Completeness Theorems for Behavioural Metrics and Equivalences

Wojciech Krzysztof Różowski

A dissertation submitted in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

of

University College London.

Department of Computer Science

University College London

March 5, 2025

I, Wojciech Krzysztof Różowski, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the work.

Abstract

My research is about stuff.

It begins with a study of some stuff, and then some other stuff and things.

There is a 300-word limit on your abstract.

Impact Statement

Outside academia: The discipline of Formal Verification enables making precise logical statements about computer systems to guarantee their correctness. This relies on the study of models of computation, which are mathematical objects representing the semantics of systems of interest. In Theoretical Computer Science, it is customary to model computations as transition systems. This thesis is part of a larger research programme aimed at providing specification languages. for representing transition systems and studying formal systems for reasoning about the equivalence or similarity of systems represented by the syntax. One of the central kind of problems attached to this field of research are completeness problems, that concern showing that every semantic equivalence or similarity of transition systems can be witnessed through the means of axiomatic manipulation. Having a complete axiomatisation allows one to fully resort to syntactic reasoning, which is particularly amenable to implementation and thus desirable from the automated reasoning point of view.

Kleene Algebra (KA) is a central example of such specification language. KA is a foundation of tools relevant to industry, such as NetKAT, which enables reasoning about the behaviour of packet-passing Software-Defined Networks, or cf-GKAT, which can be used to certify the correctness of decompilation algorithms. This thesis particularly focuses on the semantic notions of behavioural distances and probabilistic language equivalence, where it makes its main contributions. Behavioural distances replace the strict notion of equivalence of states with a more liberal quantitative measure of dissimilarity. This is particularly desirable for stochastic or probabilistic systems, where a tiny observed perturbation would lead to inequiva-

lence of states. Behavioural distances have been successfully applied to Markov Decision Processes within Reinforcement Learning (RL), with the key example being MiCO (Matching under Independent Couplings) distance. Additionally, probabilistic language equivalence, axiomatised in Chapter 4 of this thesis, underpins Apex, an automated equivalence checker for probabilistic programs.

Inside academia: This thesis makes original contributions within the field of Theoretical Computer Science. The results in Chapter 2 are an adaptation of a concrete completeness result for behavioural distance of probabilistic transition systems to an instance of an abstract coalgebraic framework. Besides the basic examples provided by Lobbia et al, the content of Chapter 3 is the first work to propose a complete axiomatisation of a quantitative calculus of string diagrams through a systematic axiomatic foundation. Finally, Chapter 4 provides an alternative axiomatisation of language equivalence of generative probabilistic transition systems through a simple generalisation of Kleene's regular expressions. The completeness result makes use of recently developed theory of proper functors and provides further evidence that the use of coalgebras for proper functors provides a good abstraction for completeness theorems.

The entire material presented in this thesis has been accepted or is under review for several highly-ranked conferences in Theoretical Computer Science. The results of this thesis have been disseminated to the computer science community through a series of seminar talks across the United Kingdom, the United States, and Germany.

Acknowledgements

Acknowledge all the things!

Contents

List of Figures

List of Tables

Chapter 1

Introduction

One of the motivations for the mathematical study of the models of computation stems from the desire for precise and formal reasoning about the correctness of computer systems. These theoretical foundations enable formal verification experts to prove that systems deployed in safety-critical areas, such as avionics or healthcare, behave as expected. In Theoretical Computer Science it is customary to model computations as state transition systems, which are discrete models where a set of states is equipped with a notion of one-step observable behaviour, describing how the system evolves. Typical examples include finite automata, Kripke frames, and Markov chains among many others.

This central topic of this thesis are axiomatisations of behaviour of transition systems. By this we mean providing expression languages for representing the behaviour of transition systems and the study of formal systems for reasoning about equivalence or similarity of behaviours represented by expressions of the interest.

The interest in axiomatising behaviour of transition systems originates from the seminal work of Kleene on regular expressions. In his influential paper from 1956, Kleene introduced deterministic finite automata (DFAs), which are the fundamental model of sequential deterministic computation. Each state of a DFA can be associated with a formal language, a collection of strings that are accepted starting from a given state. This characterises an important class of formal languages, known as regular languages. Classically, two states are equivalent if they recognise the same language. In the same paper, Kleene proposed regular expressions, which are an

algebraic specification language for DFAs and proved that both formalisms are equally expressive through a result known nowadays as Kleene's theorem. As an open problem, he left a completeness question: are there a finite number of rules that enable reasoning about language equivalence of regular expressions?

Shortly after Kleene's paper, Redko demonstrated that one cannot use a finite number of equational axioms to axiomatise the language equivalence. But the search for axiomatisation made of more expressive rules continued. The first answer came in 1966 from Salomaa, who presented two axiom systems. One was infinitary, and the other used finite equations along with an implicational rule encoding Arden's rule for formal languages. While the latter became a blueprint for inference systems for reasoning about semantic equivalence or similarity of transition systems, this axiomatisation was not algebraic. Essentially, the implicational rule relied on the productivity side-condition called empty word property (EWP) that caused the resulting axiomatisation to be unsound under substitution of letters by arbitrary expressions. This problem has motivated several researchers including Conway, Krob and Boffa to pursue the problem of obtaining algebraic axiomatisation of language equivalence of DFAs, eventually leading to a celebrated completeness result of Kozen.

Since Kleene, automata theorists have studied many variants of automata, including nondeterministic, weighted and probabilistic ones, usually focusing on the notion of language equivalence or inclusion. At the advent of process algebra in the 1980s, Milner and Park brought the concept of (strong) bisimilarity, a notion of equivalence finer than language equivalence, that was motivated by the needs of the study of concurrency theory and models such as labelled transition systems (LTSs). Essentially, language equivalence is a linear-time notion, as it hides the precise moment of resolving nondeterministic choice from the external observer. At the same time, bisimilarity allows for a more fine-grained comparison of behaviours by looking at the exact moment of resolving the nondeterministic choice.

In his seminal paper from 1984, Milner considered a variant of LTSs that he called charts and studied the associated axiomatisation problem, but this time for

strong bisimilarity. Milner observed that while the syntax of regular expression can be used to specify behaviours of charts, it is not expressive, that is there exist behaviours that cannot be specified with Kleene's syntax. Instead, he proposed a more general language called algebra of regular behaviours (ARB) featuring binders, action prefixing, and a recursion operator. Milner provided a suitable generalisation of Salomaa's non-algebraic axiomatisation and demonstrated its soundness and completeness with respect to strong bisimilarity of charts.

Aforementioned completeness results of Salomaa, Kozen and Milner are prototypical instances of the vast strain of research, that has been of particular interest to theoretical computer scientists for decades. Given a transition system model and an associated notion of semantic equivalence, having a complete axiomatisation allows one to reason about model behaviour through the syntactic manipulation of terms of the specification language, which is well-suited for implementation, automation, and formal reasoning.

This thesis provides contributions to the study of axiomatisation problems of behaviours of transition systems in two orthogonal directions.

1. The first part of the thesis is concerned with the study of formal systems for quantitative reasoning about how close the behaviour of two states of transition systems is.

2. The second part focuses on probabilistic transition systems and presents a sound and complete axiomatisation of language equivalence of behaviours specified through the syntax of Probabilistic Regular Expressions.

1.1 Behavioural metrics

In many contexts, especially when dealing with probabilistic or quantitative models, focusing on exact equivalence of behaviours such as language equivalence or bisimilarity is too restrictive. A tiny perturbation in observed probability or weights of transition can deem two states inequivalent. Instead, it is often more meaningful to measure how far apart the behaviours of two states are.

This has motivated the development of behavioural metrics, which endow the state spaces of transition systems with (pseudo)metric spaces quantifying the dissimilarity of states. In such a setting, states at distance zero are not necessarily the same, but rather equivalent with respect to some classical notion of behavioural equivalence. In a nutshell, equipping transition systems with such a notion of distance crucially relies on the possibility of lifting the distance between the states to the distance on the one-step observable behaviour of the transition system.

Behavioural distances first appeared in the context of probabilistic transition systems. Here, one-step observable behaviour forms a probability distribution. In such a setting, in order to lift distances from the state space to one-step observable behaviour, one can rely on classic Kantorovich lifting from transportation theory.

Behavioural distances are not limited to probabilistic or weighted systems; instead, they can be defined meaningfully for a variety of transition systems. The simplest instances include Deterministic Finite Automata and Labelled Transition Systems. For example, DFAs can be equipped with a shortest-distinguishing-word distance, where the longer the smallest word that can witness inequivalence of two states is, the closer the behaviour of compared states is. To illustrate that, given an alphabet $\Sigma = \{a\}$, we have that a state recognising the language $\{a, aa, aaa\}$ is closer to the one recognising $\{a, aa, aaa, aaa\}$ rather than $\{a\}$.

So far, the study of expression languages and axiomatisations of behavioural distances have mainly focused on concrete probabilistic cases. Axiomatisations of other important instances of behavioural distances are still underexplored. The main goal of the first part of this thesis is to initiate the study of axiomatisations and completeness problems for behavioural distances beyond the concrete probabilistic instances. Our starting point is the work of Bacci, Bacci, Larsen and Mardare, who gave a sound and complete axiomatisation of branching-time behavioural distance of terms of probabilistic process calculus of Stark and Smolka.

1.2 Probabilistic Language Equivalence

In his seminal work from 1963, Rabin introduced Probabilistic Automata. This model captures the simple notion of randomised computation and acts as an acceptor for probabilistic languages. Under such semantics, each word over some fixed alphabet is associated with a weight from the unit interval capturing how likely the word is to be accepted. Throughout the years, Probabilistic Automata were deeply studied from an algorithmic point of view that eventually enabled the development of practical verification tools for randomised programs.

In the process algebra community, Larsen and Skou devised a notion of strong probabilistic bisimilarity for Reactive Probabilistic Transition Systems (RPTS). One can view RPTS as Rabin's Probabilistic Automata without an explicit termination probability. Later, Stark and Smolka introduced a probabilistic specification language similar to Milner's ARB. This language is for Generative Probabilistic Transition Systems (GPTS), a type of RPTS that meets an extra normalisation condition. In the same work, Stark and Smolka provided an axiomatisation of probabilistic bisimilarity of terms of their calculus and demonstrated its completeness.

Inspired by earlier mentioned Rabinovich's result for Milner's ARB, Silva and Sokolova extended Stark and Smolka's axiomatisation of probabilistic bisimilarity with coarser axioms characterising probabilistic language equivalence and demonstrated its completeness. Their result relied crucially on the completeness of fragment of axioms for probabilistic bisimilarity and featured a syntax with binders and a general recursion operator.

While the process algebraic syntax of Silva and Sokolova is expressive for probabilistic language equivalence, it is natural to ask if one could obtain a similar picture to the one for the usual language equivalence. That is to devise a simpler, binder-free specification language in the style of Kleene's Regular Expressions and provide a more streamlined axiomatisation in the style of Salomaa.

This problem is the central motivation for the second part of this thesis. One of the main inspirations for that comes from the community of probabilistic pattern matching, where researchers already considered Regular Expression-like operations

to specify probabilistic languages. They did so by replacing the union of languages and Kleene's star from the usual Regular Expression with their probabilistic counterparts, which respectively can be seen as a convex combination and a form of the Bernoulli process. At the same time, the precise connection of such syntaxes to the transition systems model was under-explored and the topic of axiomatisation was not tackled at all.

1.3 Coalgebra

The common aspect of both parts of the thesis is the use of the theory of Universal Coalgebra to provide the main technical results. The theory of coalgebras for an endofunctor provides a uniform and abstract treatment of state transition systems. Given a type functor over some base category (such as the category of sets and functions between them), one can immediately derive the associated notion of behaviour-preserving mappings and behavioural equivalence. Under mild set-theoretic constraints, one can devise a notion of a final coalgebra, an abstract domain of behaviours of transition systems of a given type, where every other transition system can be mapped in a behaviour-preserving way. For example, the final coalgebra for an endofunctor describing the behaviour of deterministic automata is precisely the set of all formal languages over some alphabet.

It is important to note, that this canonical coalgebraic notion of behavioural equivalence is of branching-time. For simple systems without side effects, like DFAs, this captures the usual notion of language equivalence. At the same time, for more complicated systems featuring side-effects like NFAs, LTSes or various kinds of probabilistic transition systems, coalgebraic behavioural equivalence captures variants of strong bisimilarity. In order to capture linear-time notions of behaviour, one has to move to a more structured category that encompasses the side-effect in the ambient structure of the objects of the category. For a class of coalgebras over Set endofunctors subject to some technical constraints, there is a generalisation of the classical automata-theoretic powerset construction that turns Set-coalgebras into coalgebras living in categories of Eilenberg-Moore algebras, where behavioural

equivalence recovers the usual automata-theoretic notions of language equivalence and its generalisations such as probabilistic language equivalence.

At the same time, the recent work on coalgebraic behavioural metrics provided a categorical generalisation of Kantorovich lifting to lifting endofunctors to the category of pseudometric spaces and nonexpansive maps between them. This enabled the uniform treatment of behavioural distances for a whole range of transition systems, beyond just the probabilistic ones.

1.4 Overview of the thesis

1.4.1 Part One: Behavioural Metrics

Chapter Two: A Complete Quantitative Axiomatisations of Behavioural Distance of Regular Expressions

Chapter Three: A Diagrammatic Approach to Behavioural Distance of Nondeterministic Processes

1.4.2 Part Two: Probabilistic Language Equivalence

Chapter Four: Completeness Theorem for Probabilistic Regular Expressions

Chapter 2

My First Content Chapter

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.

Chapter 3

My Second Content Chapter

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.

Chapter 4

General Conclusions

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.

Appendix A

An Appendix About Stuff

(stuff)

Appendix B

Another Appendix About Things

(things)

Appendix C

Colophon

This is a description of the tools you used to make your thesis. It helps people make future documents, reminds you, and looks good.

(example) This document was set in the Times Roman typeface using LaTeX and BibTeX, composed with a text editor.

Bibliography