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

February 27, 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 syntaxes for representing transition systems and studying logical systems for reasoning about the equivalence or similarity of systems represented by the syntax. One of the central problems attached to this paradigm 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.

A central exemplar of this paradigm is Kleene Algebra, which 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 inequivalence 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 automata, Kripke frames, and Markov chains among many others.

Faced with a profusion of such objects, theoreticians have developed several paradigms concerning the study of transition systems throughout the years. The focus of this thesis is on the classical approach of providing specification languages for representing transition systems and logics for reasoning about the equivalence or similarity of systems represented by expressions of interest.

1.0.1 Regular Expressions and Deterministic Finite Automata

A central exemplar of the paradigm outlined above is 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 describing possible trajectories that one-step dynamics can emit starting from the state of interest. This yields a notion of semantic equivalence, meaning that we say

two states are language equivalent if they denote the same language. In the same paper, Kleene proposed Regular Expressions, which are an algebraic specification language for DFAs, as well as proved that both approaches are equally expressive (up to language equivalence) 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 the language equivalence of regular expressions?

This kind of problem concerning specification languages has been of particular interest to theoretical computer scientists for decades. When a complete axiomatisation is available, one may reason about model behaviour through the syntactic manipulation of terms of the specification language, which is well-suited for implementation, automation, and formal reasoning.

In the case of Regular Expressions, 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. The latter became a blueprint for axiomatisations of semantic equivalence or similarity of transition systems.

1.0.2 Process Algebra

Since Kleene, automata theorists have studied a multitude of models, including nondeterministic, weighted and probabilistic ones, usually focusing on the notion of language equivalence or inclusion. The advent of Process Algebra in the 80s shifted the perspective. New models, such as Labelled Transition Systems (LTSes), were introduced. Even though Nondeterministic Finite Automata (NFAs) are closely related to LTSes, the notions of automata theoretic equivalences are often too coarse for many situations in Concurrency Theory. In particular, the notion of a language of an NFA is of linear time; that is, it hides the moment of resolving nondeterministic choice from the external observer. Milner and Park introduced a branching-time notion of equivalence known as bisimilarity. In this concept, an external observer can see when and how nondeterminism is resolved, thus making it finer than language

equivalence.

Despite a change of focus, the paradigm of specification languages and axiomatisations remained. Milner noted that in the bisimilarity setting, Regular Expressions lose their expressiveness. Namely, there exist transition systems that cannot be specified with Kleene's syntax. The answer to that problem involved a more complex language with binders and a recursion operator. Milner provided a complete axiomatisation, following a pattern similar to Salomaa. Since then, process algebraists have been looking into various models, specification languages, and notions of equivalence or similarity within the aforementioned paradigm of axiomatisations.

1.0.3 Behavioural metrics

In many contexts, especially when dealing with probabilistic or quantitative models, focusing on exact equivalence of behaviours 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 shortest 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 for Stark and Smolka's process calculus for specifying Generative Probabilistic Transition Systems.

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