \mathrm{supp}(\lambda_{\mathbb{D}_{\leq 1}(X)}(e(t))(t))} \lambda_{\mathbb{D}_{\leq 1}(X)}(e(t))(t))(\phi) \cdot \phi(x) \\
&= \sum_{\phi \in \mathrm{supp}(\lambda_{\mathbb{D}_{\leq 1}(X)}(e(t))(t))} \ \sum_{\overline{e} \in \mathrm{supp}(e(t))\ :\ \overline{e}(t)=\phi} e(t)(\overline{e}) \cdot \phi(x) = \sum_{\overline{e} \in \mathrm{supp}(e(t))} e(t)(\overline{e}) \cdot \overline{e}(t)(x).
\end{aligned}$$

Now that we have defined a monad on $V\mathbb{D}_{\leq 1}$, we will model a continuous-time Markov chain as a coalgebra c on $V\mathbb{D}_{\leq 1}$ by modeling the transitions $(P_t)_{t \in \mathbb{R}}$.

Definition 4.19. Let $(X_t)_{t\in\mathbb{R}_{\geq 0}}$ be a continuous-time Markov chain on state space S, and let $(P_t)_{t\in\mathbb{R}_{\geq 0}}$ be the corresponding transitions. We call $c : S \to V\mathbb{D}(S)$ with $c(x)(t)(y) = P_t(x, y)$ the coalgebraic representation of $(X_t)_{t\in\mathbb{R}_{\geq 0}}$.

Since we require $(P_t)_{t\in\mathbb{R}}$ to be a semi-group, coalgebra c should abide by two conditions. Firstly, we want to have the property that, at time 0, the continuous time system $c(x)$ acts as an identity function. Secondly, we require that, evolving the system for time s, and subsequently for time t, is the same as evolving the system for time $s + t$.

Proposition 4.20. Consider a continuous-time Markov chain $(X_t)_{t\in\mathbb{R}_{\geq 0}}$ with transitions $(P_t)_{t\geq 0}$, and let $c : S \to V\mathbb{D}_{\leq 1}S$ be the coalgebraic representation of the Markov chain.
Consider the morphism $\hat{c} : \mathbb{R}_{\geq 0} \to (S \to \mathbb{D}_{\leq 1}S)$ as $\hat{c}(t)(x)(y) = c(x)(t)(y)$. Then the following statements hold for all $x, y \in S$ and all $t, s \in \mathbb{R}_{\geq 0}$.

$$c(x)(0)(y) = \delta_x(y) = \begin{cases} 1 & \text{if } y = x \\ 0 & \text{if } y \neq x \end{cases}. \tag{5.1}$$

$$\hat{c}(t + s) = \hat{c}(t) \mathbin{\#_{\mathbb{D}}} \hat{c}(s). \tag{5.2}$$

Here $\#_{\mathbb{D}}$ is the Kleisli composition defined in Definition 3.21.

Proof. **Proof of statement 5.1**
For all $x, y \in S$ we have

$$c(x)(0)(y) = P_0(x, y) = \mathbb{P}(X_0 = y \mid X_0 = x) = \delta_x(y) = \begin{cases} 1 & \text{if } y = x \\ 0 & \text{if } y \neq x \end{cases}.$$

Proof of statement 5.2
For $x, y \in S$ and $t, s \in \mathbb{R}_{\geq 0}$, we have

$$(\hat{c}(t)\#_{\mathbb{D}}\hat{c}(s))(x)(y) = \mu(\mathbb{D}(\hat{c}(t))(\hat{c}(s)(x)))(y) = \sum_{\gamma\in\text{supp}(\mathbb{D}(\hat{c}(t))(\hat{c}(s)(x))} \mathbb{D}(\hat{c}(t))(\hat{c}(s)(x))(\gamma) \cdot \gamma(y).$$

Here $\mathbb{D}(\hat{c}(t)) : \mathbb{D}(S) \to \mathbb{D}^2(S)$ such that for $\gamma \in \mathbb{D}^2(S)$, we have

$$\mathbb{D}(\hat{c}(t))(\hat{c}(s)(x))(\gamma) = \sum_{k\in\text{supp}(\hat{c}(s)(x)):\hat{c}(t)(k)=\gamma} \hat{c}(s)(x)(k).$$

Therefore, $\gamma \in \text{supp}(\mathbb{D}(\hat{c}(t))(\hat{c}(s)(x))$ if there exists a $k \in \text{supp}(\hat{c}(s)(x))$ with $\hat{c}(t)(k) = \gamma$. And thus, we can rewrite the expression above to

$$(\hat{c}(t) \mathbin{\#_{\mathbb{D}}} \hat{c}(s))(x)(y) = \sum_{\gamma\in\text{supp}(\mathbb{D}(\hat{c}(t))(\hat{c}(s)(x))} \mathbb{D}(\hat{c}(t))(\hat{c}(s)(x))(\gamma) \cdot \gamma(y)$$

$$= \sum_{\gamma\in\text{supp}(\mathbb{D}(\hat{c}(t))(\hat{c}(s)(x))} \left(\sum_{k\in\text{supp}(\hat{c}(s)(x)):\hat{c}(t)(k)=\gamma} \hat{c}(s)(x)(k) \cdot \gamma(y) \right)$$

$$= \sum_{\gamma \in \mathrm{supp}(\mathbb{D}(\hat{c}(t))(\hat{c}(s)(x))} \left(\sum_{k \in \mathrm{supp}(\hat{c}(s)(x)) : \hat{c}(t)(k) = \gamma} \hat{c}(s)(x)(k) \cdot \hat{c}(t)(k)(y) \right)$$

$$= \sum_{k \in \mathrm{supp}(\hat{c}(s)(x))} \hat{c}(s)(x)(k) \cdot \hat{c}(t)(k)(y) = \sum_{k \in S : P_s(x,k) > 0} P_s(x,k) \cdot P_t(k,y)$$

$$= P_{t+s}(x,y) = \hat{c}(t+s)(x)(y)$$

□

We have defined a coalgebra that can model the continuous-time Markov chain. In order to apply Theorem 4.16, we will need to construct some concepts needed for the requirements of the general limit theorem. We will start with the transition function.

Consider the inclusion function $i : V\mathbb{D} \Rightarrow V\mathbb{D}_{\leq 1}$ which maps each full distributions to the corresponding partial distributions, and $r : V\mathbb{D}_{\leq 1} \Rightarrow V\mathbb{D}$ which rescales each partial distribution ν by multiplying with $\frac{1}{\sum\limits_{x \in \mathrm{supp}(\nu)} \nu(x)}$.

Consider $T_c = (i_S \circ c)^{\#_{V\mathbb{D}}} : V\mathbb{D}_{\leq 1}(S) \to V\mathbb{D}_{\leq 1}(S)$, the Kleisli expansion of $i \circ c$ on $(V\mathbb{D}, \mu^\lambda, \eta^\lambda)$. For $e \in V\mathbb{D}_{\leq 1}(S)$, $t \in \mathbb{R}_{\geq 0}$ and $x \in S$ we have

$$T_c(e)(t)(x) = (i_S \circ c)^{\#_{V\mathbb{D}}}(e)(t)(x) = \mu^\lambda_X \circ V\mathbb{D}_{\leq 1}(i_S \circ c)(e)(t)(x) = \mu^\lambda_X(V\mathbb{D}_{\leq 1}(i_S \circ c)(e))(t)(x)$$

$$= \sum_{\overline{e} \in \mathrm{supp}(V\mathbb{D}_{\leq 1}(i_S \circ c)(e)(t))} V\mathbb{D}_{\leq 1}(i_S \circ c)(e)(t)(\overline{e}) \cdot \overline{e}(t)(x) = \sum_{\overline{e} \in \mathrm{supp}(\mathbb{D}_{\leq 1}(i_S \circ c)(e(t)))} \mathbb{D}_{\leq 1}(i_S \circ c)(e(t))(\overline{e}) \cdot \overline{e}(t)(x).$$

By the definition of $\mathbb{D}_{\leq 1}$, we have

$$\mathbb{D}_{\leq 1}(i_S \circ c)(e(t))(\overline{e}) = \sum_{y \in \mathrm{supp}(e(t)) : i_S \circ c(y) = \overline{e}} e(t)(y).$$

By substituting the expression of $\mathbb{D}_{\leq 1}(i_S \circ c)(e(t))(\overline{e})$ in the expression of $T_c(e)(t)(x)$, we get

$$T_c(e)(t)(x) == \sum_{\overline{e} \in \mathrm{supp}(\mathbb{D}_{\leq 1}(i_S \circ c)(e(t)))} \sum_{y \in \mathrm{supp}(e(t)) : i \circ c(y) = \overline{e}} e(t)(y) \cdot (i_S \circ c)(y)(t)(x)$$

$$= \sum_{y \in \mathrm{supp}(e(t))} e(t)(y) \cdot (i_S \circ c)(y)(t)(x).$$

Since the inclusion transformation i does not alter the output of the function, we have the equality $(i_X \circ c)(x) = c(x)$ for all $x \in X$. Therefore, we can simplify our equation to

$$T_c(e)(t)(x) = \sum_{y \in \mathrm{supp}(e(t))} e(t)(y) \cdot c(y)(t)(x).$$

Note that, when we sum over the elements in the support of $T_c(e)(t)(x)$, we get

$$\sum_{x \in \mathrm{supp}(T_c(e)(t)(x))} T_c(e)(t)(x) = \sum_{x \in \mathrm{supp}(T_c(e)(t)(x))} \sum_{y \in \mathrm{supp}(e(t))} e(t)(y) \cdot c(y)(t)(x)$$

$$= \sum_{y \in \mathrm{supp}(e(t))} \sum_{x \in \mathrm{supp}(T_c(e)(t)(x))} e(t)(y) \cdot c(y)(t)(x) = \sum_{y \in \mathrm{supp}(e(t))} e(t)(y) \sum_{x \in \mathrm{supp}(T_c(e)(t)(x))} c(x)(t)(y) = 1$$

Here we can interchange the sums due to Tonelli's theorem for interchanging summations with non-negative terms.

Let $e \in V\mathbb{D}_{\leq 1}(S)$ be normalized. Meaning, for all $t \in \mathbb{R}_{\geq 0}$, $e(t)$ is a partial distribution where the sum of probabilities is equal to 1. Since the sum of probabilities of $T_c(e)(t)$ is also 1 for all $t \in \mathbb{R}_{\geq 0}$, T_c is norm-preserving.

Lastly, note that, when we take $e = i \circ c(z)$, we get

$$T_c(i \circ c(z))(t)(x) = \sum_{y \in \mathrm{supp}(c(z)(t))} c(z)(t)(y) \cdot c(y)(t)(x) = c(z)(2t)(x).$$

The last step follows from the Chapman-Kolmogorov equation.

Partial order on Set$(A, V\mathbb{D}_{\leq 1}(B))$
Consider $f, g \in \mathbf{Set}(A, V\mathbb{D}_{\leq 1}B)$. We define $\leq_{V\mathbb{D}}$ on $\mathbf{Set}(A, V\mathbb{D}_{\leq 1}B)$ as follows:

$$f \leq_{V\mathbb{D}} g \Leftrightarrow f(a)(t)(b) \leq g(a)(t)(b) \quad \forall a \in A, \forall t \in \mathbb{R}_{>0}, \forall b \in B$$

with $\leq$ the standard order on $\mathbb{R}$.

Partial order on Set(A, R)
Let $R = V([0,1])$ and consider $\sqsubseteq$ the partial order on $\mathbf{Set}(A, R)$ with

$$f \sqsubseteq g \Leftrightarrow f(a)(t) \leq g(a)(t) \quad \forall a \in A,\ \forall t \in \mathbb{R}_{\geq 0}.$$

Pseudometric on $V\mathbb{D}_{\leq 1}(S)$
Consider $e : S \to V([0,1])$ with $e(x)(t) = 1$ for all $t \in \mathbb{R}_{\geq 0}$. We define $\overline{E} : V\mathbb{D}_{\leq 1}(R) \to R$ as

$$\overline{E}(\psi)(t) = \sum_{r \in \mathrm{supp}(\psi)} \psi(t)(r) \cdot r(t).$$

Moreover, for $f : S \to R$ with $f \sqsubseteq e$ we define $\overline{E}_f = \overline{E} \circ V\mathbb{D}_{\leq 1}(f)$. For $\delta \in V\mathbb{D}_{\leq 1}(S)$ and $t \in \mathbb{R}_{\geq 0}$, we have

$$\overline{E}_f(\delta)(t) = \sum_{s \in \mathrm{supp}(\delta(t))} \delta(t)(s) \cdot f(s)(t) = E_{f(-)(t)}(\delta(t))$$

with $E_{f(-)(t)}$ the expected value defined in Subsection 4.1.

We define the metric d_R on R with $R = V([0,1])$ as $d_R(f, g) = \sup_{t \geq 0} |f(t) - g(t)|$ for all $f, g \in R$. Note that, since $f, g : \mathbb{R}_{\geq 0} \to [0,1]$, we have $|f(t) - g(t)| \leq 1$ for all $t \in \mathbb{R}_{\geq 0}$. Therefore, the supremum exists and is bounded by 1.

Now that we have defined d_R and $\overline{E}_f$ for $f \sqsubseteq e$, we define the pseudometric d_f on $V\mathbb{D}_{\leq 1}S$ by

$$d_f(\delta, \nu) = d_R(\overline{E}_f(\delta), \overline{E}_f(\nu)).$$

Note that, since $d_R(\overline{E}_f(\delta), \overline{E}_f(\nu))$ is bounded by 1 for all $f \sqsubseteq e$, we have that $d_f(\delta, \nu)$ is bounded by 1 for all $\delta, \nu \in V\mathbb{D}_{\leq 1}(S)$.

Metric on $V\mathbb{D}_{\leq 1}(S)$
We define $\hat{d}$ on $V\mathbb{D}_{\leq 1}(S)$ as

$$\hat{d}(\nu, \delta) = \sup_{f \sqsubseteq e} d_f(\nu, \delta).$$

Note that, since $d_f(\nu, \delta)$ is bounded by 1 for all $f \sqsubseteq e$, $\sup_{f \sqsubseteq e} d_f(\nu, \delta)$ exists and is also bounded by 1.

Completeness of $\hat{d}$
For any $\delta, \nu \in V\mathbb{D}_{\leq 1}(S)$, we have

$$\hat{d}(\nu, \delta) = \sup_{f \sqsubseteq e} \sup_{t \geq 0} \left|\overline{E}_f(\nu)(t) - \overline{E}_f(\delta)(t)\right| = \sup_{t \geq 0} \sup_{f \sqsubseteq e} \left|\overline{E}_f(\nu)(t) - \overline{E}_f(\delta)(t)\right|$$

$$= \sup_{t \geq 0} \sup_{f \sqsubseteq e} \left|E_{f(-)(t)}(\nu(t)) - E^t_{f(-)(t)}(\delta(t))\right| = \sup_{t \geq 0} \sup_{g \sqsubseteq e} |E_g(\nu(t)) - E_g(\delta(t))|$$

$$= \sup_{t \geq 0} \overline{d}(\nu(t), \delta(t))$$

Here $\overline{d}$ is the metric defined in Proposition 4.1. Since $\overline{d}$ is a complete metric, and since we have the equality $\hat{d}(\nu, \delta) = \sup_{t \in \mathbb{R}_{\geq 0}} \overline{d}(\nu(t), \delta(t))$ for all $\nu, \delta : \mathbb{R}_{\geq 0} \to \mathbb{D}_{\leq 1}(S)$, by Proposition 4.13, $\hat{d}$ is complete.

Proposition 4.21. (non-expansiveness of T_k) The the transition function T_k of an arbitrary coalgebra $k : S \to V\mathbb{D}_{\leq 1}(S)$ is non-expansive under metric $\hat{d}$. Meaning, for $e_1, e_2 \in V\mathbb{D}_{\geq 1}(S)$ we have

$$\hat{d}(T_k(e_1), T_k(e_2)) \leq \hat{d}(e_1, e_2).$$

Proof. Let $e_1, e_2 \in V\mathbb{D}_{\leq 1}(S)$ be arbitrarily chosen. We have

$$\hat{d}(T_k(e_1), T_k(e_2)) = \mathrm{supp}_{f \sqsubseteq e} \mathrm{supp}_{t \geq 0} \left|\overline{E}_f(T_k(e_1))(t) - \overline{E}_f(T_k(e_2))(t)\right|.$$

We calculate $\overline{E}_f(T_k(e))(t)$ for $e \in V\mathbb{D}_{\leq 1}(S)$ and $t \in \mathbb{R}_{\geq 0}$:

$$\overline{E}_f(T_k(e))(t) = \sum_{s \in \mathrm{supp}(T_k(e(t))(t))} f(t)(s) \cdot T_k(e(t))(t)(s)$$

$$= \sum_{s \in \mathrm{supp}(T_k(e(t))(t))} f(t)(s) \left(\sum_{y \in \mathrm{supp}(e(t))} e(t)(y) \cdot c(y)(t)(s) \right)$$

$$= \sum_{y \in \mathrm{supp}(e(t))} e(t)(y) \left(\sum_{s \in \mathrm{supp}(T_k(e(t))(t))} c(y)(t)(s) \cdot f(t)(s) \right).$$

Here we interchange the summations using Tonelli's theorem for interchanging summations with non-negative terms.

Since $c(y)(t)(s) \cdot f(t)(s) > 0$ implies that $c(y)(t)(s) > 0$, we can rewrite the equation above to

$$\overline{E}_f(T_k(e))(t) = \sum_{y \in \text{supp}(e(t))} e(t)(y) \left(\sum_{s \in \text{supp}(c(y)(t))} c(y)(t)(s) \cdot f(t)(s) \right).$$

Take $g_f(t)(y) = \sum_{s \in \text{supp}(c(y)(t))} c(y)(t)(s) \cdot f(t)(s)$ for all $s \in S$. Since $f(t)(s) \leq 1$ for all $s \in S$, and since $c(y)(t)$ is a distribution over S, for all $y \in S$ we have

$$g_f(t)(y) = \sum_{s \in \text{supp}(c(y)(t))} c(y)(t)(s) \cdot f(t)(s) \leq \sum_{s \in \text{supp}(c(y)(t))} c(y)(t)(s) = 1.$$

Therefore $g_f \sqsubseteq e$. By substituting $g_f(t)(y) = \sum\limits_{s \in \text{supp}(c(y)(t))} c(y)(t)(s) \cdot f(t)(s)$ in the expression of $\overline{E}_f(T_k(e))(t)$, we get

$$\overline{E}_f(T_k(e))(t) = \sum_{y \in \text{supp}(e(t))} e(t)(y) \cdot g_f(t)(y) = \overline{E}_{g_f}(e)(t).$$

Since this holds for all $e \in V\mathbb{D}_{\leq 1}(S)$, we have

$$\hat{d}(T_k(e_1), T_k(e_2)) = \text{supp}_{t \geq 0} \text{supp}_{f \sqsubseteq e} \left| \overline{E}_f(T_k(e_1))(t) - \overline{E}_f(T_k(e_2))(t) \right|$$

$$= \text{supp}_{t \geq 0} \text{supp}_{f \sqsubseteq e} \left| \overline{E}_{g_f}(e_1)(t) - \overline{E}_{g_f}(e_2)(t) \right| \leq \text{supp}_{t \geq 0} \text{supp}_{h \sqsubseteq e} \left| \overline{E}_h(e_1)(t) - \overline{E}_h(e_2)(t) \right| = \hat{d}(e_1, e_2).$$

$\square$

Convex combinations on $\mathbf{Set}(A, R)$ and $\mathbf{Set}(A, V\mathbb{D}_{\leq 1}(S))$.
In order to apply Theorem 4.16, we need to introduce a convex combination operator on $\mathbf{Set}(A, R)$ and $\mathbf{Set}(A, V\mathbb{D}_{\leq 1}(S))$, as well as a rescaling operator on $\mathbf{Set}(A, V\mathbb{D}_{\leq 1}(S))$. We will start with the operators on $\mathbf{Set}(A, V\mathbb{D}_{\leq 1}(S))$.

$$\otimes^A_{V\mathbb{D}} : [0,1] \times \mathbf{Set}(A, V\mathbb{D}_{\leq 1}(S)) \to \mathbf{Set}(A, V\mathbb{D}_{\leq 1}(S))$$
given by $(\alpha \otimes^A_{V\mathbb{D}} e)(a)(t)(x) = \alpha \cdot e(a)(t)(x)$.

$$\odot^A_{V\mathbb{D}} : [0,1] \times \mathbf{Set}(A, V\mathbb{D}_{\leq 1}(S)) \to \mathbf{Set}(A, V\mathbb{D}_{\leq 1}(S))$$
given by $\odot^A_{V\mathbb{D}} (\alpha, e_1, e_2)(a)(t)(x) = \alpha \cdot e_1(a)(t)(x) + (1 - \alpha) \cdot e_2(a)(t)(x)$.

In order to apply Theorem 4.16, we need to verify that the following lemma holds.

Lemma 4.22. Let $c : S \to V\mathbb{D}_{\leq 1}(S)$ be a coalgebra and let $\hat{T}^A_c : \mathbf{Set}(A, V\mathbb{D}_{\leq 1}(S)) \to \mathbf{Set}(A, V\mathbb{D}_{\leq 1}(S))$ be the corresponding transition function. Suppose there exists $\alpha \in (0,1)$ and a normalized morphism $e_2 \in \mathbf{Set}(A, V\mathbb{D}_{\leq 1}(S))$ such that for all normalized morphisms $e_1 \in \mathbf{Set}(A, V\mathbb{D}_{\leq 1}(S))$ we have $\hat{T}^A_c(e_1) \geq_{V\mathbb{D}} \alpha \otimes^A_{V\mathbb{D}} e_2$. Then there exists a coalgebra $k : S \to V\mathbb{D}_{\leq 1}(S)$ such that

$$\hat{T}^A_c(e_1) = \odot^A_{V\mathbb{D}}(\alpha, e_2, \hat{T}^A_k(e_1))$$

for all normalized morphisms $e_1 \in \mathbf{Set}(A, V\mathbb{D}_{\leq 1}(S))$.

Proof. Consider $k : S \to (\mathbb{R}_{\geq 0} \to (S \to \mathbb{R}))$ defined by

$$k(x)(t)(y) = \frac{1}{1-\alpha} \cdot c(x)(t)(y) - \frac{\alpha}{1-\alpha} \cdot e_2(t)(y).$$

By taking fixed $x \in S$ and $t \in \mathbb{R}_{\geq 0}$, and summing over $y \in S$, we get

$$\sum_{y \in \mathrm{supp}(e_2(t)) \cup \mathrm{supp}(c(x)(t))} k(x)(t)(y) = \sum_{y \in \mathrm{supp}(e_2(t)) \cup \mathrm{supp}(c(x)(t))} \frac{1}{1-\alpha} \cdot c(x)(t)(y) - \frac{\alpha}{1-\alpha} \cdot e_2(t)(y)$$

$$= \frac{1}{1-\alpha} \left(\sum_{y \in \mathrm{supp}(c(x)(t))} c(x)(t)(y) \right) - \frac{\alpha}{1-\alpha} \left(\sum_{y \in \mathrm{supp}(e_2(t))} e_2(t)(y) \right) = \frac{1}{1-\alpha} - \frac{\alpha}{1-\alpha} = 1.$$

Let $x \in S$ and consider $\delta_x^A : \mathbf{Set}(A, V\mathbb{D}_{\leq 1}(S)) \to \mathbf{Set}(A, V\mathbb{D}_{\leq 1}(S))$ with

$$\delta_x^A(a)(t)(y) = \begin{cases} 1 & \text{if } x = y \\ 0 & \text{if } x \neq y \end{cases}$$

for all $a \in A$, $t \in \mathbb{R}_{\geq 0}$ and $y \in S$. Since $\hat{T}_c^A(\delta_x^A) \geq_{V\mathbb{D}} \alpha \otimes_{V\mathbb{D}}^A e_2$, for all $a \in A$, all $t \in \mathbb{R}_{\geq 0}$ and all $y \in S$ we have

$$\hat{T}_c^A(\delta_x^A)(a)(t)(y) = \sum_{z \in \mathrm{supp}(\delta_x^A(a)(t))} \delta_x^A(a)(t)(z) \cdot c(z)(t)(x) = c(x)(t)(y) \geq \alpha \cdot e_2(t)(y).$$

Since this holds for all $t \in \mathbb{R}_{\geq 0}$ and all $x, y \in S$, we have

$$k(x)(t)(y) = \frac{1}{1-\alpha} \cdot c(x)(t)(y) - \frac{\alpha}{1-\alpha} \cdot e_2(t)(y) \geq \frac{\alpha}{1-\alpha} \cdot e_2(t)(y) - \frac{\alpha}{1-\alpha} \cdot e_2(t)(y) = 0$$

for all $x, y \in S$ and all $t \in \mathbb{R}_{\geq 0}$.

We have proven that $k(x)(t)(y) > 0$ for all $x, y \in S$ and all $t \in \mathbb{R}_{\geq 0}$, and we have proven that, for fixed $t \in \mathbb{R}_{\geq 0}$ and $x \in S$, the sum over all $y \in S$ gives 1. Therefore, we have proven that $k(x)(t) \in \mathbb{D}_{\leq 1}(S)$ for all $t \in \mathbb{R}_{\geq 0}$ and all $x \in S$, and thus $k : S \to V\mathbb{D}_{\leq 1}(S)$ is a coalgebra.

Consider the transition function $\hat{T}_k^A : \mathbf{Set}(A, V\mathbb{D}_{\leq 1}(S)) \to \mathbf{Set}(A, V\mathbb{D}_{\leq 1}(S))$ of coalgebra k. For all $a \in A$, all $t \in \mathbb{R}_{\geq 0}$, all $x \in S$ and all $e_1 \in \mathbf{Set}(A, V\mathbb{D}_{\leq 1}(S))$ we have

$$\odot(\alpha, e_2, \hat{T}_k^A(e_1))(a)(t)(x) = \alpha \cdot e_2(a)(t)(x) + (1-\alpha) \cdot \hat{T}_k^A(e_1)(a)(t)(x)$$

$$= \alpha \cdot e_2(a)(t)(x) + (1-\alpha) \left(\sum_{y \in \mathrm{supp}(e_1(a)(t))} e_1(a)(t)(y) \cdot k(y)(t)(x) \right)$$

$$= \alpha \cdot e_2(a)(t)(x) + (1-\alpha) \left(\sum_{y \in \mathrm{supp}(e_1(a)(t))} \frac{1}{1-\alpha} \cdot e_1(a)(t)(y) \cdot c(x)(t)(y) - \frac{\alpha}{1-\alpha} \cdot e_1(a)(t)(y) \cdot e_2(a)(t)(x) \right)$$

$$= \alpha \cdot e_2(a)(t)(x) - \alpha \cdot e_2(a)(t)(x) \cdot \left(\sum_{y \in \mathrm{supp}(e_1(a)(t))} e_1(a)(t)(y) \right) + \left(\sum_{y \in \mathrm{supp}(e_1(a)(t))} e_1(a)(t)(y) \cdot c(x)(t)(y) \right)$$

$$= \alpha \cdot e_2(a)(t)(x) - \alpha \cdot e_2(a)(t)(x) + \hat{T}_c^A(e_1)(a)(t)(x) = \hat{T}_c^A(e_1)(a)(t)(x).$$

□

Lastly, we define the convex combination operator on $\mathbf{Set}(A, R)$.

$$\odot_R^A : [0,1] \times \mathbf{Set}(A,R)^2 \to \mathbf{Set}(A,R)$$
$$\text{given by } \odot_R^A (\alpha, f, g)(a)(t) = \alpha \cdot f(a)(t) + (1-\alpha) \cdot g(a)(t).$$

In order to apply Theorem 4.16 to the continuous-time Markov chain, the convex combination operators on $\mathbf{Set}(A, V\mathbb{D}_{\leq 1}(S))$ and $\mathbf{Set}(A, R)$ need to respect each other when taking expected values.

Lemma 4.23. For all $\alpha \in [0,1]$, all $e_1, e_2 \in \mathbf{Set}(A, V\mathbb{D}_{\leq 1}(S))$ and all $f : S \to R$ with $f \sqsubseteq e$ we have

$$\overline{E}_f \circ (\odot_{V\mathbb{D}}^A(\alpha, e_1, e_2)) = \odot_R^A(\alpha, \overline{E}_f \circ e_1, \overline{E}_f \circ e_2).$$

Proof. Let $\alpha \in [0,1]$, $a \in A$, $t \in \mathbb{R}_{\geq 0}$, $e_1, e_2 \in \mathbf{Set}(A, V\mathbb{D}_{\leq 1}(S))$ and $f : S \to R$ with $f \sqsubseteq e$ be arbitrarily chosen we have

$$\overline{E}_f \circ (\odot_{V\mathbb{D}}^A(\alpha, e_1, e_2))(a)(t) = \overline{E}_f \circ (\odot_{V\mathbb{D}}^A(\alpha, e_1, e_2)(a))(t)$$

$$= \sum_{x \in \mathrm{supp}(\odot_{V\mathbb{D}}^A(\alpha, e_1, e_2)(a)(t))} \odot_{V\mathbb{D}}^A(\alpha, e_1, e_2)(a)(t)(x) \cdot f(x)(t)$$

$$= \sum_{x \in \mathrm{supp}(\odot_{V\mathbb{D}}^A(\alpha, e_1, e_2)(a)(t))} (\alpha \cdot e_1(a)(t)(x) + (1-\alpha) \cdot e_2(a)(t)(x)) \cdot f(x)(t)$$

$$= \alpha \left(\sum_{x \in \mathrm{supp}(e_1(a)(t))} e_1(a)(t)(x) \cdot f(x)(t) \right) + (1-\alpha) \left(\sum_{x \in \mathrm{supp}(e_2(a)(t))} e_2(a)(t)(x) \cdot f(x)(t) \right)$$

$$= \alpha \cdot \overline{E}_f(e_1(a))(t) + (1-\alpha) \cdot \overline{E}_f(e_2(a))(t) = \odot_R^A(\alpha, \overline{E}_f \circ e_1, \overline{E}_f \circ e_2)(a)(t).$$

□

Corollary 4.24. (Limit-point of the continuous-time Markov chain). Assume there exists a normalized morphism $e \in \mathbf{Set}(1, V\mathbb{D}_{\leq 1}(S))$ and an $\alpha \in (0,1)$ such that $T_c(\overline{e}) \geq_{V\mathbb{D}} \alpha \otimes_{V\mathbb{D}}^1 e$ for all normalized morphisms $\overline{e} \in \mathbf{Set}(1, V\mathbb{D}_{\leq 1}(S))$. Using the constructions stated above, we satisfy all conditions stated in Subsection 4.2. Therefore $(\hat{T}_c^1)^n(i \circ \nu)$ converges to a normalized stationary point $\pi \in \mathbf{Set}(1, V\mathbb{D}_{\leq 1}(S))$ for all $\nu \in \mathbf{Set}(1, V\mathbb{D}(S))$.

5 Conclusions and Further Research

This thesis generalized a theorem for finding stationary distributions for discrete-time Markov chains to a limit point theorem for general coalgebras. By determining the assumptions needed for the limit-theorem of the discrete-time Markov chain to hold, we have isolated a set of conditions and structural restrictions of the underlying functor and category under which the limit point theorem for the general coalgebra hold. Since this list is quite extensive, it apparently restricts the scope of the systems on which we can apply the theorem. To demonstrate that the scope of the general limit theorem goes beyond the discrete-time Markov chain, we have used the limit point theorem to prove Corollary 4.24 for finding limit points of a continues-time Markov chain.

Future work. In this thesis, we chose to use the functor $\mathbb{D}$. The functor $\mathbb{D}$ sends each object X to the set of all distributions on X with countable support. Therefore, we restricted our discrete-time Markov chain to have a countable support. As a continuation of this thesis, one could try to recreate the proof and statement of Theorem 4.8 for finding the stationary point of a discrete-time Markov chain where an uncountable support is possible. This could be done by replacing $\mathbb{D}$ and monad $\mu^{\mathbb{D}}$ with the Giry functor and Giry monad from [6]. A more general theorem for the discrete-time Markov chain might result in more refined conditions for the general case.

Currently we use one specific framework to model our metric $\overline{d}_S^A$, which implicitly introduces additional assumptions, such as the existence of E. It would be interesting to investigate a more categorical characterization of our metric. Alternatively, we could investigate whether we could formulate a minimal metric structure which satisfies some conditions with respect to the underlying category. Moreover, since morphisms in a metric structure are non-expansive, we could simplify the condition that a transition function T_k^1 needs to be non-expansive under $\overline{d}_S^A$ to requiring that T_k^1 is a morphism in this metric structure.

Appendices

A Proof : $\mathbb{D} : \mathbf{Set} \to \mathbf{Set}$ is a functor

In order to prove that $\mathbb{D}$ is a functor, we must prove that $\mathbb{D}$ preserves identity and composition.

preservation of identity
Consider $X \in ob(\mathbf{Set})$, and consider the identity morphism $\mathrm{id}_X : X \to X$ on X. For $\nu \in \mathbb{D}X$ and $x \in X$, we have

$$\mathbb{D}(\mathrm{id}_X)(\nu)(x) = \sum_{y \in \mathrm{id}_X(\{x\})} \nu(y) = \nu(x)$$

preservation of composition
Consider $f : A \to B$ and $g : B \to C$. For $\nu \in \mathbb{D}A$ and $x \in C$, we have

$$\mathbb{D}(g \circ f)(\nu)(x) = \sum_{y \in \mathrm{supp}(\nu) : g \circ f(y) = x} \nu(y)$$

Note that $g \circ f(y) = x$ if there exists $z \in B$ such that $f(y) = z$ and $g(z) = x$. Therefore,

$$\mathbb{D}(g \circ f)(\nu)(x) = \sum_{z \in \mathrm{supp}(\mathbb{D}f(\nu)) : g(z) = x} \; \sum_{y \in \mathrm{supp}(\nu) : f(y) = z} \nu(y) = \sum_{z \in \mathrm{supp}(\mathbb{D}f(\nu)) : g(z) = x} \mathbb{D}f(\nu)(z) = \mathbb{D}g(\mathbb{D}f(\nu))(x).$$

B Proof : $(\mathbb{D}, \mu, \eta)$ is a distribution monad

In order to prove that $(\mathbb{D}, \mu, \eta)$ is a monad, we need to prove that the two diagrams in Definition 3.16 commutes. In the section below, we display each diagram, under which we prove that it commutes.

$$\begin{tikzcd}
\mathbb{D}X \arrow[r, "\mathbb{D}(\eta_X)"] \arrow[d, "\eta_{\mathbb{D}(X)}"'] \arrow[dr, "\mathrm{id}_{\mathbb{D}(X)}"] & \mathbb{D}^2X \arrow[d, "\mu_X"] \\
\mathbb{D}^2X \arrow[r, "\mu_X"'] & \mathbb{D}X
\end{tikzcd}$$

We will prove that the diagram commutes, by computing both paths on $\nu \in \mathbb{D}(X)$ and $x \in X$, and showing that it is equal to $\nu(x)$.

$$\mu_X \circ \mathbb{D}(\eta_X)(\nu)(x) = \mu_X(\mathbb{D}(\eta_X)(\nu))(x) = \sum_{\rho \in \mathrm{supp}(\mathbb{D}(\eta_X)(\nu))} \rho(x) \cdot \mathbb{D}(\eta_X)(\nu)(\rho)$$

$$= \sum_{\rho \in \mathrm{supp}(\mathbb{D}(\eta_X)(\nu))} \rho(x) \cdot \sum_{y \in \mathrm{supp}(\nu) : \eta_X(y) = \rho} \nu(y).$$

For all $y \in X$ we have $\eta_X(y) = \delta_y$ with $\delta_y(k) = 1 \iff k = y$. Therefore we have

$$\mu_X \circ \mathbb{D}(\eta_X)(\nu)(x) = \sum_{\rho \in \mathrm{supp}(\mathbb{D}(\eta_X)(\nu))} \rho(x) \cdot \sum_{y \in \mathrm{supp}(\nu) : \rho = \delta_y} \nu(y) = \delta_x(x) \cdot \nu(x) = \nu(x)$$

Note that, in the second to last step, the term of the summation is non-zero if and only if $\rho = \delta_x$. Therefore, we can simplify the equation to $\delta_x(x) \cdot \nu(x)$.

Below, we will compute the second path on $\nu \in \mathbb{D}(X)$ and $x \in X$.

$$\mu_X \circ \eta_{\mathbb{D}(X)}(\nu)(x) = \mu_X(\eta_{\mathbb{D}(X)}(\nu))(x) = \sum_{\rho \in \text{supp}(\eta_{\mathbb{D}(X)}(\nu))} \eta_{\mathbb{D}(X)}(\nu)(\rho) \cdot \rho(x).$$

Since $\eta_{\mathbb{D}(X)}(\nu)(\rho) = 1 \;\Leftrightarrow\; \rho = \nu$, we have

$$\mu_X \circ \eta_{\mathbb{D}(X)}(\nu)(x) = \nu(x).$$

Since both paths computed on ν and x lead to $\nu(x)$, the diagram commutes.

$$\begin{tikzcd}
\mathbb{D}^3 X \arrow[r, "\mu_{\mathbb{D}(X)}"] \arrow[d, "\mathbb{D}(\mu_X)"'] & \mathbb{D}^2 X \arrow[d, "\mu_X"] \\
\mathbb{D}^2 X \arrow[r, "\mu_X"'] & \mathbb{D} X
\end{tikzcd}$$

In order to prove that the diagram commutes, we compute both paths on $\nu \in \mathbb{D}^3(X)$ and $x \in X$.

$$\mu_X \circ \mu_{\mathbb{D}(X)}(\nu)(x) = \mu_X(\mu_{\mathbb{D}(X)}(\nu))(x) = \sum_{\rho \in \text{supp}(\mu_{\mathbb{D}(X)}(\nu))} \rho(x) \cdot \mu_{\mathbb{D}(X)}(\nu)(\rho)$$

$$= \sum_{\rho \in \text{supp}(\mu_{\mathbb{D}(X)}(\nu))} \rho(x) \left(\sum_{\delta \in \text{supp}(\nu)} \delta(\rho) \cdot \nu(\delta) \right) = \sum_{\delta \in \text{supp}(\nu)} \sum_{\rho \in \text{supp}(\mu_{\mathbb{D}(X)}(\nu))} \rho(x) \cdot \delta(\rho) \cdot \nu(\delta).$$

In the last step, we interchanged the summations using Tonelli's theorem for interchanging summations with positive terms. Lastly, we have

$$\mu_{\mathbb{D}(X)}(\nu)(\rho) = \sum_{\delta \in \text{supp}(\nu)} \delta(\rho) \cdot \nu(\delta).$$

Therefore, $\rho \in \text{supp}(\mu_{\mathbb{D}(X)}(\nu)) \subseteq \text{supp}(\delta)$. Using this, we can simplify the expression to

$$\mu_X \circ \mu_{\mathbb{D}(X)}(\nu)(x) = \sum_{\delta \in \text{supp}(\nu)} \sum_{\rho \in \text{supp}(\delta)} \rho(x) \cdot \delta(\rho) \cdot \nu(\delta).$$

Below, we will compute the second path on $\nu \in \mathbb{D}^3(X)$ and $x \in X$.

$$\mu_X \circ \mathbb{D}(\mu_X)(\nu)(x) = \mu_X(\mathbb{D}(\mu_X)(\nu))(x) = \sum_{\phi \in \text{supp}(\mathbb{D}(\mu_X)(\nu))} \phi(x) \cdot \mathbb{D}(\mu_X)(\nu)(\phi)$$

$$= \sum_{\phi \in \text{supp}(\mathbb{D}(\mu_X)(\nu))} \phi(x) \cdot \sum_{\delta \in \text{supp}(\nu) : \mu_X(\delta) = \phi} \nu(\delta) = \sum_{\phi \in \text{supp}(\mathbb{D}(\mu_X)(\nu))} \sum_{\delta \in \text{supp}(\nu) : \mu_X(\delta) = \phi} \nu(\delta) \cdot \phi(x)$$

$$= \sum_{\delta \in \text{supp}(\nu)} \nu(\delta) \cdot \mu_X(\delta)(x) = \sum_{\delta \in \text{supp}(\nu)} \nu(\delta) \left(\sum_{\rho \in \text{supp}(\delta)} \rho(x) \cdot \delta(\rho) \right) = \sum_{\delta \in \text{supp}(\nu)} \sum_{\rho \in \text{supp}(\delta)} \rho(x) \cdot \delta(\rho) \cdot \nu(\delta)$$

Since both paths lead to the same result, the diagram commutes.

C Proof : λ is a distributive law between $\mathbb{D}$ and V

In order to prove that $\lambda : \mathbb{D}_{\leq 1}V \Rightarrow V\mathbb{D}_{\leq 1}$ is a distributive law on $(\mathbb{D}_{\leq 1}, \mu^{\mathbb{D}}, \eta^{\mathbb{D}})$ and (V, μ^R, ν^R), we need to prove that the four diagrams in Definition 3.20 commute. In the section below, we display each diagram, under which we prove that it commutes.

$$\begin{tikzcd}
\mathbb{D}_{\leq 1}(X) \arrow[d, "\mathbb{D}_{\leq 1}(\eta^R_X)"'] \arrow[dr, "\eta^R_{\mathbb{D}_{\leq 1}(X)}"] & \\
\mathbb{D}_{\leq 1}V(X) \arrow[r, "\lambda_X"'] & V\mathbb{D}_{\leq 1}(X)
\end{tikzcd}$$

In order to prove that the diagram commutes, we compute both paths on $\nu \in \mathbb{D}_{\leq 1}(X)$, $t \in \mathbb{R}_{\geq 0}$ and $x \in X$.

$$\lambda_X \circ \mathbb{D}_{\leq 1}(\eta^R_X)(\nu)(t)(x) = \lambda_X(\mathbb{D}_{\leq 1}(\eta^R_X)(\nu))(t)(x) = \sum_{e \in \mathrm{supp}(\mathbb{D}_{\leq 1}(\eta^R_X)(\nu)) : e(t) = x} \mathbb{D}_{\leq 1}(\eta^R_X)(\nu)(e)$$

$$= \sum_{e \in \mathrm{supp}(\mathbb{D}_{\leq 1}(\eta^R_X)(\nu)) : e(t) = x} \quad \sum_{y \in \mathrm{supp}(\nu) : \eta^R_X(y) = e} \nu(y).$$

For all $y \in X$ and all $t \in \mathbb{R}_{\geq 0}$, we have $\eta^R_X(y)(t) = y$. Therefore

$$\lambda_X \circ \mathbb{D}_{\leq 1}(\eta^R_X)(\nu)(t)(x) = \sum_{e \in \mathrm{supp}(\mathbb{D}_{\leq 1}(\eta^R_X)(\nu)) : e(t) = x} \quad \sum_{y \in \mathrm{supp}(\nu) : e(t) = y} \nu(y) = \nu(x).$$

Below, we compute the second path on $\nu \in \mathbb{D}_{\leq 1}(X)$, $t \in \mathbb{R}_{\geq 0}$ and $x \in X$. This computation follows directly from the definition of η^R.

$$\eta^R_{\mathbb{D}_{\leq 1}(X)})(\nu)(t)(x) = \nu(x).$$

Since both path lead to the same result, the diagram commutes.

$$\begin{tikzcd}
V(X) \arrow[d, "\eta^{\mathbb{D}}_{V(X)}"'] \arrow[dr, "V(\eta^{\mathbb{D}}_X)"] & \\
\mathbb{D}_{\leq 1}V(X) \arrow[r, "\lambda_X"'] & V\mathbb{D}_{\leq 1}(X)
\end{tikzcd}$$

In order to prove that the diagram commutes, we compute both paths on $e \in V(X)$, $t \in \mathbb{R}_{\geq 0}$ and $x \in X$.

$$\lambda_X \circ \eta^{\mathbb{D}}_{V(X)}(e)(t)(x) = \lambda_X(\eta^{\mathbb{D}}_{V(X)}(e))(t)(x) = \sum_{\overline{e} \in \mathrm{supp}(\eta^{\mathbb{D}}_{V(X)}(e)) : \overline{e}(t) = x} \eta^{\mathbb{D}}_{V(X)}(e)(\overline{e}).$$

Since $\eta_{V(X)}(e)(\overline{e}) = 1 \iff e = \overline{e}$, and since we sum over all $\overline{e} \in \mathrm{supp}(\eta^{\mathbb{D}}_{V(X)}(e)))$ for which $\overline{e}(t) = x$, we have

$$\lambda_X \circ \eta^{\mathbb{D}}_{V(X)}(e)(t)(x) = \sum_{\overline{e} \in \mathrm{supp}(\eta^{\mathbb{D}}_{V(X)}(e)) : \overline{e}(t) = x} \eta^{\mathbb{D}}_{V(X)}(e)(\overline{e}) = \begin{cases} 1 & \text{if } e(t) = x \\ 0 & \text{if } e(t) \neq x \end{cases}.$$

Below, we compute the second path on $e \in V(X)$, $t \in \mathbb{R}_{\geq 0}$ and $x \in X$.

$$V(\eta_X^{\mathbb{D}})(e)(t)(x) = \eta_X^{\mathbb{D}}(e(t))(x) = \begin{cases} 1 & \text{if } e(t) = x \\ 0 & \text{if } e(t) \neq x \end{cases}$$

Since both paths lead to the same result, the diagram commutes.

$$\begin{tikzcd}
\mathbb{D}_{\leq 1}VV(X) \arrow[r, "\lambda_{V(X)}"] \arrow[d, "\mathbb{D}_{\leq 1}(\mu_X^R)"'] & V\mathbb{D}_{\leq 1}V(X) \arrow[r, "V(\lambda_X)"] & VV\mathbb{D}_{\leq 1}(X) \arrow[d, "\mu^R_{\mathbb{D}_{\leq 1}(X)}"] \\
\mathbb{D}_{\leq 1}V(X) \arrow[rr, "\lambda_X"'] & & V\mathbb{D}_{\leq 1}(X)
\end{tikzcd}$$

In order to prove that the diagram commutes, we compute both paths on $\nu \in \mathbb{D}_{\leq 1}VV(X)$, $t \in \mathbb{R}_{\geq 0}$ and $x \in X$.

$$\mu^R_{\mathbb{D}_{\leq 1}(X)} \circ V(\lambda_X) \circ \lambda_{V(X)}(\nu)(t)(x) = \mu^R_{\mathbb{D}_{\leq 1}(X)}(V(\lambda_X(\lambda_{V(X)}(\nu))))(t)(x) = V(\lambda_X(\lambda_{V(X)}(\nu)))(t)(t)(x)$$

$$= \lambda_X(\lambda_{V(X)}(\nu)(t))(t)(x) = \sum_{e\,\mathrm{supp}(\lambda_{V(X)}(\nu)(t)):e(t)=x} \lambda_{V(X)}(\nu)(t)(e)$$

$$= \sum_{e \in \mathrm{supp}(\lambda_{V(X)}(\nu)(t)):e(t)=x} \; \sum_{\overline{e} \in \mathrm{supp}(\nu):\overline{e}(t)=e} \nu(\overline{e}) = \sum_{\overline{e} \in \mathrm{supp}(\nu):\overline{e}(t)(t)=x} \nu(\overline{e}).$$

Below, we compute the second path on $\nu \in \mathbb{D}_{\leq 1}VV(X)$, $t \in \mathbb{R}_{\geq 0}$ and $x \in X$.

$$\lambda_X \circ \mathbb{D}_{\leq 1}(\mu_X^R)(\nu)(t)(x) = \lambda_X(\mathbb{D}_{\leq 1}(\mu_X^R)(\nu))(t)(x) = \sum_{e \in \mathrm{supp}(\mathbb{D}_{\leq 1}(\mu_X^R)(\nu)):e(t)=x} \mathbb{D}_{\leq 1}(\mu_X^R)(\nu)(x)$$

$$= \sum_{e \in \mathrm{supp}(\mathbb{D}_{\leq 1}(\mu_X^R)(\nu)):e(t)=x} \; \sum_{\overline{e} \in \mathrm{supp}(\nu):\mu_X^R(\overline{e})=e} \nu(\overline{e}).$$

For $\overline{e} \in VV(X)$ and $t \in \mathbb{R}_{\geq 0}$, we have $\mu^R(\overline{e})(t) = \overline{e}(t)(t)$. Therefore

$$\lambda_X \circ \mathbb{D}_{\leq 1}(\mu_X^R)(\nu)(t)(x) = \sum_{e \in \mathrm{supp}(\mathbb{D}_{\leq 1}(\mu_X^R)(\nu)):e(t)=x} \; \sum_{\overline{e} \in \mathrm{supp}(\nu):\overline{e}(s)(s)=e(s) \ \forall s \in \mathbb{R}_{\geq 0}} \nu(\overline{e})$$

$$= \sum_{\overline{e} \in \mathrm{supp}(\nu):\overline{e}(t)(t)=x \ \forall s \in \mathbb{R}_{\geq 0}} \nu(\overline{e}).$$

Since both paths lead to the same result, the diagram commutes.

$$\begin{tikzcd}
\mathbb{D}_{\leq 1}\mathbb{D}_{\leq 1}V(X) \arrow[r, "\mathbb{D}_{\leq 1}\lambda_X"] \arrow[d, "\mu^{\mathbb{D}}_{V(X)}"'] & \mathbb{D}_{\leq 1}V\mathbb{D}_{\leq 1}(X) \arrow[r, "\lambda_{\mathbb{D}_{\leq 1}(X)}"] & V\mathbb{D}_{\leq 1}\mathbb{D}_{\leq 1}(X) \arrow[d, "V(\mu_X^{\mathbb{D}})"] \\
\mathbb{D}_{\leq 1}V(X) \arrow[rr, "\lambda_X"'] & & V\mathbb{D}_{\leq 1}(X)
\end{tikzcd}$$

In order to prove that the diagram commutes, we compute both paths on $\nu \in \mathbb{D}_{\leq 1}\mathbb{D}_{\leq 1}V(X)$, $t \in \mathbb{R}_{\geq 0}$ and $x \in X$.

$V(\mu_X^{\mathbb{D}}) \circ \lambda_{\mathbb{D}_{\leq 1}(X)} \circ \mathbb{D}_{\leq 1}\lambda_X(\nu)(t)(x) = V(\mu_X^{\mathbb{D}})(\lambda_{\mathbb{D}_{\leq 1}(X)}(\mathbb{D}_{\leq 1}\lambda_X(\nu)))(t)(x) = \mu_X^{\mathbb{D}}\lambda_{\mathbb{D}_{\leq 1}(X)}(\mathbb{D}_{\leq 1}\lambda_X(\nu))(t))(x)$.

For $\nu \in \mathbb{D}_{\leq 1}\mathbb{D}_{\leq 1}V(X)$ and $e \in V\mathbb{D}_{\leq 1}(X)$, we have

$$\mathbb{D}_{\leq 1}(\lambda_X)(\nu)(e) = \sum_{\psi \in \mathrm{supp}(\nu) : \lambda_X(\psi) = e} \nu(\psi) = \sum_{\psi \in \mathrm{supp}(\nu) : \lambda_X(\psi) = e} \nu(\psi).$$

Moreover, for $\rho \in \mathbb{D}_{\leq 1}V\mathbb{D}_{\leq 1}(X)$, $t \in \mathbb{R}_{\geq 0}$ and $\phi \in \mathbb{D}_{\leq 1}(X)$ we have

$$\lambda_{\mathbb{D}_{\leq 1}X}(\rho)(t)(\phi) = \sum_{e \in \mathrm{supp}(\rho) : e(t) = \phi} \rho(e).$$

By substituting ρ with $\mathbb{D}_{\leq 1}(\lambda_X)(\nu)$ we get

$$\lambda_{\mathbb{D}_{\leq 1}X}(\mathbb{D}_{\leq 1}(\lambda_X)(\nu))(t)(\phi) = \sum_{e \in \mathrm{supp}(\mathbb{D}_{\leq 1}(\lambda_X)(\nu)) : e(t) = \phi} \; \sum_{\psi \in \mathrm{supp}(\nu) : \lambda_X(\psi) = e} \nu(\psi) = \sum_{\psi \in \mathrm{supp}(\nu) : \lambda_X(\psi)(t) = \phi} \nu(\psi).$$

Lastly, we apply $\mu_X^{\mathbb{D}}$ on $\lambda_{\mathbb{D}_{\leq 1}X}(\mathbb{D}_{\leq 1}(\lambda_X)(\nu))(t)$.

$$V(\mu_X^{\mathbb{D}}) \circ \lambda_{\mathbb{D}_{\leq 1}(X)} \circ \mathbb{D}_{\leq 1}\lambda_X(\nu)(t)(x) = \mu_X^{\mathbb{D}}(\lambda_{\mathbb{D}_{\leq 1}X}(\mathbb{D}_{\leq 1}(\lambda_X)(\nu))(t))(x)$$

$$= \sum_{\phi \in \mathrm{supp}(\lambda_{\mathbb{D}_{\leq 1}X}(\mathbb{D}_{\leq 1}(\lambda_X)(\nu))(t))} \left(\sum_{\psi \in \mathrm{supp}(\nu) : \lambda_X(\psi)(t) = \phi} \nu(\psi) \right) \cdot \phi(x) = \sum_{\psi \in \mathrm{supp}(\nu)} \nu(\psi) \cdot \lambda_X(\psi)(t)(x)$$

$$= \sum_{\psi \in \mathrm{supp}(\nu)} \nu(\psi) \left(\sum_{e \in \mathrm{supp}(\psi) : e(t) = x} \psi(e) \right).$$

Below, we compute the second path on $\nu \in \mathbb{D}_{\leq 1}\mathbb{D}_{\leq 1}V(X)$, $t \in \mathbb{R}_{\geq 0}$ and $x \in X$.

$$\lambda_X \circ \mu_{V(X)}^D(\nu)(t)(x) = \lambda_X(\mu_{V(X)}^D(\nu))(t)(x) = \sum_{e \in \mathrm{supp}(\mu_{V(X)}^D(\nu)) : e(t) = x} \mu_{V(X)}^D(\nu)(e)$$

$$= \sum_{e \in \mathrm{supp}(\mu_{V(X)}^D(\nu)) : e(t) = x} \; \sum_{\psi \in \mathrm{supp}(\nu)} \nu(\psi) \cdot \psi(e) = \sum_{\psi \in \mathrm{supp}(\nu)} \nu(\psi) \left(\sum_{e \in \mathrm{supp}(\mu_{V(X)}^D(\nu)) : e(t) = x} \psi(e) \right).$$

In the last step, we can interchange the summations due to Tonelli's theorem for interchanging summations with non-negative terms. Since both paths lead to the same result, the diagram commutes.

D Proof : $\mathbf{Alg}(F)$ is a category

Consider $\mathbf{Alg}(F)$ with $F : \mathcal{C} \to \mathcal{C}$ an endofunctor. In order to prove that $\mathbf{Alg}(F)$ is a category, we need to prove that, for each algebra (A, α), an identity morphisms exists, and for each pair of morphisms $f : (A, \alpha) \to (B, \beta)$ and $g : (B, \beta) \to (C, \gamma)$ there exists a unique composite morphism. These identity morphisms and composite morphisms must satisfy certain conditions (see Definition 3.1 of Subsection 3.1).

Identity morphism
Let (A, α) be an algebra and consider $\mathrm{id}_A : (A, \alpha) \to (A, \alpha)$ the identity morphism of $\mathcal{C}$ on object A. Since F is a functor, we have $F(id_A) = id_{F(A)}$. By combining this with the fact that the morphisms are invariant under composition with an identity function of $\mathcal{C}$, we get

$$\alpha \circ F(\mathrm{id}_A) = \alpha \circ \mathrm{id}_{F(A)} = \alpha = \mathrm{id}_A \circ \alpha.$$

This proves that $\mathrm{id}_A : (A, \alpha) \to (A, \alpha)$ is a morphism of $\mathbf{Alg}(F)$.

Composite morphism
Consider $f : (A, \alpha) \to (B, \beta)$ and $g : (B, \beta) \to (C, \gamma)$, and consider the composition $g \circ f$ of category $\mathcal{C}$. Since F is an endofunctor, we have $F(g \circ f) = F(g) \circ F(f)$. Moreover, since f and g are morphisms of $\mathbf{Alg}(F)$, for both morphisms the diagram in Definition 3.12 commutes. Therefore, we have

$$\gamma \circ F(g \circ f) = \gamma \circ F(g) \circ F(f) = g \circ \beta \circ F(f) = g \circ f \circ \alpha.$$

This proves that $g \circ f$ is a morphism in $\mathbf{Alg}(F)$.

In order for $\mathbf{Alg}(F)$ to be a category, we need to prove that composition is invariant under composition with the identity, and composition is associative. Since we use the same identity morphism and composite morphism as category $\mathcal{C}$, these properties are directly transferred from category $\mathcal{C}$.

E Proof : $\mathbf{Coalg}(F)$ is a category

Consider $\mathbf{Coalg}(F)$ with $F : \mathcal{C} \to \mathcal{C}$ an endofunctor. In order to prove that $\mathbf{Coalg}(F)$ is a category, we need to prove that, for each coalgebra (A, α), an identity morphisms exists, and for each pair of morphisms $f : (A, \alpha) \to (B, \beta)$ and $g : (B, \beta) \to (C, \gamma)$ there exists a unique composite morphism. These identity morphisms and composite morphisms must satisfy certain conditions (see Definition 3.1 of Subsection 3.1).

Identity morphism
Let (A, α) be a coalgebra and consider $\mathrm{id}_A : (A, \alpha) \to (A, \alpha)$ the identity morphism of $\mathcal{C}$ on object A. Since F is a functor, we have $F(id_A) = id_{F(A)}$. By combining this with the fact that the morphisms are invariant under composition with an identity function of $\mathcal{C}$, we get

$$F(\mathrm{id}_A) \circ \alpha = \mathrm{id}_{F(A)} \circ \alpha = \alpha = \alpha \circ \mathrm{id}_A.$$

This proves that $\mathrm{id}_A : (A, \alpha) \to (A, \alpha)$ is a morphism of $\mathbf{Coalg}(F)$.

composite morphisms
Consider $f : (A, \alpha) \to (B, \beta)$ and $g : (B, \beta) \to (C, \gamma)$, and consider the composition $g \circ f$ of category $\mathcal{C}$. Since F is an endofunctor, we have $F(g \circ f) = F(g) \circ F(f)$. Moreover, since f and g are morphisms of $\mathbf{Coalg}(F)$, for both morphisms the diagram in Definition 3.12 commutes. Therefore, we have

$$F(g \circ f) \circ \alpha = F(g) \circ F(f) \circ \alpha = F(g) \circ \beta \circ f = \gamma \circ g \circ f.$$

This proves that $g \circ f$ is a morphism in $\mathbf{Alg}(F)$.

In order for $\mathbf{Alg}(F)$ to be a category, we need to prove that composition is invariant under composition with the identity, and composition is associative. Since we use the same identity morphism and composite morphism as category $\mathcal{C}$, these properties are directly transferred from category $\mathcal{C}$.

References

[1] J. Adámek and J. Reiterman. Banach's fixed-point theorem as a base for data-type equations. *Appl. Categ. Structures*, 2(1):77–90, 1994.

[2] A. Agore. *A First Course in Category Theory*. Universitext. Springer International Publishing, 2023.

[3] A. D. Ames, S. Mattenet, and J. Moeller. Categorical lyapunov theory ii: Stability of systems, 2025.

[4] J. Beck. Seminar on Triples and Categorical Homology Theory. *Repr. Theory Appl. Categ.*, (18):303, 2008. Lectures from the seminar held at Forschungsinstitut für Mathematik, ETH, Zürich, 1966/1967, Reprint of the 1969 original, With a preface to the reprint by Michael Barr.

[5] Arij Benkhadra and Isar Stubbe. A logical analysis of fixpoint theorems. *Cah. Topol. Géom. Différ. Catég.*, 64(1):97–121, 2023.

[6] M. Giry. A categorical approach to probability theory. In B. Banaschewski, editor, *Categorical Aspects of Topology and Analysis*, pages 68–85, Berlin, Heidelberg, 1982. Springer Berlin Heidelberg.

[7] G. Grimmett and D. J. A. Welsh. *Probability : an introduction / Geoffrey Grimmett, Dominic Welsh.* Oxford University Press, Oxford, [England, second edition. edition, 2014 - 2014.

[8] M. Hairer. Ergodic properties of markov processes, 2006. Lecture notes, University of Warwick.

[9] B. Jacobs. *Introduction to coalgebra : towards mathematics of states and observation.* Cambridge tracts in theoretical computer science ; 59. Cambridge University Press, Cambridge, 2017.

[10] M. Kick, J. Power, and A. Simpson. Coalgebraic semantics for timed processes. *Information and Computation*, 204(4):588–609, April 2006.

[11] E. Kreyszig. *Introductory Functional Analysis with Applications*. John Wiley & Sons, New York, 1989.

[12] E. Di Lavore, J. Forster, and M. Román. Graded monad coalgebras for continuous-time transition systems, 2026.

[13] J. R. Norris. *Markov chains*. Cambridge series on statistical and probabilistic mathematics. Cambridge University Press, Cambridge, 1st pbk. ed. edition, 1997.

[14] E. Seneta. *Non-negative Matrices and Markov Chains*. Springer Series in Statistics. Springer New York, 2006.

[15] A. Sokolova. Probabilistic systems coalgebraically: A survey. *Theoretical Computer Science*, 412(38):5095–5110, 2011. CMCS Tenth Anniversary Meeting.

[16] C. Villani. *Optimal Transport: Old and New*. Grundlehren der mathematischen Wissenschaften. Springer Berlin Heidelberg, 2008.