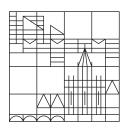
[Draft] Probabilistic Automata Semantics, Equivalence & Minimization

Master Thesis by

Fabian Klopfer

Universität Konstanz



Faculty of Sciences

Department of Computer and Information Science

Modeling of Complex, Self-Organizing Systems Group

1st Reviewer: Prof. Dr. Tatjana Petrov

2nd Reviewer: ToDo



Some New Abstract Text

Contents

1	Intro	ntroduction			
2	2.1 2.2 2.3 2.4 2.5 2.6	Catego Proba Auton Forma	d minaries ory Theory ory Theory organization at Theory organization at Language Theory organization at Equations organization at Equation 2007.	6 10 10 11 12 12	
3	Sem	antics	Equivalence & Minimization	13	
J	3.1		ntics	13	
	3.1	3.1.1	Parametrization and Initialization	13	
		3.1.1	Trace Semantics	13	
		3.1.3	Transient Semantics (incl. Reward/weighted semantics)	13	
		3.1.4	Threshold semantics	13	
		3.1.4 $3.1.5$	Funtion Transient Semantics (?)	13	
		3.1.6	Languages & Grammars	13	
	3.2		alence	13	
	5.2	3.2.1	Trace semantics is decidable in P for all	14	
		3.2.1	Transient semantics	14	
		3.2.3	Word Function-based	14	
	3.3		nization	14	
	3.3	3.3.1	Approaches	14	
		0.0	3.3.1.1 Partition Refinement - Coalgebraic Approach	14	
			3.3.1.2 Schützenberger's Construction and Arnoldi Iteration with Housh-		
			older Reflectors &	14	
		3.3.2	Wrt. different semantics	14	
			3.3.2.1 Decidability	14	
			3.3.2.2 Complexity	15	
4	Con	clusion	s	16	
-					

Contents 3

1 Introduction

2 Background

As one aim of this thesis is to connect aspects from different domains — namely automata theory, systems of equations and some notions of category theory and algebra — we introduce all terms used in the following sections and chapters carefully. The intention is that every individual that has received basic training in one of the natural sciences is able to understand this work without having to look up concepts in other literature. We start out by defining the terminology from category theory, continue with some notions from algebra and probability theory and finally introduce formalisms to describe dynamical systems, namely automata theory and systems of equations.

TODO minimize category theroy & coalgebraic stuff, semirings, \dots ; Focus on the important stuff

2.1 Preliminaries

The following paragraphs introduce the most basic mathematical objects and notions around them: The set theory of vno Neumann Bernay and Gödel. This is followed by the definition of relations — like equivalence — and some definitions of algebraic structures: Monoids, groups, rings and semi-rings. Then we give some examples of algebraic structures in the form of the probability semiring, the ring of formal power series, the polynomial ring, the semiring of NFAs,

The axiomatic definition of sets by Zermalo-Frenkel [10] with the extension of the axiom of choice (in short ZFC) or the formally equivalent definition of sets by von Neumann, Bernays and Gödel (NBG) [4] can be used. While ZFC [7] is the standard definition, NBG matches here more closely due to the extension of classes which are defined afterwards.

The following constructions are required [1]:

Definition 2.1.1 (Set). • We can form subsets, that have certain properties: Let X be a set and P be a formula in first-order logic over one variable called "'property"', then $\forall X \forall P : \{x \in X | P(x)\}$

- We can form the set of all subsets, called the **Power Set** \mathcal{P} : $\forall X : \mathcal{P}(X)$
- For a family of sets $(X_i)_{i\in I}$ we can form the following sets:
 - the set containing two sets: $\{X_1, X_2\}$
 - for $n \in \mathbb{N}$ sets X_1, X_2, \dots, X_n the **ordered** n-tuple (X_1, X_2, \dots, X_n)
 - the **image** of the indexing function ²: $\{X_i | i \in I\}$
 - the Cartesian product with f being a function $\Pi_{i \in I} X_i = \{f : I \to \bigcup_{i \in I} | \forall i \in I : f(i) \in X_i\}$
 - the union $\bigcup_{i \in I} X_i = \{x | x \in X_i, i \in I\}$
 - the disjoint union $\uplus_i X_i = \bigcup_{i \in I} (X_i \times \{i\})$
 - the intersection $\bigcap_i X_i = \{x | \forall i : x \in X_i\}$
 - the relative complement $X_1 \setminus X_2 = \{x | x \in X_1 \land x \notin X_2\}$

¹see 2.1.2 with for a definition

 $^{^{2}}$ see 2.1.4 for a definition

³²

- the set of all functions⁴ $X_2^{X_1}$, that is $\{f|f:X\to Y\}$.
- We can also form the following sets: \mathbb{N} of all natural numbers, \mathbb{Z} of all integers, \mathbb{Q} of all rational numbers, \mathbb{R} of all real numbers, \mathbb{C} of all complex numbers.

Every automaton is a set when using this definition.

In order to construct functions between classes of automata we need another level of abstraction to be able to avoid certain problems with sets of sets and functions on these [5].

Definition 2.1.2 (Class [1]). A class is a "large" collection of sets. It satisfies the following axioms:

- the members of a class are sets.
- every set is a class. A class is called **proper** if it is not a set.
- There is no surjection from a set to a proper class.
- We can form The class of all sets with property P:
 Let X be a set and P be a formula in first-order logic over one set called "'property"', then
 ∀P∀S: {S|P(S)}}.
- If X_1, X_2, \ldots, X_n are classes then then the n-tuple (X_1, X_2, \ldots, X_n) is a class.
- classes are closed under countable union, intersection and Cartesian product.

Definition 2.1.3 (Universe, Family, Large Classes, Small Classes, Proper Classes). The class of all sets is called **universe** \mathcal{U} . All classes are sub-collections of \mathcal{U} .

A family of sets $(A_i)_{i\in I}$ is a function $f: I \to U$. If I is a set, then the family is said to be set-indexed.

A proper class is also called large, while a set is called a small class.

The class of all automata is a proper class. The notion of classes and sets is needed to evade Russell's paradox and the charachterization in [1] largely coincides with the definition by Gödel [4] and the more pedagogically oriented restatement by Mendlsson [8]. Since we are interested in automata only we do not consider the notion of conglomerates. If interested, the reader is referred to Herrlich & Strecker [5].

As we want to investigate how to minimize automata and how to check their equivalence, it is useful to define the notions of an equivalence relation, which is tightly related to a category theoretical point of view, as we will see.

Definition 2.1.4 (Binary Relation). A binary relation R over sets X, Y is a subset of the Cartesian product $R \subseteq (X \times Y)$. R is a set of ordered pairs (x, y) with $x \in X, y \in Y$. Thus two elements are related, if the ordered pair of those elements is contained in the set $xRy \Leftrightarrow (x, y) \in R$. X is called the **domain**, while Y is called the **co-domain**.

Definition 2.1.5 (Function). A function f over two sets X, Y is a binary relation which maps every element of the domain to exactly one element of the co-domain.

$$\forall x \in X \exists_1 y \in Y : (x, y) \in f$$

We write $f: X \to Y$ to define function, domain and co-domain, while we write f(x) = y to refer to a specific pair in the set or $x \mapsto f(x)$. The **image** of a function is the set defined with $X_0 \subseteq X$ by

$$f(X_0) = \{ f(x) \in Y | x \in X_0 \}$$

2.1. PRELIMINARIES 7

⁴2

The pre-image of f on a subset of the co-domain $Y_0 \subseteq Y$ is given by

$$f^{-1}(Y_0) = \{ x \in X | f(x) \in Y_0 \}$$

A function is called **injective**, if no two elements of the domain map to the same element in the co-domain.

$$\forall x \in X \forall z \in X : (x \neq z) \Rightarrow f(x) \neq f(z)$$

A function is called **surjective**, if every element in the co-domain is mapped to.

$$\forall y \in Y \exists x \in X : y = f(x)$$

A function is called bijective, if every element in the co-domain is mapped to exactly once or in different terminology if it is surjective and injective.

$$\forall y \in Y \exists_1 x \in X : y = f(x)$$

Functions can be composed, if the image of the co-domain of the first applied function is equal to the domain of the second applied one. For $x \in X$, $f: X \to Y$ and $g: Y \to Z$ the composition $g \circ f: X \to Z$ is defined by

$$(g \circ f)(x) = g(f(x)) \in Z$$

We are going to consider functions between different types of automata that preserve their structure. We are especially interested in surjective functions between two automata where the co-domain automata is a minimal realization of the domain automata. To be sure that the this function produces indeed the same behaviour, we have to define special notions of equivalence, thus introduce here what a equivalence relation is in general.

Definition 2.1.6 (Equivalence Relation, Equivalence Class). An equivalence relation $\subseteq X \times X$ is a relation that is for $\forall x, y \in X$

- reflexive: x x.
- $symmetric: x y \Rightarrow y x$
- and transitive $x \ y \wedge y \ z \Rightarrow x \ z$

We will be looking for cannonical surjections between automata defined over special equivalence relations, that we define in the next chapter.

Definition 2.1.7 (Equivalence Class). Let $\subseteq X \times X$ be a equivalence relation. An **equivalence** class is the class that is induced by the property of the equivalence relation : For $x \in X$

$$[x] = \{ z \in X | x \ z \}$$

Equivalence classes are disjoint. The set of all equivalence classes

$$X \setminus = \{ [x] | x \in X \}$$

is called quotient set of X and . The surjective function $q: X \to X \setminus$ is called canonical surjection and is defined by $x \mapsto [x]$ for.

Introduce matrix notations for sparse MM in next chapter

In order to define the word structures of different types of automata, we define the notions of monoid, which M.P. Schützenberger used when he first defined the notion of weighted automata — also presenting some fundamental theorems for minimization we consider in the next chapter [9].

Definition 2.1.8 (Monoid). A **Monoid** is a 3-tuple $M = (S, \circ, 0)$ consisting of a set S and a binary relation \circ , where M has the following properties, with $\forall x, y, z \in S$:

- M has a neutral element: $\exists 0 \in S : x \circ 0 = 0 \circ x = x$
- M is associative: $(x \circ y) \circ z = x \circ (y \circ z)$

If a monoid is also commutative, that is

$$\forall x, y \in S : x \circ y = y \circ x$$

we call it a commutative monoid.

In order to define the notion of ring later on we also need to know what a group is.

Definition 2.1.9 (Group). A group is a 3-tuple $G = (S, \circ, 0)$ consisting of a set S and a binary relation \circ , where G is a monoid and has an inverse for each element:

$$\forall x \in S \exists i \in S : x \circ i = i \circ x = 0$$

A group is commutative if G is a commutative monoid.

Many authors in the field of weighted automata considered semirings as the algebraic structure over which a weighted automaton operates [3, 2]. Explicitly the interpretation of labels and weights, as well as the executions, paths, traces and the like, differ depending on the set and operations that are chosen. In order to leverage their results in the next chapter we define:

Definition 2.1.10 (Semiring). A **Semiring** is defined by a 5-tuple $\mathbb{S} = (S, +, \cdot, 0, 1)$ where

- (S, +, 0) is a commutative monoid
- $(S, \cdot, 1)$ is a monoid
- the distributive law is valid: $\forall a, b, c \in S : a \cdot (b+c) = (a \cdot b) + (a \cdot c) \wedge (b+c) \cdot a = (b \cdot a) + (c \cdot a)$
- 0 is an annihilator: $\forall a \in S : 0 \cdot a = a \cdot 0 = 0$

 \mathbb{S} is a **commutative semiring**, if $(S,\cdot,1)$ is a commutative monoid.

A semiring is closed [6], if

- + is idempotent: $\forall a \in S : a + a = a$
- the infinitary application of + for every countable sequence $(s_i)_{i \in I}$ (notation $\Sigma_{i \in I} s_i$) is associative, commutative, idempotent and distributive

In a closed semiring one can define * by [6]

$$b* = \sum_{k \le n} b^k$$

such that $ab^*c = \Sigma_n ab^n c$, thus * is the algebraic equivalent to the Kleene star in regular expressions as we will see in an example.

Definition 2.1.11 (Ring). A ring is defined by a 5-tuple $\mathbb{S} = (S, +, \cdot, 0, 1)$ where

- S is a semiring
- (S, +, 0) is an abelian group

A ring is called commutative if $(S, \cdot, 1)$ is commutative.

Remark. A ring is also a semiring. This follows immediately from the definition of an abelian group.

2.1. PRELIMINARIES 9

Semiring	Name	Informal Description	
		using automata	
Bool = $(\{0,1\}, \vee, \wedge, 0, 1)$	Boolean Semiring	Represents the boolean	
		algebra	
$Nat = (\mathbb{N}, +, \cdot, 0, 1)$	Ring of Natural Num-	Unsigned integer arith-	
	bers	metic	
$FPS = (R[X], +, \cdot, 0, 1)$	Ring of Formal Power	Formal Power Series	
	Series	over Ring R with	
		variables X	
$\operatorname{Lang}_{\Sigma} = (\mathcal{P}(\Sigma *), \cup, \circ, \emptyset, \varepsilon)$	Languages over Σ	closed semiring, repre-	
		sents semantics of NFAs	
$Trop = (\mathbb{N} \cup \mathbb{N})$	Tropical Semiring	e.g. Shortest paths in a	
$\{\infty\}, \min, +, \infty, 0)$		graph	
$Arct = (\mathbb{N} \cup \mathbb{N})$	Arctic Semiring	e.g. analysis of synchro-	
$\{-\infty\}, \max, +, -\infty, 0\}$		nized systems	
NonDetProb =	Non-deterministic	used when non-	
$(\mathbb{R}_{\geq ec{arphi}},+,\cdot,0,1)$	Probabilistic Semiring	determinism is required	
$Vit = ([0,1], \max, \cdot, 0, 1)$	Viterbi Semiring	e.g. used for modelling	
		RNA	
$Fuzzy = ([0,1], \min, \max, 0, 1)$	Fuzzy Semiring	used in fuzzy logic	

Table 1 Examples of different semirings

For a more thouroughly treatment of monoids and semirings, the reader is referred to Droste et al. and Kozen [3, 6]

Example. /2/

Several examples of semirings are given in 1. One can replace \mathbb{N} by $\mathbb{Z}, \mathbb{Q}, \mathbb{R}, \mathbb{C}$ to get integer, rational, real or complex instead of the natural semiring. Further \mathbb{N} can replaced by \mathbb{R}^+ in the arctic and tropical semiring. Note that the Viterbi semiring defines also defines a semirings over probabilities, but neglects addition and is thus not able to model non-determinism.

Example. [3] Given the closed Semiring Lang. Kleene iteration is defined as $L* = \bigcup_{n\geq 0} L^n$ for all $L\subseteq \Sigma^*$

needed for schützenbergers theorems

Definition 2.1.12 (Semimodule).

Theorem 2.1.12.1 (Semiring of Matrices & Vectors).

2.2 Category Theory

Definition 2.2.1 (Coalgebra, F-Coalgebra).

2.3 Probability Theory

Definition 2.3.1 (Sample space, events, σ -algebra, measurable space). Let Ω be a set, namely the set of possible outcomes of a chance experiment, called **sample space**.

Let $\mathcal{F} \subseteq \mathcal{P}(\Omega)$ with \mathcal{P} the power set, a set of subsets of the sample space, whose elements are called **events**.

Let Ω be a sample space, \mathcal{F} a set of events. \mathcal{F} is called a σ -algebra over Ω , if and only if

• the sample space is contained in the set of events,

$$\Omega \in \mathcal{F}$$

• the set of events is closed under complement,

$$A \in \mathcal{F} \Rightarrow \Omega \setminus A \in \mathcal{F}$$

• and the set of events is closed under countable union:

$$\forall i \geq 0 : A_i \in F \Rightarrow (\bigcup_{n \in \mathbb{N}A_n) \in \mathcal{F}}$$

The pair (Ω, \mathcal{F}) is called a **measurable space**.

Let $(\Omega_1, \mathcal{F}_1), (\Omega_2, \mathcal{F}_2)$ measurable spaces. A function $f : \Omega_1 \to \Omega_2$ is called a **measurable** function if and only if for every $A \in \mathcal{F}_2$ the pre-image of A under f is in \mathcal{F}_1 .

$$\forall A \in \mathcal{F}_2 : f^{-1}(A) \in \mathcal{F}_1$$

Definition 2.3.2 (Probability Space, Probability Measure, Discrete and Continuous Probability Space and Measure).

Definition 2.3.3 (Random Variable, Probability Distribution, Distribution Funtion and Cumulative Density Function).

Definition 2.3.4 (Mean, Median, Mode, Expectation, Variance).

Definition 2.3.5 (Joint Probability Distribution, Conditional Probability, Bayes Rule, Independence, Conditional Independence).

Definition 2.3.6 (Stochastic Process, Bernouilli & Binomial Process).

Definition 2.3.7 (Geometric Distribution).

Definition 2.3.8 ((Negative) Exponential Distr.).

Theorem 2.3.8.1 (Memoryless property).

Definition 2.3.9 (Markov Property, Markov Process, Time Homogeneity).

2.4 Automata Theory

Definition 2.4.1 (Transition System).

Definition 2.4.2 (Labelled Transition System).

Example (DFA).

Definition 2.4.3 (Path, Trace, Cylinder Sets, Prefix, Postfix).

Definition 2.4.4 (Determinism and Non-Determinism, Internal vs. External, Adversaries/Policies).

Example (DFA & NFA).

Definition 2.4.5 (Weighted Automata).

Example (WAs over Boolean Semiring, Rational Semiring, Polynomial Semiring, plus-max Semiring, Probability Semiring).

Definition 2.4.6 (Probabilistic Automata, Initial Distribution, transition probability function, stochastic matrix, transition probability matrix).

Remark (Probabilistic vs. Non-Deterministic Choice).

Remark (Sokolova's System Types Paper and focus on MC-like models).

Definition 2.4.7 (Markov Chains, Discrete-Time, Continuous-Time, Labelled, MDP).

Definition 2.4.8 (Lumpability, exact, ordinary, weak, finite-horizon, buchholz).

2.5 Formal Language Theory

Definition 2.5.1 (Abstract Rewriting Systems).

Definition 2.5.2 (Semi-Thue Systems or String Rewriting Systems).

Definition 2.5.3 (Stochastic Rewrite Systems).

Definition 2.5.4 (Grammar, Weighted, Probabilistic).

2.6 Systems of Equations

Definition 2.6.1 (Algebraic, Recurrence, Difference, Differential).

Definition 2.6.2 (linear, non-linear, polynomial, stochastic).

Definition 2.6.3 (Ordinary, partial, homogeneous, inhomogeneous, order (first, higher)).

Remark (Transformation higher order to first order).

Definition 2.6.4 (initial value problem, boundary value problem, laplace & fourier transform, phase space).

3 Semantics, Equivalence & Minimization

The general case, may help with some things, e.g. if proven independent of the semiring used. So for each subsection here starting with WAs independent of the Semiring that is used, continue with PA results that are as independent as possible from the concrete transition structure. Finally apply the aforementioned results to LMCs/MCs. Also discriminate between DTMC, MDP, PA model(s) I.e. per subsection apply the folling structure

WA - General Case for arbitrary Semirings

PA - Results for other transition structures

MC-like models - Application of above and other literature on specific example of LMC

3.1 Semantics

3.1.1 Parametrization and Initialization

relation to initial value problem & phase space Connection to Systems of Equations ODEs as central limit of CTMCs & their SDEs [Prof. Thomas G. Kurtz, wisc]

3.1.2 Trace Semantics

Execution as Sparse Matrix Multiplication, (constrained) reachability (pr.), path/trace distributions, ergodicity, state residency time, Uniformization

3.1.3 Transient Semantics (incl. Reward/weighted semantics)

Definition 3.1.1 (Transient probability distribution).

3.1.4 Threshold semantics

3.1.5 Funtion Transient Semantics (?)

word functions instead of probability functions?

3.1.6 Languages & Grammars

Threshold Language

Transient/Population Language

3.2 Equivalence

Definition 3.2.1 (Bisimulation Relations, Strong, weak, forward, backward Prob., Buchholz).

Equivalence classes of initial distrs. (rubino sericola)

3.2.1 Trace semantics is decidable in P for all

Definition 3.2.2 (trace Equivalence, branching bisim.).

paz p.36, Kiefer, Tzeng, Schützenberger, Bollig & Zeitoun, Kiefer WA should hold for PA when using sufficient eps (theoretically unclean), Doyen,

Let automata $A_i = (S_i, \Sigma_i, M^{(i)}, \pi_i, \eta_i)$ for $i \in \{1, 2\}$ Equivalence of initialized automata:

$$A_1(\pi_1)$$
 $A_2(\pi_2) \Leftrightarrow \forall \sigma \in \Sigma : \pi_1 \eta_1 = \pi_2 \eta_2 \wedge \pi_1 M_{\sigma}^{(1)} = \pi_2 M_{\sigma}^{(2)}$

How about uninitialized? "unique minimal realization" due to [1], i.e. minimize uninitialized, then compare

3.2.2 Transient semantics

Definition 3.2.3. transient equivalence, , finite-horizon bisimulation (bisimulation up to time-step k) by katoen

Finite horizon bisim. What is T here? A Trace with a certain property (e.g. ending with σ or up to k steps?)

$$A_1(\pi_1) \ A_2(\pi_2) \Leftrightarrow \forall \sigma \in \Sigma : \ \pi_1 M_{\sigma}^{(1)} \eta_1 = \pi_2 M_{\sigma}^{(2)} \eta_2 \wedge \pi_1 T_1 = \pi_2 T_2$$

3.2.3 Word Function-based

finite-horizon + reward fn

$$A_1(\pi_1)$$
 $A_2(\pi_2) \Leftrightarrow \forall \sigma \in \Sigma : \pi_1 M_{\sigma}^{(1)} \eta_1 \mu_{\sigma} = \pi_2 M_{\sigma}^{(2)} \eta_2 \mu_{\sigma} \wedge \pi_1 T_1 = \pi_2 T_2$

3.3 Minimization

3.3.1 Approaches

3.3.1.1 Partition Refinement - Coalgebraic Approach

Proof: PA minimization is in P for uninitialized/strict lumpability

Deiffel, Wissmann, Paz p.24ff (IntroToProbabilisticAutomata) valmari for all in O(n log n) as partition refinement is minimal (see sources above) for uninitialized Automata.

Rubino sericola for initializations

in case ODEs: Tribastones work & boreale

3.3.1.2 Schützenberger's Construction and Arnoldi Iteration with Housholder Reflectors &

Show either that is also in P oder correct Kiefers runtime analysis. For initialized see kiefer;

3.3.2 Wrt. different semantics

3.3.2.1 Decidability

trace semantics: is decidable, see e.g. Kiefer or Mateus, Qiu, Li;;;; Trace equivalence PSpace Complete (according to TU Eindhoven's j.f. Groote

Transient semantics? finite horizon bisimulation minimization using partition refinement weighted transient semantics?

3.3.2.2 Complexity

trace semantics: TODO figure out if kiefers reduction is flawed or if the runtime analysis of his algo is flawed. NP from complexity proof, O(sigma n^3) in algo rt. check against rubino sericola n log m for uninitialized/strong

Transient semantics? n log m, if finite horizon bisim is right weighted transient semantics? as above

3.3. MINIMIZATION 15

4 Conclusions

Maybe some practical aspects? e.g. Numerics, Matmuls, concurrency (NC) Summarize Outlook

Bibliography

- [1] Ji Adámek, Horst Herrlich, and George E Strecker. "Abstract and concrete categories. The joy of cats". In: (2004).
- [2] Benedikt Bollig and Marc Zeitoun. "Weighted Automata". In: LSV, ENS Marc Zeitoum (2011).
- [3] Manfred Droste, Werner Kuich, and Heiko Vogler. *Handbook of weighted automata*. Springer Science & Business Media, 2009.
- [4] Kurt Gödel. Consistency of the Continuum Hypothesis. (AM-3). Princeton University Press, 1940. ISBN: 9780691079271. URL: http://www.jstor.org/stable/j.ctt1b9rzx4.
- [5] Horst Herrlich and George E Strecker. Category theory: an introduction. Vol. 1. Heldermann, 1979.
- [6] Dexter Kozen. "On Kleene algebras and closed semirings". In: *International Symposium on Mathematical Foundations of Computer Science*. Springer. 1990, pp. 26–47.
- [7] K KUNEN. "Set Theory". In: Handbook of Mathematical Logic (1982), p. 317.
- [8] Elliott Mendelson. Introduction to mathematical logic. CRC press, 1987.
- [9] Marcel Paul Schützenberger. "On the definition of a family of automata". In: *Inf. Control.* 4.2-3 (1961), pp. 245–270.
- [10] Ernest Zermelo. "Über Grenzzahlen und Mengenbereiche". In: Fundamenta Mathematicae 16.1 (1930), pp. 29–47.

Bibliography 17