

Doctoral dissertation

*Logical Foundations of Metric Behavioural  
Theory for Markov Processes*

by

Radu Mardare, Ph.D.

Department of Computer Science  
Aalborg University  
Selma Lagerlöfs Vej 300  
9220 Aalborg



# Contents

<b>I</b>	<b>Introduction</b>	<b>7</b>
<b>1</b>	<b>Research Context and Hypothesis</b>	<b>9</b>
<b>2</b>	<b>The Monograph</b>	<b>17</b>
2.1	An Overview . . . . .	17
2.1.1	Preliminary research supporting this Monograph . . . . .	17
2.1.2	Research included in this Monograph . . . . .	18
2.1.3	The publications included in the Monograph . . . . .	20
2.2	The Structure . . . . .	22
<b>3</b>	<b>Preliminaries</b>	<b>29</b>
3.1	Set Theory - Basic Notations . . . . .	29
3.2	Basic concepts of Measure Theory . . . . .	29
3.3	Basic concepts of Topology . . . . .	32
3.4	Metric and Pseudometric Spaces . . . . .	34
3.5	Analytic Spaces . . . . .	36
3.6	Rasiowa-Sikorski Lemma . . . . .	36
3.7	Lindenbaum's Lemma and Non-Compact Logics . . . . .	37
3.8	Markov Processes . . . . .	40
<b>II</b>	<b>Stochastic Process Algebras</b>	<b>47</b>
<b>4</b>	<b>Stochastic-CCS</b>	<b>51</b>
4.1	Introduction . . . . .	51
4.2	Basic Concepts and Notations . . . . .	55
4.3	A minimal Stochastic Process Algebra . . . . .	56
4.4	Structural Operational Semantics . . . . .	60
4.5	Stochastic bisimulation is a congruence . . . . .	67
4.6	Concluding remarks . . . . .	76
<b>5</b>	<b>Stochastic Pi-Calculus</b>	<b>77</b>
5.1	Introduction . . . . .	77

5.2	Stochastic Pi-Calculus . . . . .	78
5.2.1	Syntax . . . . .	79
5.2.2	Rate environments . . . . .	81
5.2.3	The class of indexed measures . . . . .	82
5.2.4	Semantics . . . . .	86
5.3	Stochastic bisimulation . . . . .	94
5.4	Concluding remarks . . . . .	101
<b>6</b>	<b>Metrics for Stochastic Process Algebras</b>	<b>103</b>
6.1	Introduction . . . . .	103
6.2	Distances Between Stochastic Process Algebra Terms . . . . .	104
6.3	Concluding Remarks . . . . .	109
<b>III</b>	<b>Markovian Logics</b>	<b>115</b>
<b>7</b>	<b>Markovian Logics</b>	<b>121</b>
7.1	Markov Processes . . . . .	121
7.2	Syntax of Markovian Logics . . . . .	122
7.3	Semantics of Markovian logics . . . . .	123
7.4	Hilbert-style Axiomatizations . . . . .	127
7.4.1	Axioms of Probabilistic Markovian Logic . . . . .	128
7.4.2	Axioms of Sub-probabilistic Markovian Logic . . . . .	130
7.4.3	Axioms of the General Markovian Logic . . . . .	130
7.5	Finite Model Property and Weak-Completeness . . . . .	133
7.6	Canonical Models and Strong Completeness . . . . .	139
7.7	Concluding Remarks . . . . .	147
<b>8</b>	<b>The Topologies of Markovian Logics</b>	<b>149</b>
8.1	Introduction . . . . .	149
8.2	Preliminaries . . . . .	151
8.2.1	Pseudometric Spaces . . . . .	151
8.2.2	Markov Processes and Markovian Logics - Basic Notations . . . . .	154
8.3	Dynamic-Continuous Pseudometric Bisimulations . . . . .	158
8.4	The Topological Space of Logical Formulas . . . . .	159
8.4.1	The topology of Discrete Markovian Logic . . . . .	160
8.4.2	The topology of Continuous Markovian Logic . . . . .	163
8.5	Concluding Remarks . . . . .	167
<b>IV</b>	<b>Stone Dualities for Markov Processes</b>	<b>173</b>
<b>9</b>	<b>Classic Stone Duality for Markov Processes</b>	<b>177</b>

9.1	Introduction . . . . .	177
9.1.1	A Summary for Experts . . . . .	178
9.2	Stone Duality - An Overview . . . . .	179
9.3	Markov Processes and Markovian Logic . . . . .	181
9.4	Aumann Algebras . . . . .	183
9.4.1	Definition of Aumann Algebras . . . . .	183
9.4.2	Extended Satisfiability Principles for Aumann Algebras . . . . .	185
9.5	Stone Markov Processes . . . . .	187
9.5.1	MPs with Distinguished Base . . . . .	187
9.5.2	Saturation . . . . .	188
9.5.3	Definition of SMP . . . . .	188
9.6	Stone Duality . . . . .	188
9.6.1	From AAs to SMPs . . . . .	189
9.6.2	Construction of $\mathbb{M}(\mathbb{A})$ . . . . .	191
9.6.3	From SMPs to AAs . . . . .	196
9.6.4	Duality . . . . .	197
9.6.5	Duality in Categorical Form . . . . .	198
9.7	Concluding Remarks . . . . .	201
<b>10</b>	<b>Metrized Stone Duality for Markov Processes</b>	<b>203</b>
10.1	Introduction . . . . .	203
10.2	Extending the Duality to Metrized Markov Processes . . . . .	204
10.3	Metrized Markov Processes . . . . .	204
10.4	Metrized Aumann Algebras . . . . .	206
10.5	Metrized Stone Duality . . . . .	207
10.6	A canonic metric for Metrized Aumann Algebras . . . . .	210
10.7	Metric Duality in Categorical Form . . . . .	213
10.8	Concluding Remarks . . . . .	214
<b>V</b>	<b>Computational Aspects of Behavioural Distances</b>	<b>221</b>
<b>11</b>	<b>Behavioural Distances for Markov Chains</b>	<b>227</b>
11.1	Introduction . . . . .	227
11.2	Preliminaries: The Transportation Problem . . . . .	229
11.3	Markov Chains and Bisimilarity Pseudometrics . . . . .	229
11.4	Alternative Characterization of the Pseudometric . . . . .	232
11.5	Exact Computation of Bisimilarity Distance . . . . .	234
11.5.1	Computing the $\lambda$ -Discrepancy . . . . .	234
11.5.2	Greedy Search Strategy for Computing an Optimal Coupling . . . . .	235
11.6	The On-the-Fly Algorithm for Exact Computation . . . . .	240
11.7	Experimental Results . . . . .	244
11.8	Conclusive Remarks . . . . .	248

<b>12</b>	<b>Efficient Computation of Bisimilarity Distances</b>	<b>249</b>
12.1	Introduction . . . . .	249
12.2	The BISIMDIST Library . . . . .	250
12.3	Conclusive Remarks . . . . .	252
<b>13</b>	<b>Compositional Distance for Markov Decision Processes</b>	<b>253</b>
13.1	Introduction . . . . .	253
13.2	Markov Decision Processes and Behavioral Metrics . . . . .	254
13.3	Compositional Operators for MDPs . . . . .	257
13.4	Non-Extensive Compositional Operators . . . . .	258
13.5	Alternative Characterization of the Pseudometric . . . . .	262
13.6	Exact Computation of Bisimilarity Distance . . . . .	265
13.7	A Compositional On-the-Fly Algorithm . . . . .	269
13.8	Concluding Remarks . . . . .	272
<b>14</b>	<b>A Distance for Continuous-Time Markov Chains</b>	<b>273</b>
14.1	Introduction . . . . .	273
14.2	CTMCs and Bisimilarity Pseudometrics . . . . .	275
14.2.1	Bisimilarity Pseudometrics on CTMCs . . . . .	277
14.3	Complexity and Linear Programming representation . . . . .	283
14.3.1	Iterative method . . . . .	283
14.3.2	Linear Program Characterization . . . . .	284
14.4	Alternative Characterization of the Pseudometric . . . . .	287
14.5	Greedy Computation of the Bisimilarity Distance . . . . .	289
14.5.1	Computing the $\lambda$ -Discrepancy . . . . .	289
14.5.2	Greedy Strategy for Optimal Coupling Structures . . . . .	290
14.6	The On-the-Fly Algorithm . . . . .	292
14.7	Experimental Results . . . . .	298
14.8	Conclusive Remarks . . . . .	300
<b>15</b>	<b>Total Variation Distances of Semi-Markov Chains</b>	<b>303</b>
15.1	Introduction . . . . .	303
15.2	Preliminaries . . . . .	305
15.3	Semi-Markov Chains and Trace Distance . . . . .	307
15.4	Trace Distance and Probabilistic Model Checking . . . . .	309
15.4.1	Model Checking for MTL Formulas . . . . .	312
15.4.2	Model Checking for Timed Automata . . . . .	316
15.5	General Convergence Criteria . . . . .	319
15.6	An Approximation Algorithm . . . . .	330
15.6.1	Fixed Point Characterization of the Discrepancy . . . . .	331
15.6.2	Construction of an Optimal Coupling Structure . . . . .	336
15.7	Conclusive Remarks . . . . .	338

*CONTENTS*

7

**16 Dansk résumé**

**347**



**Part I**  
**Introduction**



# Chapter 1

## Research Context and Hypothesis

The investigation of the fundamental relations between computational (theoretical) models and real, physical systems goes back to the first works of Turing on morphogenesis [42] and artificial intelligence [41]. With the development of computer science, the questions related to the existence of the *real computation* in nature specialised and revealed concrete computational phenomena in the real world. In such a context, also due to the evolution of the technology, it became increasingly important to model and understand the possible interactions between a real and a human-made computational systems. Such systems are called *cyber-physical systems*.

Complex cyber-physical systems in areas as diverse as aerospace, automotive engineering, chemical processes, civil infrastructures, energy, healthcare, manufacturing, transportation, and consumer appliances, are often represented as *stochastic processes* to model ignorance, uncertainty or inherent randomness.

A stochastic process uses real-valued parameters (probabilities of transitions, distributions characterizing the residence-time in a state, etc.) to abstract missing information about the analyzed system. Such systems are frequently modular in nature, consisting of parts which are systems in their own right. Their global behavior depends on the behavior of their parts and on the links which connect them.

To define the concept of behavior for a stochastic or probabilistic systems were introduced various concepts of *stochastic/probabilistic behavioral equivalences*, which are relations that equates systems with identical behaviors. These concepts follow the original definition of *probabilistic bisimulation* introduced by Larsen and Skou [29]. On the other hand, to characterize the behavior of a system, logics such as *probabilistic modal logics* or *probabilistic-timed-temporal logics* have been used to encode characteristic behavioral properties [2, 17, 18, 29, 43].

Since these behavioral equivalences relate stochastic systems with identical probabilistic behaviors, they are too “exact” for most purposes. In applications, one needs instead to know whether two processes that may differ by a small amount in the real-valued parameters have sufficiently similar behaviors.

The main motivation of the research presented in this monograph was to study a

relaxation of the notion of behavioral equivalence for stochastic systems that supports a metric theory of *behavioral “nearness”* which measures the dissimilarities between two models. Since the equivalence-based theory enjoys interesting properties in relation to logics, our intention was to verify to what extent a metric theory can support similar properties. However, the development of the behavioural metric was mainly driven by a series of topological properties that intuitively any approximation theory for stochastic systems must satisfy. The study of the logical foundation of this metric theory provided concepts and techniques for designing analysis tools for cyber-physical systems.

A series of fundamental questions regarding the similarity of behaviours arise from the attempts of analysing concrete stochastic systems.

In Figure 1.1 is represented a famous modelling problem from systems biology [36]: the cell-cycle switch, a Nobel prize winning network, which is a fundamental mechanism in all Eukaryotic cells.

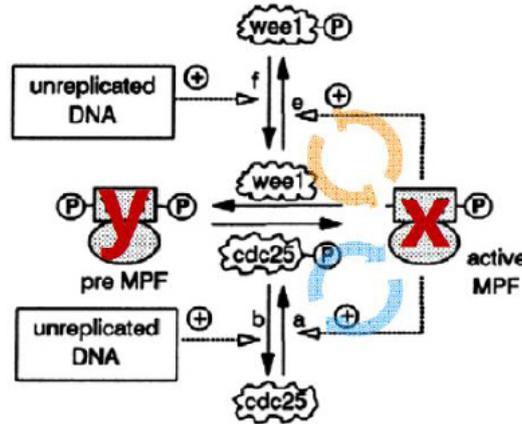


Figure 1.1: The Cell-Cycle Switch

In Figure 1.2 we have, according to [8], three large continuous-time Markov chains that abstract, at different levels, the chemical representation of Figure 1.1. The intention is to study the cell-cycle switch by analyzing its models and use these results to make trustful predictions about the stochastic behavior of the cell-cycle switch. Since  $m_1$ ,  $m_2$  and  $m_3$  are increasingly accurate stochastic models of the chemical network, we expect that the predictions we can make using each model will be increasingly accurate as well, in the sense that the errors obtained from  $m_3$  are significantly less relevant than the ones we get from  $m_1$ .

A behavioural metric is expected, in this case, to provide the framework for verifying that  $m_3$  is indeed “more similar” to the real system than  $m_1$  or  $m_2$ . Concretely, such a metric shall be used to verify, for instance, the following properties:

- *if the models show oscillatory behaviours, then the real system oscillates;*

- if the approximants have steady states, then a similar property can be indeed inferred for the real system.

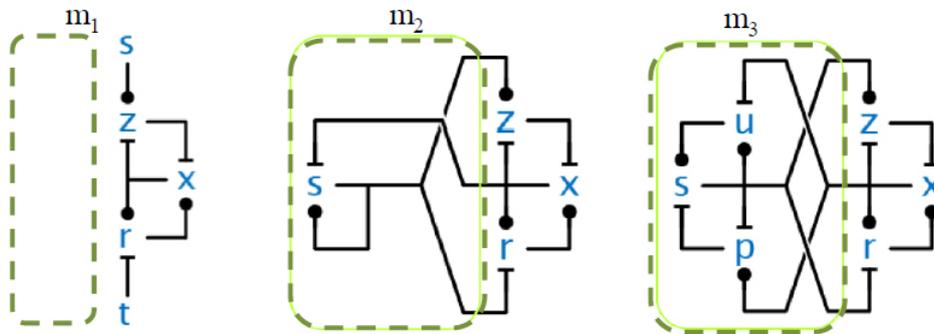


Figure 1.2: A sequence of increasingly accurate models for the Cell-Cycle Switch

In other situations, as represented in Figure 1.3, the model  $n_1$  of the cell-cycle switch is too complex for certain analyses, so that simpler versions of it, such as  $n_2$  and  $n_3$ , have been proposed [8]. In this context, a behavioural metric must guarantee that these are “good approximations” of  $n_1$  and that

- if the simplified models (approximations) show certain behaviour in a given context, then the approximated system has a very similar behaviour in the same context.

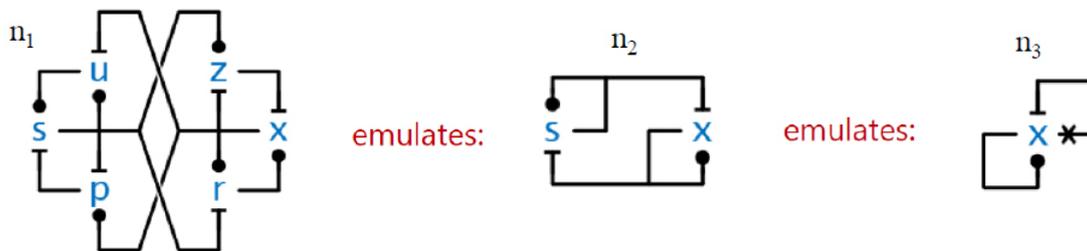


Figure 1.3: A model and its refinements

Another interesting case of a cyber-physical system is that of a pacemaker interacting with the heart. To design such a system, one has firstly to abstract the electrical mechanisms of the heart and to design the pacemaker such that to guarantee its optimal functionality in whatever imaginable stochastic context.

According to [30], the basic electrical mechanism of the heart can be represented as in Figure 1.4, where the graph represents the abstraction of the electrical conduction system of the heart. Abstracted as a stochastic system, the electrical system of the heart should interact with the components of the pacemaker, themselves represented as a stochastic systems, in [12]. The pacemaker system contains five interactive components,

represented as stochastic timed-automata in Figure 1.5. These are behavioral specifications with hard timing constraints that the components of the pacemaker must guarantee. However, the design process for such devices is an iterative procedure, where component specifications are repeatedly being updated and refined until eventually a final implementation may be readily (or even automatically) derived. To support such a refinement process, one needs to evaluate the dissimilarities between various specifications and guarantee that

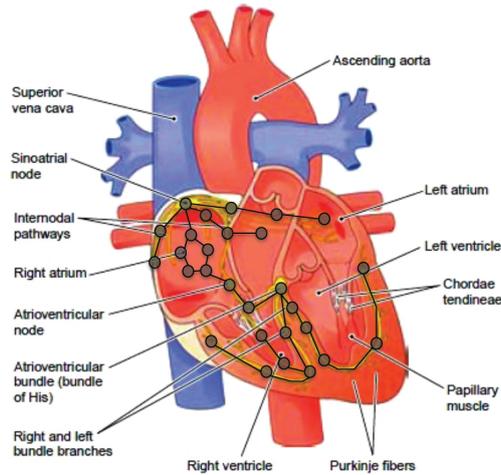


Figure 1.4: Electrical conduction system of the heart

- *by replacing a component (specification) with another one that is sufficiently similar, we get an overall behavior (as a result of the interaction with the other components) that is not very different from the original one.*

Another interesting example of stochastic system is described in Figure 1.6, where  $m$  represents a wireless communication protocol in a hostile environment, where the probability of a safe communication is  $p$ . The models  $m_i$  are better and better models of  $m$  obtained from repeated experiments, i.e., the sequence  $(e_i)_{i \in I}$  converges to 0. Observe that  $m_i$  and  $m$  always differ by a value  $e_i$  that eventually will be very small. Nevertheless, it prevents the two systems from being behavioural equivalent. This limitation can be compensated by the use of behavioural metric: the infinite sequence  $(m_i)_{i \in I}$  of models should converge to  $m$  in the topology induced by the metric. This means that we expect that the distance between  $m_i$  and  $m$  should converge to 0. This example emphasizes a series of topological properties that we expect to be satisfied by an appropriate behavioural distance. For instance, we expect that the distance is defined such that it will answer questions such as

- *if each model  $m_i$  has a certain property, is this property also satisfied by the real system  $m$ ?*

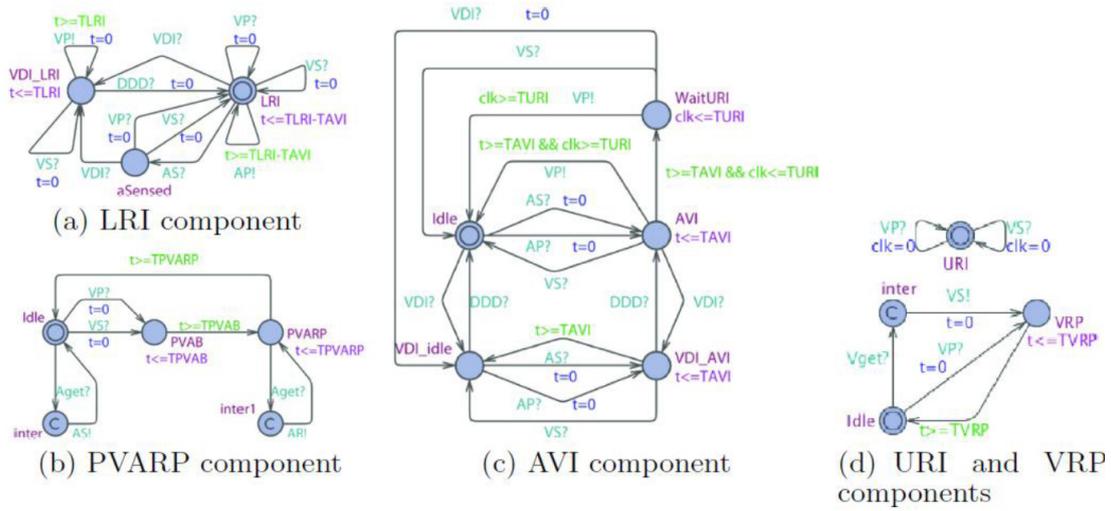


Figure 1.5: The components of a Pacemaker

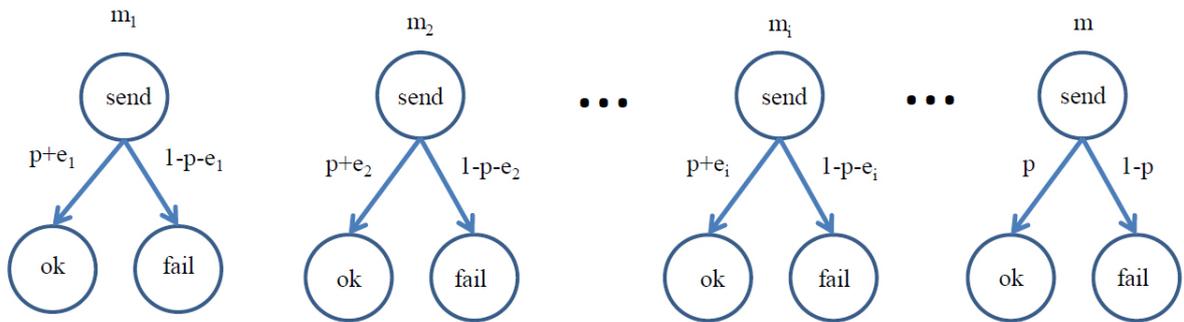


Figure 1.6: Models of a wireless communication protocol

All these problems are cornerstones for modelling, simulation and verification of complex systems establishing the limits of knowledge transfer between models and the systems being modelled. Due to the insufficient understanding of the role of a behavioural metric, the research in this field evolved on two parallel tracks. On one hand, computability and complexity issues were studied for metrics defined as the fixed points of some operators, usually based on the Kantorovich metric [4, 11, 14, 22, 43–45]. On the other hand, the research on model verification focused on metrics from a logical perspective [17, 37], where the distance between two models is defined as the variation difference with respect to the degree of satisfaction of logical properties.

At the moment the research presented in this monograph was initiated, there was no result known to connect the two research directions and consequently, no authentic approximation theory for stochastic systems that allows one to answer questions similar to the ones emphasized by our examples. In fact, as we will emphasize in this monograph, some particular measures used in the literature fail to correctly reflect our intuition of “nearness”.

In this context, we have taken the challenge of developing an appropriate theory that will support such analysis for a general concept of stochastic systems. There are various levels of abstraction that one can consider in the definition of a stochastic processes:

1. the state space can be modelled by using particular types of structures that can vary from discrete finite spaces to topological or measurable spaces;
2. the indeterminacy can be modelled by using probability or sub-probability distributions over the state space to describe the probability of transitions or by assuming certain distributions to characterize the time durations between transitions.

To cope with this level of generality, we have adopted a general notion of *Markov process* (MP), following [18, 20, 37], which encapsulates various notions of Markovian stochastic systems such as *Markov chain* with discrete or continuous time [27]. A Markov process in this understanding, is a coalgebraic structure that encodes stochastic behaviors and can be seen, following the *transition-systems-as-coalgebras* paradigm [16, 28], as a generalisation of the notion of transition system: a transition system associates to each state of a system an action-indexed set of functions over the state space; functions with *boolean values* define *labelled transition systems* while *probabilistic distributions* define labelled Markov processes [7, 17, 18] and *Harsanyi type spaces* [26, 35]; further, one can associate to states arbitrary distributions to also encode continuous-time systems. This paradigm is particularly appropriate when one is interested in systems with a complex state space where transitions cannot be represented from one state to another, but from a state to a measurable set of states or to a (topological) neighbourhood. In earlier papers, these models were called *labeled* Markov processes to emphasize the fact that there were multiple possible actions that can label the transitions, but in this Monograph we will suppress the labels, as they do not contribute any relevant structure for our results.

To complete this research, one needs to deeply understand the metric-logical principles behind the concept of metrized behaviour for MPs. The main hypothesis that motivated and driven our research presented in this Monograph can be summarized as follows.

- (i) *There exists a fundamental duality between the metric space of stochastic systems and the metric space of (stochastic) logical properties; this generalizes the duality between stochastic systems and stochastic properties in the way the behavioural metrics generalize the behavioural equivalences.*
- (ii) *The aforementioned duality reflects fundamental approximated reasoning principles that are for the metric theory what the Boolean reasoning principles are for the behavioural theory.*
- (iii) *The approximate reasoning principles are the basis for an approximation theory for stochastic systems where concrete questions related to the knowledge transfer between models and the systems being modelled can be formally addressed and answered.*

- (iv) *The metric reasoning for the concrete cases of continuous-time Markov chains and stochastic timed-automata can be devised with proof systems, deductive techniques and algorithmic methods in a similar way in which the bisimulation relation has been axiomatized and shown computable.*



# Chapter 2

## The Monograph

### 2.1 An Overview

This Monograph contains a series of research results achieved by the candidate in the last years, focussed on topics related to the logical foundation of metric behavioural theory for Markov processes. This research addresses the hypothesis listed in the previous section and it is summarized in what follows.

This Monograph is based on a series of articles listed below but it is not, however, just the sum of these articles. It was conceived as a proper monograph that constructs iteratively the arguments from the first to the last Chapter. It also relies on previous research results that have not been included explicitly in this Monograph. More details about the structure of this Monograph can be found in Section 2.2.

#### 2.1.1 Preliminary research supporting this Monograph

The research included in this Monograph was initiated during the period the candidate worked as a researcher at the Microsoft Research CoSbi center (Italy), where he had studied the logical and computational structures of biological and physical systems. Many of these firsts results are not included in the Monograph, nor the practical results that the candidate published in fields such as systems biology, membrane computing, colonies of synchronizing agents, or quantum information. Nevertheless, this research experience was the promoter and a powerful impetus for the results summarized in this Monograph.

A key achievement obtained while working at CoSbi has been the use of logical analysis and model-checking techniques for validating models and predicting behaviours of natural systems. The paper [34] of the candidate was one of the first papers to introduce model-verification techniques for biological processes; a technique that has now become standard in the field. The candidate has also used probabilistic model-checking techniques to prove oscillatory behaviours for bio-chemical systems and pinpoint the qualitative differences between stochastic models and the nondeterministic models based

on ordinary differential equations [5]. In the same period Mardare has coordinated a project aiming to develop a statistical model-checking technique to predict properties of very large stochastic systems by combining Gillespie-based simulations with quantitative (Monte Carlo-based) model checking and statistical analysis. All these results contributed to consolidate Mardare's intuition about the role of logical analysis in the study of cyber-physical systems.

Mardare has tested the potentiality of applying axiomatized logics in the field of membrane computing (P-systems), which studies computation in hierarchical formal systems. The candidate has introduced a completely axiomatized logic for membrane systems and emphasized the advantage of using theorem proving techniques for such applications [9]. This logic is designed to specify local information at membrane-levels and to use this information for proving global properties of the overall computation. This work inspired a new research direction in the field and eventually guided the candidate towards the problems of compositionality and modularity in reasoning, which is a key issue in this Monograph.

The first intuitions regarding the necessity of a metric behavioural theory for stochastic systems arose in the period when Mardare developed a model inspired by the intercellular and intracellular interactions in tissues called *colonies of synchronizing agents* (CSA), which generalizes the concept of cellular automata [10, 33]. The key novelty of this model is the focus on local interactions between arbitrary populations of agents, which nevertheless produces robust behaviour at a global level. The paper that introduced the CSA model [33] was ranked between the 25 "hottest papers" in the journal *Theoretical Computer Science*. Based on CSA models the candidate proposed a definition for the concept of robustness for a class of systems [32]. Despite its importance for the study of biological systems, this concept had thus far evaded all attempts at a formal definition. Eventually, it has been demonstrated that this definition recovers many key concepts in ecology, such as stability, reliability, resistance and resilience, as well as the concepts of species and mutants in biology. This research experience was a promoter for the ideas of metric reasoning that emerged from the attempt of catching the concept of robustness for a stochastic system.

### 2.1.2 Research included in this Monograph

The first investigations on the general structure of stochastic systems were done in collaboration with **Luca Cardelli** (Microsoft Research Cambridge, UK). This concluded with a series of publications regarding the structural operational semantics for stochastic process algebras and the first observations regarding the possibility to provide a metric semantics to stochastic processes [I,V,XV].

The algebraic structure of stochastic processes leads to logical analysis. In an attempt of comprehending the logical structure of stochastic processes, the candidate together with Cardelli and **Kim G. Larsen** (Aalborg University, Denmark) pioneered the model theoretical studied of the Markovian logics proposing axiomatic systems for var-

ious semantics and studying metatheoretical aspects that include a metric metatheory [XI,XIII,XIV].

In the process of identifying the most general concept of stochastic systems, Mardare learned about the Markov processes in the definition of **Prakash Panangaden** (McGill University, Canada). Mardare, Larsen and Panangaden teamed up with the intention of apprehending the logical-topological principles of the Markovian logics defined for labelled continuous-space Markov processes [XII].

The aforementioned research opened the perspective of a Stone-like duality for Markov processes, which was eventually demonstrated with the contribution of **Dexter Kozen** (Cornell University, USA). This work, that represent one of the major contribution of our Monograph, identified the concept of Aumann algebra, that is the Boolean algebra with operators corresponding to Markovian logic [VI]. A residual of the duality is the strong-completeness result for the Markovian logic [VIII]. This result completes the previous work by Goldblatt who studied strongly-complete axiomatizations for logics on coalgebras defined by polynomial functors constructed from a standard monad on the category of measurable spaces.

At the time the duality research was concluded, even if it has clarified a series of open problems regarding the relation between Markov processes, logics, topology and Boolean algebras, the issues related to metric reasoning for Markov processes was not integrated. In collaboration with Panangaden and Kozen, Mardare has continued the investigation of the mathematical structures that support the development of the theory of Markov processes in order to understand if the concept of bisimilarity can be relaxed to a concept of pseudometric on Markov processes while preserving the elegant connection with the logic. This is how they arrived to the concept of metrized Markov process and metrized Aumann algebra [II]. In this context, a new type of duality was proven: a Stone-like duality that extends the classic Stone duality for Markov processes. This research provides a novel perspective on the research of behavioural distances and clarifies a series of problems regarding the relation between the Stone topology induced by the classic Stone duality and the open-ball topology defined by a “proper” behavioural distance.

All the theoretical research has been done in parallel with practically-oriented investigations aimed to comprehend particular examples of behavioural distances for discrete-time and continuous-time Markov chains or Markov decision processes. This research is the result of a collaboration with Larsen, **Giorgio Bacci** (Aalborg University, Denmark) and **Giovanni Bacci** (Aalborg University, Denmark). Bacci and Bacci were initially hired as research assistants for Mardare’s fellowship Sapere Aude. This part of our research was mainly oriented towards computability and algorithm developments, with the intention of developing tools for metric reasoning [III,IV,VII,IX,X]. This investigation played a central role in the research landscape described in this Monograph, as it provided the case studies on which Mardare based his research.

The research presented in this Monograph was supported by the prestigious fellowship *Sapere Aude: DFF Young Researchers Grant* and by an *Individual Research Grant*,

both awarded to Mardare by *The Danish Council for Independent Research* in 2010. The research was partially supported also by the *VKR Center of Excellence MT-LAB* and by the *Sino-Danish Basic Research Center IDEA4CPS*.

### 2.1.3 The publications included in the Monograph

The research results presented in this monograph have been mainly published in the papers listed chronologically below.

- [I] L. Cardelli, R. Mardare. *The Measurable Space of Stochastic Processes*. *Fundamenta Informaticae*, FI volume 131(3-4), pages:351-371, 2014.
- [II] D. Kozen, R. Mardare, P. Panangaden. *A Metrized Duality Theorem for Markov Processes*. In Proc. of 30th Conference Mathematical Foundation of Programming Semantics, MFPS2014, *Electronic Notes in Theoretical Computer Science*, ENTCS vol. 308, pages: 211-227, 2014.
- [III] G. Bacci, G. Bacci, K.G. Larsen, R. Mardare. *On-The-Fly Computation of Bisimilarity Distances*. *Logical Methods in Computer Science*, LMCS, to appear, 2015.
- [IV] G. Bacci, G. Bacci, K.G. Larsen, R. Mardare. *On the Total Variation Distance of Semi-Markov Chains*. under submission 2014.
- [V] L. Cardelli, R. Mardare. *Stochastic Pi-calculus Revisited*. In Proc. of 10th International Colloquium Theoretical Aspects of Computing, ICTAC 2013, LNCS 8049, pages: 1-21, 2013.
- [VI] D. Kozen, K. G. Larsen, R. Mardare, P. Panangaden. *Stone Duality for Markov Processes*. In Proc. of 28th Annual ACM/IEEE Symposium on Logic in Computer Science, LICS 2013, IEEE Computer Society, pages: 321-330, 2013.
- [VII] G. Bacci, G. Bacci, K.G. Larsen, R. Mardare. *Computing Behavioral Distances, Compositionally*. In Proc. of 38th International Symposium Mathematical Foundations of Computer Science, MFCS 2013, *Lecture Notes in Computer Science*, LNCS 8087, pages: 74-85, 2013.
- [VIII] D. Kozen, R. Mardare, P. Panangaden. *Strong Completeness for Markovian Logics*. In Proc. of 38th International Symposium Mathematical Foundations of Computer Science, MFCS 2013, *Lecture Notes in Computer Science*, LNCS 8087, pages: 655-666, 2013.
- [IX] G. Bacci, G. Bacci, K.G. Larsen, R. Mardare. *The BisimDist Library: Efficient Computation of Bisimilarity Distances for Markovian Models*. In Proc. of 10th International Conference Quantitative Evaluation of Systems, QEST 2013, *Lecture Notes in Computer Science*, LNCS 8054, pages: 278-281, 2013.

- [X] G. Bacci, G. Bacci, K.G. Larsen, R. Mardare. *On-the-Fly Exact Computation of Bisimilarity Distances*. In Proc. of 19th International Conference Tools and Algorithms for the Construction and Analysis of Systems, TACAS 2013, Lecture Notes in Computer Science, LNCS 7795, pages:1-15, 2013.
- [XI] R. Mardare, L. Cardelli, K. G. Larsen. *Continuous Markovian Logics - Axiomatization and Quantified Metatheory*. Logical Methods in Computer Science, LMCS vol.8(4):1-28, 2012.
- [XII] K. G. Larsen, R. Mardare, P. Panangaden. *Taking it to the Limit: Approximate Reasoning for Markov Processes*. In Proc. of 37th International Symposium Mathematical Foundations of Computer Science, MFCS 2012, Lecture Notes in Computer Science, LNCS 7464, pages: 681-692, 2012.
- [XIII] L. Cardelli, K.G. Larsen, R. Mardare. *Continuous Markovian Logics - From Complete Axiomatization to the Metric Space of Formulas*. In Proc. of Computer Science Logic, CSL2011, LIPIcs 12 Schloss Dagstuhl - Leibniz-Zentrum fuer Informatik, vol.12, pages: 144-158, 2011.
- [XIV] L. Cardelli, K.G. Larsen, R. Mardare. *Modular Markovian Logic*. in Proc. of the 38th International Colloquium on Automata, Languages and Programming, ICALP 2011, Lecture Notes in Computer Science, LNCS 6756: 380-391, 2011.
- [XV] L. Cardelli, R. Mardare. *The Measurable Space of Stochastic Processes*. In Proc. of Seventh International Conference on the Quantitative Evaluation of Systems, QEST2010, IEEE Computer Society, pages: 171-180. 2010.

## 2.2 The Structure

The Monograph is divided into five distinct Parts, each reflecting a particular aspect of the theory we are developing. The Parts complete each other and contribute altogether to the description of the mathematical universe of Markov processes. Part I, II, III and IV are dedicated to theoretical developments, while in Part V the theory is applied for developing dedicated algorithms.

Each Part is organized in several chapters, each chapter summarizing research results published in one or more articles. In each Chapter the notations are uniformly and consistently used. However, the notation might differ from a Chapter to an other. When this is the case, the Chapter contains an introductory section where the notation is fixed. Each Chapter is written with the intention of being, as much as possible, self-contained. Nevertheless, we did not repeat basic definitions or concepts that are repeatedly used in more Chapters. Instead, many of these recurrent concepts have been collected in Chapter 3: Preliminaries.

Each Part contains an independent bibliography.

### Part I: Introduction

**Chapter 3: Preliminaries** fixes the basic notation and presents a series of concepts and results from set theory, measure theory, topology and model theory that are used in this Monograph. Some of these results are original being developed for the particular settings in which they are applied. This is, for instance, the case with Section 3.7 where we demonstrate a series of results from model theory that, at the best of our knowledge, have not been used nor stated before, despite their potentiality in applications.

### Part II: Stochastic Process Algebras

The research summarized in Part II represents the first step towards a deeper understanding of the Markov processes. This research, inspired by the experience of the applicant with real examples of stochastic systems from Systems Biology, is an attempt to answer a long-standing open question in concurrency theory regarding the possibility to develop a stochastic process algebra on the mathematical pattern used to develop nondeterministic process algebras. The major challenge that the researchers had to face was that, in the attempt of providing a general structural operational semantic (GSOS) format for a stochastic or probabilistic calculus, almost always one arrives to prove anomalies or is required to provide an infinite set of behavioural axioms to handle measurability issues. It is a classic example in this respect Priami's stochastic Pi-calculus [38], where the parallel composition of processes is not associative.

In the research presented in Part II of this monograph we have tried to explain this situation from a mathematical perspective and offer a solution to the aforementioned

problem. In our opinion, the problem is generated by the fact that most of the researchers who have tried to adapt the Turi-Plotkin GSOS format [40] for the stochastic or probabilistic case used the wrong functors. The Turi-Plotkin GSOS format was designed for a freely generated monad, while the case of the stochastic or probabilistic process algebra the underline monad is equational, being defined by the structural congruence. This is essential to understand, because a measurable set of processes must be closed under structural congruence. Consequently, the measurable space of processes is not organized by the powerset, but by the  $\sigma$ -algebra generated by the structural congruence; thus, pointwise probabilistic or stochastic transitions between processes shall not be defined. Instead, the structure of the GSOS should include transitions from processes to distributions over the measurable set of processes. And measure theory shall play a central role in the development of process algebras.

We demonstrate that our idea solve, on block, the problems related to the definition of SOS systems for probabilistic and stochastic process algebras.

**Chapter 4** is a proof of concept to support our ideas. Here we develop such a GSOS for a finite fragment of Stochastic-CCS in order to understand how the algebraic structure of processes can be lifted to the level of distributions over the space of processes. Central is the observation that these structures are, in fact, Markov processes in the sense of [37].

**Chapter 5** develops further this idea and demonstrates that the work presented in the previous Chapter can be, in fact, extended to the entire Pi-calculus that includes the name passing mechanism, the private name management and recursion (in the form of replication). This work revealed a series of unexpected results from the process of lifting the complex algebraic structure of the Pi-calculus to the level of distributions.

In addition to providing an example of a correct and canonical way to define GSOS formats for stochastic process algebras, our idea of organizing the space of processes as a measurable space of structural congruence-closed sets and considering the distributions over this space, gave us a the technical tools needed to go beyond the concept of bisimulation. In **Chapter 6** we show how one can extend this setting to define pseudometrics between processes that measure the dissimilarities between non-bisimilar processes. Thus, we propose a simple way of discussing about the robustness of the processes.

The results summarized in Chapter 6 determined us to look deeper into the relation between bisimulation and metrics for Markov processes and inspired most of the research that we will further discuss in this Monograph.

## Part III: Markovian Logics

Inspired by the work we have done on stochastic process algebras, we started to be interested in the logical properties of these processes. We realized that, in fact, the setting of stochastic process algebras is too restrictive and we focused our interest towards

the Markov processes in general. From the lesson we learned while studying process algebras, we realized that the discrete processes are not sufficiently robust from a mathematical perspective and that, instead, one needs a definition of Markov processes with continuous state space. This is how we eventually arrived to the concept of Markov process (MPs) introduced by Panangaden et al. [18, 37] that we use in the rest of this Monograph.

Regarding the logical properties of MPs, Desharnais et al. [18, 37] have demonstrated that the logical equivalence induced by a simple version of probabilistic modal logic on the class of MPs coincides with bisimulation equivalence. Moreover, the same equivalence can already be induced by the negation-free fragment of this logic. These results were very surprising because for the nondeterministic transition system a similar characterization is only possible for systems that are image-finite.

Being this exciting context, we have decided to focus our research on the study of these probabilistic/stochastic logics, that we have called Markovian logics, and try to develop a model theory for them that will help us understanding their relation to the space of MPs. **Chapter 7** summarizes our results obtained for sub-probabilistic, probabilistic and stochastic MPs. We have extended the proof of Desharnais and shown that it works in larger settings. We have developed complete axiomatizations for these logics and initially proved the weak-completeness of the axiomatic systems<sup>1</sup>. We have proven that they enjoy the finite model property and we have constructed canonical models based on the maximal-consistent sets of formulas.

The construction of the canonical models helped us explaining why the negation-free fragment of the Markovian logic can characterize the bisimilarity of MPs. This is because the mentioned fragment forms a field in the  $\sigma$ -algebra generated by the (sets of models of the) logical formulas and because a finitely-additive and countable-subadditive set function on a field can be uniquely extended to a measure on the  $\sigma$ -algebra generated by that field. These arguments emphasized one more time the central role of topology and measure theory in the study of MPs.

Last but not least, involving some topological techniques in model theory, due to Rasiowa and Sikorski [39], we succeed to also develop a strongly-complete axiomatization for the Markovian logics, thus improving the previous results of [24, 46].

In **Chapter 8** we return to the problem of behavioural distances for MPs and try to address a general question.

Assume the Markovian logics with their semantics and suppose that we have a “well-behaved” behavioural distance. Let  $\varphi$  be a logical property and  $(m_i)_{i \in \mathbb{N}}$  be a sequence of MPs that converges to a limit MP  $m$  in the topology of the distance.

*If we can prove that for each  $i \in \mathbb{N}$ ,  $m_i$  satisfies  $\varphi$ , can we also claim that  $\varphi$  is satisfied at the limit  $m$ ?*

---

<sup>1</sup>The Markovian logics are non-compact and for this reason the weak-completeness does not imply the strong completeness.

This is a question that in natural sciences is implicitly assumed to have a positive answer since we often take better and better approximations of a system, verify properties on the approximations and infer the results to the approximated system.

In Chapter 8 we demonstrate that for the Markovian logics this question does not have always a positive answer. We classify the logical properties from this perspective and identify major differences between probabilistic and stochastic systems.

## Part IV: Stone Dualities for Markov Processes

The results presented in the previous Part of the Monograph motivated us to look deeper into the relation between bisimulations on MPs, consistent formulas of Markovian logics (seen as the sets of their models) and the possibility of defining a behavioural pseudometric on MPs that guarantees certain topological properties that we expect to see when we speak about approximations of MPs. This leads towards dualities.

The classical Stone duality theorem states that for every Boolean algebra one can construct a certain kind of topological space (called a Stone space) and for every Stone space one can construct a Boolean algebra. If one goes back and forth in either direction one gets the “same” object back. These things were known in the 1930s and extensions to other kinds of mathematical objects are now numerous. In computer science there are many examples as well: forward and backward flow in static analysis, state-transformer semantics and predicate-transformer semantics, a probabilistic analogue of the latter due to Kozen and, in general between transition systems and logics for reasoning about them. Such a relationship is called a Stone-type duality. In the case of logics and transition systems it subsumes completeness theorems.

For this reason, after proving the completeness theorems for Markovian logics we decided that, in order to understand deeper the relation between logics and bisimulation we should try to identify a Stone-like duality for MPs. We proved such a Stone-type duality for Markov processes and these results are presented in **Chapter 9**. This gave us a stronger completeness theorem than was hitherto known for any Markovian logic or similar probabilistic modal logic. The work was far from a routine extension: we had to replace many of the standard ideas and had to establish some rather intricate lemmas relying on results from measure theory, topology and model theory, in order to complete the program.

The Stone duality presented in Chapter 9 revealed for us deeper relations between bisimulation and logic in the sense that the Stone topology induced by the Boolean core of Markovian logic has the property that its separability relation refines any concept of bisimulation on MPs. This observation was a breakthrough that helped us understanding better what should be the role of a behavioural pseudometric in such a context. We have understood at this point what are the right conditions that characterize a distance that behaves properly from the topological perspectives of Chapter 8. Following this intuitions, we arrived to the concept of metrized Markov process which is, essentially, an

MP devised with an axiomatized behavioural distance. This axiomatization guarantees an enforcement of the concept of dynamic-continuity studied in Chapter 8. All these results are comprised in **Chapter 10** where we proposed a metrized Stone duality for Markov processes. To summarize, the concepts in Chapter 10 establish the most general conditions that any behavioural metric and any metric semantics for Markov processes should satisfy in order to guarantee a correct “behavioural nearness”.

## Part V: Computational Aspects of Behavioural Distances

In parallel with the theoretical results presented in the previous parts of this Monograph, in order to enforce our intuitions about what a behaviour distance should be, we have also done more concrete research centered on key studies. In the Part V of the Monograph we present, in evolution, some of the more practical research done in these years. Mainly, we dedicate this part to a class of algorithms that we have developed and implemented to compute some concrete behavioural distances for some particular classes of Markov processes.

In **Chapter 11** we approach the pseudometric introduced by Desharnais et al. [19] for discrete-time Markov chains (MCs). At the time we start working on this, van Breugel et al. [44] have presented a fixed point characterization of this pseudometric and several iterative algorithms have been developed in order to compute its approximation up to any degree of accuracy. In this Chapter, we propose an alternative approach, which allows us to compute the pseudometric exactly and efficiently in practice. This is inspired by the characterization of the undiscounted pseudometric using *couplings*. We aim at finding an optimal coupling using a greedy strategy that starts from an arbitrary coupling and repeatedly looks for new couplings that improve the discrepancy function. This strategy will eventually find an optimal coupling. We use it to support the design of an on-the-fly algorithm for computing the exact behavioural pseudometric that can be either applied to compute all the distances in the model or to compute only some particular distances. By using an on-the-fly approach, we avoid to exhaustively explore the state space. Instead, we only consider those fragments that are needed in the local computation. The efficiency of our algorithm has been evaluated on a significant set of randomly generated Markov chains. The results show that our algorithm performs orders of magnitude better than the corresponding iterative algorithms in [14, 22]. Moreover, we provide empirical evidence for the fact that our algorithm enjoys good execution running times.

Our on-the-fly technique was proven to be sufficiently robust for even more complex models, such as the Markov Decision Processes with rewards (MDPs), for which a similar pseudometric was proposed in [22]. In **Chapter 12** we present the BISIMDIST library, composed of two Mathematica packages which implement our on-the-fly algorithm for computing the bisimilarity distances both for MCs and MDPs. BISIMDIST is available

at <http://people.cs.aau.dk/~mardare/tools.htm> together with simple tutorials presenting useful examples that show all the features of the library.

However, the case of MDPs is interesting in practice also because there are many examples reflecting the compositional features of such models. Realistic models are usually specified compositionally by means of operators that describe the interactions between the subcomponents. These specifications may thus suffer from an exponential growth of the state space. Moreover, when we consider approximate behaviors, the concept of distances requires *non-extensive* composition operators [19]. In **Chapter 13** we study to which extent compositionality on MDPs can be exploited in the computation of the behavioral pseudometrics of [22], hence how the compositional structure of processes can be used in an approximated analysis of behaviors. This research has been inspired by our experience with probabilistic and stochastic process algebras and by our study of the interaction between the operational semantics of such an algebra and the behavioural pseudometrics, detailed in Part I of this Monograph. As a result, we provide an algorithm to compute the bisimilarity pseudometric by exploiting both the on-the-fly state space exploration and the compositional structure of MDPs built over “safe” operators. Experimental results show that the compositional optimization yields a significant additional improvement on top of that obtained by the on-the-fly method.

In **Chapter 14** we push further the idea of the on-the-fly algorithm for computing the behavioural distances, and we move from discrete-time models to continuous time models. Continuous-time Markov chains (CTMCs) are one of the most prominent models in performance and dependability analysis, constituting the underlying semantics of many modeling formalisms for real-time probabilistic systems. We extend the distance considered in the previous Chapters to CTMCs and we show that it can be computed in polynomial time in the size of the CTMC. This is obtained by reducing the problem of computing the distance to that of finding an optimal solution of a linear program that can be solved using the ellipsoid method. However, our linear program characterization has a number of constraints that is bounded by a polynomial in the size of the CTMC. This, in particular, allows us to avoid the use of the ellipsoid algorithm in favor of the simplex or the interior point methods. Nevertheless, we propose to follow an on-the-fly approach for computing the distance, by relying on the concept of coupling structure generalized for CTMCs. With respect to the previous algorithms, in this Chapter we have a series of innovations required by the fact that we need, in addition, to handle the information regarding the residence-time in a state of the system.

The first proposals of behavioral distances in the literature are based on the Kantorovich metric and are branching-time. In **Chapter 15** we consider a linear-time metric, motivated by the fact that in many applications, such as in systems biology, modeling/testing, and machine learning, the system to be modeled cannot be internally accessed, but only tested via observations performed over a set of random executions. In this respect, we introduce a general class of models, the semi-Markov chains (SMCs),

which are continuous-time probabilistic transition systems where the residence time on states is governed by generic distributions on the positive real line. SMCs subsume many probabilistic models including the discrete-time Markov chains and the continuous-time Markov Chains. Regarding the behavioural distance, we study the total variation between probability measures induced by an SMC over infinite timed traces, which corresponds to the largest possible difference between the probabilities that the measures assign to the same event. In applications, the events are specified either as metric temporal logic (MTL) formulas [2, 3], or languages accepted by timed automata (TAs) [1]. We prove that each of these classes of specifications characterize the total variation.

However, the problem of computing the total variation is known to be NP-hard already for the case of MCs [13, 31] and there is no proof that it is, in fact, decidable. In this context, in Chapter 15, we also prove that the problem of approximating the total variation distance with arbitrary precision is computable. This is done providing two sequences that converge from below and above to the total variation distance. Our results are based on a duality that characterizes the total variation between two measures as the minimal discrepancy associated with their couplings. The computability of the converging sequences provides us with a computable procedure to approximate the total variation distance on SMCs with arbitrary precision.

# Chapter 3

## Preliminaries

In this chapter we introduce notation and establish basic terminology that will be used in this Monograph. We also present a set of results that will be applied in our research. Some of these are classic well-known facts and some others are genuine results that we proved for our particular settings. This is, for instance, the case of Section 3.7 that contains, at the best of our knowledge, original results.

### 3.1 Set Theory - Basic Notations

For arbitrary sets  $A$  and  $B$ ,  $2^A$  denotes the powerset of  $A$ ,  $A \cup B$ ,  $A \cap B$ ,  $A \uplus B$  and  $A \times B$  denote their union, intersection, disjoint union and cartesian product respectively.

Both  $[A \rightarrow B]$  and  $B^A$  are used to denote the class of functions from  $A$  to  $B$ .

For an arbitrary function  $f : A \rightarrow B$ ,  $f^{-1} : 2^B \rightarrow 2^A$  denotes the pre-image of  $f$ .

Given a set  $A$  and an equivalence relation  $R \subseteq A \times A$ , the set of  $R$ -equivalence classes will be denoted by  $A/R$  and for arbitrary  $x \in A$ ,  $[x]_R$  denotes the equivalence class of  $x$ .

Given a set  $A$  and a relation  $\mathfrak{R} \subseteq A \times A$ , a set  $B \subseteq A$  is said to be  $\mathfrak{R}$ -closed if and only if

$$\{x \in A \mid \exists y \in B, (x, y) \in \mathfrak{R}\} \subseteq B.$$

In what follows we use  $\mathbb{N}$ ,  $\mathbb{Q}$  and  $\mathbb{R}$  to denote the set of natural, rational and real numbers respectively.  $\mathbb{Q}^+$  and  $\mathbb{R}^+$  are used to denote the sets of positive rationals and positive reals (including 0).

### 3.2 Basic concepts of Measure Theory

In this section we introduce a few concepts and results from measure theory. For more details, we refer the reader to [6, 21].

Let  $M$  be an arbitrary nonempty set.

A nonempty family of subsets  $\Pi \subseteq 2^M$  closed under finite intersection is called a  $\pi$ -system.

A nonempty family of subsets  $\Lambda \subseteq 2^M$  is a  $\lambda$ -system if it contains  $M$  and is closed under complementation and countable union of pairwise disjoint sets.

A *field* (of sets) over  $M$  is a family  $\mathcal{F} \subseteq 2^M$  that contains  $M$  and is closed under complement and union.

A *semiring* of subsets of  $M$  is a nonempty family  $\mathcal{S} \subseteq 2^M$  that contains  $\emptyset$ , is closed under finite intersection, and it is such that whenever  $A, B \in \mathcal{S}$  and  $A \subseteq B$ , the set difference  $B \setminus A$  is a finite union of elements of  $\mathcal{S}$ . Clearly any field is a semiring.

A  $\sigma$ -algebra (also called a  $\sigma$ -field) over  $M$  is a field of sets  $\Sigma$  over  $M$  closed under countable union. The tuple  $(M, \Sigma)$  where  $\Sigma$  is a  $\sigma$ -algebra over  $M$ , is called a *measurable space* and the elements of  $\Sigma$  *measurable sets*.

If  $\Omega \subseteq 2^M$ , the  $\sigma$ -algebra generated by  $\Omega$ , denoted in literature by  $\sigma(\Omega)$  or  $\Omega^\sigma$ , is the smallest  $\sigma$ -algebra containing  $\Omega$ .

Every topological space has a natural  $\sigma$ -algebra associated with it, namely the one generated by the open sets. This is called the *Borel algebra* of the space, and the measurable sets are called *Borel sets*.

Given two measurable spaces  $(M, \Sigma)$  and  $(N, \Theta)$ , a function  $f : M \rightarrow N$  is *measurable* if  $f^{-1}(T) \in \Sigma$  for all  $T \in \Theta$ .

A nonnegative real-valued set function  $\mu$  is *finitely additive* if

$$\mu(A \cup B) = \mu(A) + \mu(B) \text{ whenever } A \cap B = \emptyset.$$

We say that  $\mu$  is *countably subadditive* if

$$\mu\left(\bigcup_i A_i\right) \leq \sum_i \mu(A_i)$$

for a countable family of measurable sets, and we say that  $\mu$  is *countably additive* if

$$\mu\left(\bigcup_i A_i\right) = \sum_i \mu(A_i)$$

for a countable *pairwise-disjoint* family of measurable sets. Finite additivity implies monotonicity and countable additivity implies certain continuity properties.

**Theorem 3.2.1.** *Let  $\mathcal{F} \subseteq 2^M$  be a field of sets. If  $\mu : \mathcal{F} \rightarrow \mathbb{R}^+$  is finitely additive, then*

- (i)  $\mu(\emptyset) = 0$ ;
- (ii)  $\mu$  is monotone: if  $A \subseteq B$  then  $\mu(A) \leq \mu(B)$ ; and
- (iii) if  $A \subseteq B$  then  $\mu(B \setminus A) = \mu(B) - \mu(A)$ .

The next theorem provides five equivalent properties that a finitely additive function over a field of sets can have.

**Theorem 3.2.2.** *Let  $\mathcal{F} \subseteq 2^M$  be a field of sets. Let  $\mu : \mathcal{F} \rightarrow \mathbb{R}^+$  be finitely additive. The following are equivalent:*

(i)  $\mu$  is countably subadditive: for any countable collection  $A_i$  such that  $\bigcup_i A_i \in \mathcal{F}$ ,

$$\mu\left(\bigcup_i A_i\right) \leq \sum_i \mu(A_i);$$

(ii)  $\mu$  is countably additive: for any countable collection  $A_i$  such that  $\bigcup_i A_i \in \mathcal{F}$  and the  $A_i$  are pairwise disjoint,

$$\mu\left(\bigcup_i A_i\right) = \sum_i \mu(A_i);$$

(iii)  $\mu$  is  $\omega$ -continuous from below: for any countable chain  $A_0 \subseteq A_1 \subseteq \dots$  such that  $\bigcup_i A_i \in \mathcal{F}$ ,

$$\mu\left(\bigcup_i A_i\right) = \sup_i \mu(A_i);$$

(iv)  $\mu$  is  $\omega$ -continuous from above: for any countable chain  $A_0 \supseteq A_1 \supseteq \dots$  such that  $\bigcap_i A_i \in \mathcal{F}$ ,

$$\mu\left(\bigcap_i A_i\right) = \inf_i \mu(A_i);$$

(v)  $\mu$  is  $\omega$ -continuous from above at  $\emptyset$ : for any countable chain  $A_0 \supseteq A_1 \supseteq \dots$  such that  $\bigcap_i A_i = \emptyset$ ,

$$\mu\left(\bigcap_i A_i\right) = 0.$$

Given a measurable space  $(M, \Sigma)$ , a countably additive set function  $\mu : \Sigma \rightarrow \mathbb{R}^+$  such that  $\mu(\emptyset) = 0$  is a *measure* on  $(M, \Sigma)$ . A measure  $\mu : \Sigma \rightarrow [0, 1]$  is a *subprobability measure*. Thus, for a subprobability measure  $\mu(M) \leq 1$ ; if in addition  $\mu(M) = 1$ ,  $\mu$  is a *probability measure*. Observe that all measures satisfy the five equivalent properties stated in Theorem 3.2.2.

The *null measure*  $\omega$  on  $(M, \Sigma)$  is the unique measure on  $(M, \Sigma)$  such that  $\omega(M) = 0$ . From monotonicity, we get that  $\omega(A) = 0$  for any  $A \in \Sigma$ .

For  $m \in M$ , the *Dirac measure at  $m$*   $D_m$  is defined, for arbitrary  $S \in \Sigma$ , by  $D_m(S) = 1$  if  $m \in S$  and  $D_m(S) = 0$  otherwise. Similarly, one can define the  *$r$ -Dirac measure at  $m$*   $D_{m,r}$  for arbitrary  $r \in \mathbb{R}^+$ , by  $D_{m,r}(S) = r$  if  $m \in S$  and  $D_{m,r}(S) = 0$  otherwise.

We denote by  $\Delta(M, \Sigma)$  the set of measures on  $(M, \Sigma)$ .

The next theorem is repeatedly used for various constructions in what follows.

**Theorem 3.2.3.** *Let  $\mathcal{F} \subseteq 2^M$  be a field of sets. Let  $\mu : \mathcal{F} \rightarrow \mathbb{R}^+$  be finitely additive and countably subadditive. Then  $\mu$  extends uniquely to a measure on  $\sigma(\mathcal{F})$ .*

*Proof.* This argument is known as the *Carathéodory construction*. For  $B \in 2^M$ , define

$$\mu^*(B) = \inf_{\substack{B \subseteq \bigcup \mathcal{C} \\ \mathcal{C} \text{ countable}}} \sup_{\substack{F \subseteq \mathcal{C} \\ F \text{ finite}}} \mu(\bigcup F),$$

where the infimum ranges over countable  $\mathcal{F}$ -covers  $\mathcal{C}$  of  $B$ . The map  $\mu^*$  is called an *outer measure* of  $\mu$ . One can show that  $\mu^*$  satisfies the following properties:

- $\mu^*$  and  $\mu$  agree on  $\mathcal{F}$ ;
- $\mu$  is monotone;
- $\mu$  is countably subadditive.

Now define a set  $B$  to be *measurable* with respect to  $\mu^*$  if for all  $A \in 2^M$ ,

$$\mu^*(A) = \mu^*(A \cap B) + \mu^*(A \setminus B).$$

One now shows that the set of measurable sets with respect to  $\mu^*$  is a  $\sigma$ -algebra, therefore contains  $\sigma(\mathcal{F})$ , and all properties of measures are satisfied.  $\square$

Given a measurable function  $f: (M, \Sigma) \rightarrow (N, \Theta)$ , any measure  $\mu$  on  $(M, \Sigma)$  defines a measure  $\mu[f]$  on  $(N, \Theta)$  by  $\mu[f](E) = \mu(f^{-1}(E))$ , for all  $E \in \Theta$ .  $\mu[f]$  is called the *push forward of  $\mu$  under  $f$* .

Given two measurable spaces  $(M, \Sigma)$  and  $(N, \Theta)$ , the *product space*,  $(M, \Sigma) \otimes (N, \Theta)$ , is the measurable space  $(M \times N, \Sigma \otimes \Theta)$ , where  $\Sigma \otimes \Theta$  is the  $\sigma$ -algebra generated by the *rectangles*  $E \times F$  for  $E \in \Sigma$  and  $F \in \Theta$ .

Given  $\mu$  and  $\nu$  measures on  $(M, \Sigma)$  and  $(N, \Theta)$ , respectively, the *product measure*  $\mu \times \nu$  on  $(M, \Sigma) \otimes (N, \Theta)$  is *uniquely* defined by  $(\mu \times \nu)(E \times F) = \mu(E) \cdot \nu(F)$ , for all  $(E, F) \in \Sigma \times \Theta$ .

A measure  $\omega$  on  $(M, \Sigma) \otimes (N, \Theta)$  is a *coupling* for  $(\mu, \nu)$  if for all  $E \in \Sigma$  and  $F \in \Theta$ ,  $\omega(E \times Y) = \mu(E)$  and  $\omega(X \times F) = \nu(F)$ . In such a case,  $\mu$  is called the *left marginal* and  $\nu$  the *right marginal* of  $\omega$ .

### 3.3 Basic concepts of Topology

Given a set  $X$ , a collection  $\tau$  of subsets of  $X$  is a *topology* on  $X$  if it satisfies the following conditions.

- $\tau$  contains the empty set and  $X$ ;
- arbitrary unions of sets in  $\tau$  belong to  $\tau$ ;
- finite intersections of elements of  $\tau$  belong to  $\tau$ .

The elements of  $\tau$  are called open sets and the tuple  $(X, \tau)$  is called a *topological space*.

The complements of the sets in  $\tau$  are called *closed sets*.

A subset of  $X$  may be neither closed nor open, either closed or open, or both. A set that is both closed and open is called a *clopen set*.

A set is a  $G_\delta$  set if it is a countable intersection of open sets; it is a  $F_\sigma$  set if it is a countable union of closed sets.

A *base* (or *basis*)  $B$  for a topological space  $(X, \tau)$  is a collection of open sets in  $\tau$  such that every open set can be written as a union of elements of  $B$ . We say that the base generates the topology  $\tau$ .

Given a topological space  $(X, \tau)$  and a set  $A \subseteq X$ , the *closure* of  $A$  is the smallest closed set that contains  $A$  and the *interior* of  $A$  is the largest open set included in  $A$ . The *boundary* of  $A$  is the set difference between its closure and its interior.

Given a topological space  $(X, \tau)$  and a subset  $A \subseteq X$ , the *relative topology* (*subspace topology*) on  $A$  is the topology  $\tau' = \{S \cap A \mid S \in \tau\}$ .

The *inseparability* (*indistinguishability*) relation induced by a topology over its support set is the equivalence relation that pairs the points that belong to exactly the same open sets.

Given two topological spaces  $(X, \tau)$  and  $(Y, \theta)$ , a function  $f : X \rightarrow Y$  is *continuous* if the inverse image of an open set in  $Y$  is open in  $X$ , i.e., for any  $S \in \theta$ ,  $f^{-1}(S) \in \tau$ . Equivalently,  $f$  is continuous if the inverse image of a closed set in  $Y$  is closed in  $X$ .

Every topological space has a natural  $\sigma$ -algebra associated with it, namely the one generated by the open sets. This is called the *Borel algebra* of the space, and the measurable sets are called *Borel sets*.

A topological space  $(X, \tau)$  is *compact* if any cover of the space with open sets contains a finite subcover. Equivalently, the space is compact if any family of closed sets with empty intersection contains a finite subfamily with empty intersection.

A subset  $A \subseteq X$  of a topological space  $(X, \tau)$  is *compact* if any cover of  $A$  with open sets contains a finite subcover. Equivalently,  $A$  is compact if its relative topology form a compact topological space on  $A$ .

A topological space  $(X, \tau)$  is *Kolmogorov* ( $T_0$ ) if for every pair of distinct points  $x, y \in X$ , there exists an open set  $A \in \tau$  containing exactly one of the two points.

A topological space  $(X, \tau)$  is *Frechét* ( $T_1$ ) if for every pair of distinct points  $x, y \in X$ , there exists an open set  $A_x \in \tau$  containing  $x$  but not  $y$ .

A topological space  $(X, \tau)$  is *Hausdorff* ( $T_2$ ) if for every pair of distinct points  $x, y \in X$ , there exist disjoint open sets  $A_x, A_y \in \tau$  such that  $x \in A_x$  and  $y \in A_y$ .

A topological space  $(X, \tau)$  is *totally disconnected* if for every pair of distinct points  $x, y \in X$ , there exist two disjoint open sets  $A_x, A_y \in \tau$  such that  $x \in A_x$ ,  $y \in A_y$  and  $A_x \cup A_y = X$ .

A subset of a topological space is a *dense set* if its closure is the entire space. For example, the set  $\mathbb{Q}$  of rational numbers is dense in the topological space  $\mathbb{R}$  of real number with the topology generated, for instance, by the open intervals – which is the usual topology on reals.

A topological space is *zero-dimensional* (with respect to the small inductive dimension) if it has a base consisting of clopen sets.

A topological space  $(X, \tau)$  is *separable* if it contains a countable dense subset. It is *second countable* if its topology has a countable base. Second countability implies separability, but separability does not imply second countability. However, the two concepts coincide for metric spaces.

### 3.4 Metric and Pseudometric Spaces

Given a set  $X$ , a *pseudometric on  $X$*  is a function  $d: X \times X \rightarrow \mathbb{R}^+$  such that for arbitrary  $x, y, z \in X$  the following conditions are satisfied

- (1):  $d(x, x) = 0$
- (2):  $d(x, y) = d(y, x)$
- (3):  $d(x, y) \leq d(x, z) + d(z, y)$

The function  $d$  is a *hemimetric on  $X$*  if it is not symmetric, i.e., it does not satisfy the condition (2);  $d$  is a *metric on  $X$*  if, it satisfies (1)-(3) and, in addition, it also satisfy the condition (4) below.

- (4):  $d(x, y) = 0$  implies  $x = y$

If  $d$  is a pseudometric on  $X$ , the pair  $(X, d)$  is a *pseudometric space*; if  $d$  is a metric on  $X$ , the pair  $(X, d)$  is a *metric space*.

In a pseudometric or metric space  $(X, d)$ , the *open ball* with center  $x \in X$  and radius  $\varepsilon > 0$  is the set  $\{y \in X \mid d(x, y) < \varepsilon\}$ . The collection of open balls forms a base for a topology called the *open-ball topology induced by  $d$* .

Given a (pseudo-) metric space  $(X, d)$ , a sequence  $(x_i)_{i \in \mathbb{N}}$  converges to  $x \in X$  if for any  $\varepsilon > 0$  there exists  $k \in \mathbb{N}$  such that for any  $j \geq k$ ,  $d(x_j, x) < \varepsilon$ . In such a case we say that  $x$  is the limit of the sequence  $(x_i)_{i \in \mathbb{N}}$ .

In pseudometric spaces the limit of a sequence is not necessarily unique: if  $x$  is a limit and  $y \in X$  is such that  $d(x, y) = 0$ , then  $y$  is also a limit of the sequence. In the case of metric spaces and Hausdorff spaces in general (any metric space with the open-ball topology is Hausdorff) one can prove that if a sequence has a limit, then the limit is unique.

A sequence  $(x_i)_{i \in \mathbb{N}}$  is a *Cauchy sequence* if for any  $\varepsilon > 0$  there exists  $k \in \mathbb{N}$  such that for any  $i, j \geq k$ ,  $d(x_i, x_j) < \varepsilon$ .

A metric space is *complete* when every Cauchy sequence converges.

A pseudometric  $d$  on  $X$  defines an equivalence relation on  $X$ , called the *inseparability relation*, by pairing the elements at distance zero, i.e., the inseparability relation is the kernel

$$\ker(d) = \{(x, y) \in X^2 \mid d(x, y) = 0\}.$$

Observe that the inseparability relation induced by a pseudometric coincides with the inseparability relation of the open-ball topology induced by the pseudometric.

Given a pseudometric space  $(X, d)$ , we can lift the pseudometric structure from the points in  $X$  to sets of points in  $2^X$  in a manner analogous to the way in which one extends metrics to compact sets. We define the *Hausdorff pseudometric*  $d^H$  on the class of subsets of  $X$  by

$$d^H(X, Y) = \max(\sup_{x \in X} \inf_{y \in Y} d(x, y), \sup_{y \in Y} \inf_{x \in X} d(x, y)).$$

**Lemma 3.4.1.** *If  $d : M \times M \rightarrow [0, 1]$  is a pseudometric on  $M$ , then  $d^H$  is a pseudometric on subsets of  $M$ .*

*Proof.* Obviously,  $d^H(X, X) = 0$  for any  $X \subseteq M$ , since  $\inf_{x' \in X} d(x', x) = 0$ .

We prove now the triangle inequality for arbitrary subsets  $X, Y, Z \subseteq M$ . Note that since  $d$  takes values in  $[0, 1]$  the inf and sup are well defined.

We have that  $d(x, y) \leq d(x, z) + d(z, y)$ . From it we derive the following inequalities.

$$\begin{aligned} d(x, y) &\leq d(x, z) + \sup_{z' \in Z} d(z', y) \\ \inf_{y' \in Y} d(x, y') &\leq d(x, z) + \sup_{z' \in Z} d(z', y) \\ \inf_{y' \in Y} d(x, y') &\leq d(x, z) + \sup_{z' \in Z} \inf_{y'' \in Y} d(z', y''). \end{aligned}$$

We define  $d^*(X, Y) = \sup_{x \in X} \inf_{y \in Y} d(x, y)$  and  $d_*(X, Y) = \sup_{y \in Y} \inf_{x \in X} d(x, y)$ . The last inequality is equivalent to:

$$\inf_{y' \in Y} d(x, y') \leq d(x, z) + d^*(Z, Y)$$

And from it we derive the following chain of inequalities.

$$\begin{aligned} \inf_{y' \in Y} d(x, y') &\leq \sup_{x' \in X} d(x', z) + d^*(Z, Y) \\ \inf_{y' \in Y} d(x, y') &\leq \inf_{z' \in Z} \sup_{x' \in X} d(x', z') + d^*(Z, Y) \\ \sup_{x \in X} \inf_{y' \in Y} d(x, y') &\leq \inf_{z' \in Z} \sup_{x' \in X} d(x', z') + d^*H(Z, Y) \end{aligned}$$

This last inequality is equivalent to

$$d^*(X, Y) \leq d^*(X, Z) + d^*(Z, Y).$$

A similar proof can be carried out for  $d_*$ . The triangle inequality for  $d^H$  follows immediately.  $\square$

### 3.5 Analytic Spaces

Measure theory works most smoothly in conjunction with certain topological assumptions; in fact a standard restriction is to consider so called *Polish spaces*.

A *Polish space* is the topological space underlying a complete separable metric space.

An *analytic space*<sup>1</sup> is a continuous image of a Polish space in a Polish space. More precisely, if  $X$  and  $Y$  are Polish spaces and  $f : X \rightarrow Y$  is continuous, then the image  $f(X)$  is an analytic space. Remarkably, one does not get a broader class by allowing  $f$  to be merely measurable instead of continuous, nor by taking the image of a Borel subset of  $X$  instead of  $X$ . That is, the measurable image of any Borel subset of a Polish space in a Polish space is analytic.

Analytic spaces enjoy remarkable properties that were crucial in proving the logical characterization of probabilistic bisimulation as well as in establishing the completeness theorems for probabilistic modal logics – results that are central in what follows.

### 3.6 Rasiowa-Sikorski Lemma

The *Baire category theorem* is a topological result with important applications in logic. It is used to prove the Rasiowa-Sikorski lemma [25, 39] which is a crucial result for this monograph.

The Rasiowa-Sikorski lemma is a model-theoretic result that exploits the Baire category theorem and the Stone duality for boolean algebras with operators. Applied to logics, the Rasiowa-Sikorski lemma states that given a multimodal logic (possibly involving an infinite set of modalities) for which the provability relation admits an axiomatization such that the set of instances of the infinitary proof rules (if any) is countable, then for any consistent formula  $\varphi$ , there exists a maximally-consistent set of formulas containing  $\varphi$ .

Recall that a subset  $D$  of a topological space  $X$  is *dense* if its closure  $\overline{D}$  is all of  $X$ . Equivalently, a dense set is one intersecting every nonempty open set.

A set  $N \subseteq X$  is *nowhere dense* if every nonempty open set contains a nonempty open subset disjoint from  $N$ .

A set is *of the first category* or *meager* if it is a countable union of nowhere dense sets. The term “meager” is meant to suggest that these sets are small in a topological sense. A basic fact that we use is that the boundary of an open set is nowhere dense.

A *Baire space* is a topological space in which the intersection of countably many dense open sets is dense. It follows from these definitions that the complement of a first category set is dense in any Baire space. Baire originally proved that the real line is a Baire space. More generally, every Polish space is Baire and every locally compact Hausdorff

---

<sup>1</sup>This concept is called in some mathematical communities an *analytic set* since they use the concept of analytic space for another concept.

space is Baire. For us in this Monograph, the relevant version is the following special case: *every compact Hausdorff space is Baire*.

A *Boolean algebra* (BA) over the set  $B \neq \emptyset$  can be given abstractly as a structure  $\mathcal{B} = (B, \top, \perp, \neg, \vee, \wedge, \leq)$  that satisfies the equational axioms in table 3.1, where  $\top, \perp \in B$ ,  $\neg : B \rightarrow B$  is a monadic operation on  $B$ ,  $\vee, \wedge : B \times B \rightarrow B$  are dyadic operations on  $B$  and  $\leq \subseteq B \times B$  is a relation on  $B$ .

(BA1):	$a \vee (b \vee c) = (a \vee b) \vee c$
(BA2):	$a \wedge (b \wedge c) = (a \wedge b) \wedge c$
(BA3):	$a \vee b = b \vee a$
(BA4):	$a \wedge b = b \wedge a$
(BA5):	$a \vee (b \wedge a) = a \wedge (b \vee a) = a$
(BA6):	$a \vee (b \wedge c) = (a \vee b) \wedge (a \vee c)$
(BA7):	$a \wedge (b \vee c) = (a \wedge b) \vee (a \wedge c)$
(BA8):	$a \vee (\neg a) = \top$
(BA8):	$a \wedge (\neg a) = \perp$
(BA9):	$a \leq b$ iff $a = a \wedge b$ iff $b = a \vee b$

Table 3.1: Boolean Algebra

Let  $\mathcal{B}$  be a boolean algebra and let  $T \subseteq \mathcal{B}$  be such that  $T$  has a greatest lower bound  $\bigwedge T$  in  $\mathcal{B}$ . An ultrafilter (maximal filter)  $U$  is said to *respect*  $T$  if  $T \subseteq U$  implies that  $\bigwedge T \in U$ . If  $\mathcal{T}$  is a family of subsets of  $\mathcal{B}$ , we say that an ultrafilter  $U$  *respects*  $\mathcal{T}$  if it respects every member of  $\mathcal{T}$ .

**Lemma 3.6.1** (Rasiowa–Sikorski lemma [39]). *For any boolean algebra  $\mathcal{B}$  and any countable family  $\mathcal{T}$  of subsets of  $\mathcal{B}$ , each member of which has a meet in  $\mathcal{B}$ , and for any nonzero  $x \in \mathcal{B}$ , there exists an ultrafilter in  $\mathcal{B}$  that contains  $x$  and respects  $\mathcal{T}$ .*

This lemma was later proved by Tarski in a purely algebraic way. See [25] for a discussion of the role of the Baire category theorem in the proof of this lemma.

## 3.7 Lindenbaum's Lemma and Non-Compact Logics

In this section we instantiate the Rasiowa-Sikorski lemma to get a series of model theoretic results regarding Boolean logics<sup>2</sup>. We eventually prove an extension of the Rasiowa-Sikorski lemma and use it to demonstrate that countably axiomatized logics, even if they are non-compact or have infinitary axiomatizations, enjoy the so called Lindenbaum's

---

<sup>2</sup>In this context, by a Boolean logic we mean a logic that obeys the Boolean laws; it can be modal or first order, etc.

lemma that states that any consistent set of formulas can be extended to a maximally-consistent set.

Given a Boolean logic  $\mathcal{L}$ , a relation  $\vdash \subseteq 2^{\mathcal{L}} \times \mathcal{L}$  is a *deducibility relation on  $\mathcal{L}$*  if it satisfies the following conditions.

- (R1) If  $F \vdash \varphi$  and  $F \subseteq F'$ , then  $F' \vdash \varphi$ ;
- (R2) If  $\varphi \in F$ , then  $F \vdash \varphi$ ;
- (R3) If  $F \vdash \varphi$  and  $F \cup \{\varphi\} \vdash \perp$ , then  $F \vdash \perp$ ;
- (R4)  $F \cup \{\neg\varphi\} \vdash \perp$  iff  $F \vdash \varphi$ .

We call the elements of  $\vdash$  *inferences*. The previous rules guarantee that a deducibility relation is, in particular, a Boolean deducibility, i.e.,

$$\varphi \rightarrow \psi \text{ is a Boolean validity iff } \{\varphi\} \vdash \psi.$$

Given a deducibility relation  $\vdash$ , a set  $\Phi \subseteq \mathcal{L}$  of formulas is  $\vdash$ -*consistent* if it is not the case that  $\Phi \vdash \perp$ ;  $\Phi$  is  $\vdash$ -*inconsistent* if  $\Phi \vdash \perp$ . Similar concepts are defined for a formula  $\varphi \in \mathcal{L}$ : it is *consistent/inconsistent* if  $\{\varphi\}$  is *consistent/inconsistent*.

We say that a deducibility relation on  $\mathcal{L}$  is *countably-axiomatized* if there exists a countable set  $(\Gamma_i \vdash \varphi_i)_{i \in \mathbb{N}}$  of inferences such that from them using the rules (R1)-(R4) we can derive all the elements of  $\vdash$ .

The next theorem is a direct consequence of the Rasiowa-Sikorski Lemma.

**Theorem 3.7.1.** *Given a Boolean logic  $\mathcal{L}$  and a countably-axiomatized deducibility relation  $\vdash$ , for any  $\vdash$ -consistent formula  $\varphi \in \mathcal{L}$  there exists a  $\vdash$ -maximally consistent set of formulas that contains  $\varphi$ .*

*Proof.* Observe that the Lindenbaum algebra  $\mathcal{L}^{\leftrightarrow}$  (the quotient of  $\mathcal{L}$  w.r.t. logical equivalence) is a Boolean algebra. Each of the inferences in the class  $(\Gamma_i \vdash \varphi_i)_{i \in \mathbb{N}}$  state that the meet of the elements in  $\Gamma_i$  exists in  $\mathcal{L}^{\leftrightarrow}$  and it is equal to  $\varphi_i$ , for each  $i = 1, n$ . Consequently, a  $\vdash$ -maximally consistent set of formulas is an ultrafilter that respects each element of the class  $(\Gamma_i)_{i \in \mathbb{N}}$ . Since this class is countable, applying Rasiowa-Sikorski lemma we get that there must exist an ultrafilter of  $\mathcal{L}^{\leftrightarrow}$  that contains  $\varphi$  and respects each element in  $(\Gamma_i)_{i \in \mathbb{N}}$  - this is a  $\vdash$ -maximally consistent set of formulas.  $\square$

Observe that the previous theorem states that each consistent formula is contained in a maximally-consistent set. This does not mean that each consistent set of formulas has a maximally-consistent extension (Lindenbaum's lemma). While in the case of compact logics this can be proved using Zorn's lemma, in the case of non-compact logics one cannot prove constructively the existence of a maximal-consistent extension.

In what follows we prove that under the assumptions of the previous theorem Lindenbaum's lemma is satisfied also for the case of non-compact logics that admit a countable (possibly infinitary) axiomatization.

In order to prove this, we firstly introduce a few concepts.

A *completion* of a Boolean algebra  $\mathbb{A}$  is a Boolean algebra  $\mathcal{B}$  enjoying the following properties:

1.  $\mathbb{A}$  is a subalgebra of  $\mathcal{B}$ ;
2. every subset of  $\mathbb{A}$  has a supremum in  $\mathcal{B}$ ;
3. every element in  $\mathcal{B}$  is the supremum (in  $\mathcal{B}$ ) of some subset of  $\mathbb{A}$ .

Now, we state a few basic facts about the completion of a Boolean algebra (for more see [23]). First of all, any Boolean algebra  $\mathbb{A}$  has a completion, namely (an isomorphic copy of) the Boolean algebra of complete ideals in  $\mathbb{A}$ . The completion is a complete Boolean algebra. Every element in the completion is the supremum (in the completion) of the set of elements in  $\mathbb{A}$  that are below it. Condition (3) in the previous definition is equivalent to the assertion that  $\mathbb{A}$  is a *dense* subset of  $\mathcal{B}$  in the sense that every non-zero element in  $\mathcal{B}$  is above a non-zero element in  $\mathbb{A}$ .

These results will support us in proving the following lemma.

**Lemma 3.7.2** (Extended Rasiowa-Sikorski Lemma). *Let  $\mathcal{B}$  be a Boolean algebra,  $\overline{\mathcal{B}}$  its completion and  $\mathcal{T}$  a countable family of subsets of  $\mathcal{B}$  each member of which has a meet in  $\mathcal{B}$ . If  $S \subseteq \mathcal{B}$  is such that  $\bigwedge S \neq 0$  in  $\overline{\mathcal{B}}$ , then there exists an ultrafilter  $\mathcal{U}$  of  $\mathcal{B}$  such that  $S \subseteq \mathcal{U}$  and  $\mathcal{U}$  respects  $\mathcal{T}$ .*

*Proof.* Since  $\bigwedge S \neq 0$  in  $\overline{\mathcal{B}}$ , there exists an ultrafilter  $\overline{\mathcal{U}}$  of  $\overline{\mathcal{B}}$  such that  $\bigwedge S \in \overline{\mathcal{U}}$  and  $\overline{\mathcal{U}}$  respects  $\overline{\mathcal{T}}$ .

Let  $\mathcal{U} = \overline{\mathcal{U}} \cap \mathcal{B}$ . Because  $\mathcal{B}$  is a subalgebra of  $\overline{\mathcal{B}}$ ,  $\mathcal{U}$  is an ultrafilter of  $\mathcal{B}$ .

Since  $\bigwedge S \in \overline{\mathcal{U}}$  and for any  $s \in S$ ,  $\bigwedge S \leq s$ , we obtain that  $S \subseteq \overline{\mathcal{U}}$ , because  $\overline{\mathcal{U}}$  is a filter. But we also know that  $S \subseteq \mathcal{B}$ . Hence,  $S \subseteq \mathcal{U}$ .

Consider an arbitrary set  $T \in \mathcal{T}$  and assume that  $T \subseteq \mathcal{U}$ . Then,  $T \subseteq \overline{\mathcal{U}}$ . Because  $\overline{\mathcal{U}}$  respects  $T$ , we get that  $\bigwedge T \in \overline{\mathcal{U}}$ . But  $\bigwedge T \in \mathcal{B}$  from hypothesis. Hence,  $\bigwedge T \in \mathcal{U}$ . Consequently,  $\mathcal{U}$  respects each elements of  $\mathcal{T}$  and contains  $S$ .  $\square$

A direct consequence of the previous lemma is the Lindenbaum's lemma for a countably-axiomatized deductibility relation.

**Theorem 3.7.3** (Lindenbaum's Lemma for Countably-Axiomatized Logics). *Given a Boolean logic  $\mathcal{L}$  and a countably-axiomatized deductibility relation  $\vdash$ , then for any  $\vdash$ -consistent set for formula  $\Phi \subseteq \mathcal{L}$  there exists a  $\vdash$ -maximally consistent set of formulas that extends  $\Phi$ .*

*Proof.* As in the case of Theorem 3.7.1, we consider the Lindenbaum algebra  $\mathcal{L}^{\leftrightarrow}$ , which is a Boolean algebra. Each of the countably many inferences  $(\Gamma_i \vdash \varphi_i)_{i \in \mathbb{N}}$  that axiomatize  $\vdash$  state that the meet of the elements in  $\Gamma_i$  exists in  $\mathcal{L}^{\leftrightarrow}$  and it is equal to  $\varphi_i$ , for each  $i = 1, n$ . Consequently, a  $\vdash$ -maximally consistent set of formulas is an ultrafilter

that respects each element of the class  $(\Gamma_i)_{i \in \mathbb{N}}$ . Since this class is countable and  $\Phi$  is consistent, meaning that in the completion  $\mathcal{L}^{\leftrightarrow}$  of  $\mathcal{L}^{\leftrightarrow}$  it has a meet which is different than  $\perp$ , we can apply the Extended Rasiowa-Sikorski lemma 3.7.2 to get that there must exist an ultrafilter of  $\mathcal{L}^{\leftrightarrow}$  that contains  $\varphi$  and respects each element in  $(\Gamma_i)_{i \in \mathbb{N}}$  - this is a  $\vdash$ -maximally consistent set of formulas. □

### 3.8 Markov Processes

To gain the maximum level of generality we will consider, in the same time, three types of Markov processes: the *probabilistic*, the *sub-probabilistic* and the so-called *stochastic* Markov process. The first two classes are continuous-state space models of what we usually call discrete-time processes, while the stochastic case generalizes, for the continuous-state space the continuous-time processes. Each of these types of systems is defined for a measurable space of states and the transitions are from states to measurable sets of states. In the sub-probabilistic and probabilistic cases, these measures represent the probability that a transition from the initial state can be performed to a random state in the target measurable set, hence this value is in the interval  $[0, 1]$ . In the stochastic case the transition from a state to a measurable set is described by a positive real that represents the rate of an exponentially-distributed random variable that characterizes the time when the transition is fired.

Formally, we assume an analytic space  $(M, \Sigma)$ , where  $\Sigma$  is the Borel algebra induced by the underlying topology. And let  $\Delta(M, \Sigma)$ ,  $\Pi(M, \Sigma)$  and  $\Pi^*(M, \Sigma)$  be the classes of general, probabilistic and subprobabilistic measures on  $(M, \Sigma)$  respectively.

The space of measures on  $(M, \Sigma)$  is organized as a measurable space as follows. We define a  $\sigma$ -algebra on  $\Delta(M, \Sigma)$ , which is the one generated by the sets

$$\{\mu \in \Delta(M, \Sigma) \mid \mu(S) \geq r\} \text{ for } S \in \Sigma \text{ and } r \in \mathbb{Q}^+.$$

This is the least  $\sigma$ -algebra on  $\Delta(M, \Sigma)$  such that all maps  $\mu \mapsto \mu(S) : \Delta(M, \Sigma) \rightarrow \mathbb{R}^+$  for  $S \in \Sigma$  are measurable, where the set of positive reals is endowed with the  $\sigma$ -algebra generated by all rational intervals, i.e. the Borel  $\sigma$ -algebra. This is the stochastic case.

Similarly, the set  $\Pi^*(M, \Sigma)$  of subprobabilistic measures on  $(M, \Sigma)$  and the set  $\Pi(M, \Sigma)$  of probabilistic measures on  $(M, \Sigma)$  can be organized as measurable spaces by defining the  $\sigma$ -algebras generated by the sets

$$\{\mu \in \Pi^*(M, \Sigma) \mid \mu(S) \geq r\}$$

and

$$\{\mu \in \Pi \mid \mu(S) \geq r\}$$

respectively, defined for  $S \in \Sigma$  and  $r \in [0, 1] \cap \mathbb{Q}$ .

A *Markov process* (MP) is a tuple  $\mathcal{M} = (M, \Sigma, \theta)$ , where  $\theta : (M, \Sigma) \rightarrow \Omega(M, \Sigma)$  is a measurable function defined, in various cases in what follows, for  $\Omega \in \{\Delta, \Pi^*, \Pi\}$ .

Thus, we will have stochastic, probabilistic and sub-probabilistic MPs.

In a Markov process  $\mathcal{M} = (M, \Sigma, \theta)$ ,  $M$  is the *support set*, denoted by  $\text{supp}(\mathcal{M})$ , and  $\theta$  is the *transition function*.

For the probabilistic (sub-probabilistic) case,  $\theta(m) : \Sigma \rightarrow [0, 1]$  for  $m \in M$ , is a probability (sub-probability) measure on the state space  $(M, \Sigma)$ ; and for  $N \in \Sigma$ , the value  $\theta(m)(N) \in [0, 1]$  represents the probability of a transition from  $m$  to a state in  $N$ .

In the stochastic case,  $\theta(m) : \Sigma \rightarrow \mathbb{R}^+$  for  $m \in M$ , is a measure on the state space  $(M, \Sigma)$  and for  $N \in \Sigma$ , the value  $\theta(m)(N) \in \mathbb{R}^+$  represents the rate of an exponentially-distributed random variable that computes the probability that a transition from  $m$  to a state in  $N$  is fired within a given time.

In all these cases, the condition that  $\theta$  is a measurable function from  $M$  to  $\Delta(M, \Sigma)$  is equivalent to the condition that for fixed  $N \in \Sigma$ , the function  $m \mapsto \theta(m)(N)$  is a measurable function from  $M$  to  $\mathbb{R}$ . (see e.g. Proposition 2.9 of [20]).

Given two Markov processes (of the same type)  $\mathcal{M}_i = (M_i, \Sigma_i, \theta_i)$ ,  $i = 1, 2$ , a surjective measurable function  $f : M_1 \rightarrow M_2$  is a *zig-zag* if for any  $m \in M_1$  and  $B \in \Sigma_2$ ,  $\theta_1(m)(f^{-1}(B)) = \theta_2(f(m))(B)$ . Such a map is essentially a functional version of bisimulation [18].

A *span* in a category is a pair of morphisms  $f : A \rightarrow B$  and  $g : A \rightarrow C$  with a common domain. A *cospan* is a pair of morphisms  $f : B \rightarrow A$  and  $g : C \rightarrow A$  with a common codomain.

Two Markov processes  $\mathcal{M}_1, \mathcal{M}_2$  are said to be *bisimilar* if there is a third Markov process  $\mathcal{M}$  and a span of zig-zags  $f_i : \mathcal{M} \rightarrow \mathcal{M}_i$ ,  $i = 1, 2$ .

Two states  $m_i \in \text{supp}(\mathcal{M}_i)$ ,  $i = 1, 2$ , are said to be *bisimilar* if there exist a span of zig-zags  $f_i : \mathcal{M} \rightarrow \mathcal{M}_i$ ,  $i = 1, 2$  and  $m \in \text{supp}(\mathcal{M})$  such that  $m_i = f_i(m)$ ,  $i = 1, 2$ .

We write  $(\mathcal{M}_1, m_1) \approx (\mathcal{M}_2, m_2)$  to indicate that  $m_1$  and  $m_2$  are bisimilar in this sense.

In order to show that this definition actually is transitive one needs to restrict the class of measurable spaces one is working with to analytic spaces. This is a restriction but any space that one encounters in practice is likely to be Polish and almost certainly analytic. This notion of bisimulation is equivalent to the following on analytic spaces.

Two Markov processes  $\mathcal{M}_1, \mathcal{M}_2$  are **bisimilar** if there is a third Markov process  $\mathcal{M}$  and a co-span of zig-zags  $f : \mathcal{M}_1 \rightarrow \mathcal{M}$  and  $g : \mathcal{M}_2 \rightarrow \mathcal{M}$ .

In the context of analytic spaces, one can alternatively define bisimulation between the states of a Markov process using a relational definition [18, 29, 37]. This concept is equivalent to the previous two concepts of bisimilarity defined in categorial settings.

Given a binary relation  $\mathfrak{R} \subseteq M \times M$  on a set  $M$ , we call a subset  $N \subseteq M$   $\mathfrak{R}$ -closed iff

$$\{m \in M \mid \exists n \in N, (n, m) \in \mathfrak{R}\} \subseteq N.$$

If  $(M, \Sigma)$  is a measurable space,  $\Sigma(\mathfrak{R})$  denotes the set of measurable  $\mathfrak{R}$ -closed subsets of  $M$ .

Let  $\mathcal{M} = (M, \Sigma, \theta)$  be an  $A$ -Markov process. A *bisimulation relation* is an equivalence relation  $\mathfrak{R} \subseteq M \times M$  such that  $(m, n) \in \mathfrak{R}$  iff for any  $C \in \Sigma(\mathfrak{R})$  and any  $\alpha \in A$ ,

$$\theta(\alpha)(m)(C) = \theta(\alpha)(n)(C).$$

Two MPs  $(\mathcal{M}, m)$  and  $(\mathcal{M}, n)$  are *bisimilar*, written  $m \sim_{\mathcal{M}} n$ , if  $m$  and  $n$  are related by a rate-bisimulation relation. The *bisimilarity* is the largest rate-bisimulation.

Observe in the previous definition the role of  $\mathfrak{R}$  as it is reflected in the structure of  $C$ . This definition reflects subtle coalgebraic facts that we will not discuss here. For a detailed analysis of the concept of stochastic bisimulation the reader is referred to [15].

This last definition of bisimulation can be further extended to define bisimulation between states of different Markov processes by simply taking the disjoint union of the two and use the relational definition in this larger MP.

# Bibliography

- [1] R. Alur and D. Dill. *Automata for Modeling real-time Systems*. In M. S. Paterson, editor, *Automata, Languages and Programming*, volume 443 of *Lecture Notes in Computer Science*, pages 322–335. Springer Berlin Heidelberg, 1990.
- [2] R. Alur and T. A. Henzinger. *A Really Temporal Logic*. *Journal of the ACM*, 41(1):181–204, 1994.
- [3] R. Alur and T. A. Henzinger. *Real-Time Logics: Complexity and Expressiveness*. *Information and Computation*, 104(1):35–77, 1993.
- [4] G. Bacci, G. Bacci, K.G. Larsen, R. Mardare. *Computing Behavioral Distances, Compositionally*. In *Proc. MFCS2013*:74-85, 2013.
- [5] P. Ballarini, R. Mardare, I. Mura. *Analysing Biochemical Oscillations through Probabilistic Model Checking*. *Electronic Notes in Theoretical Computer Science*, ENTCS 229(1):3-19, 2009.
- [6] P. Billingsley. *Probability and Measure*. Wiley-Interscience, 1995.
- [7] R. Blute et al. *Bisimulation for labeled Markov processes*. In *Proc. LICS'97*, IEEE Press, 1997.
- [8] L. Cardelli, and A. Csikasz-Nagy. *The Cell Cycle Switch Computes Approximate Majority*. *Scientific Reports* 2:656, 13 September 2012. Nature Publishing Group, 2012
- [9] M. Cavaliere, R. Mardare. *Playing with partial knowledge in membrane systems: A logical approach*. In *Proc. of the seventh Workshop on Membrane Computing - At the Crossroads of Cell Biology and Computation*, WMC2006, *Lecture Notes in Computer Science*, LNCS 4361:279-297, Springer, 2006.
- [10] M. Cavaliere, R. Mardare, S. Sedwards. *Colonies of synchronizing agents: An abstract model of intracellular and intercellular processes*. In *Proc. of International Workshop Automata for Cellular and Molecular Computing*, page:35-51, Computer and Automation Research Institute of the Hungarian Academy of Sciences, 2007.
- [11] K. Chatterjee and L. de Alfaro and R. Majumdar and V. Raman. *Algorithms for Game Metrics*. *Logical Methods in Computer Science*, 6.3, 2010.

- [12] T. Chen, M. Diciolla, M. Kwiatkowska, and A. Mereacre. *A simulink hybrid heart model for quantitative verification of cardiac pacemakers*. In Proc. of HSCC 2013: 131–136.
- [13] T. Chen and S. Kiefer. *On the Total Variation Distance of Labelled Markov Chains*. In Proc. of CSL-LICS '14, CSL-LICS '14, pages 33:1–33:10, New York, NY, USA, 2014. ACM.
- [14] T. Chen, F. van Breugel, and J. Worrell. *On the Complexity of Computing Probabilistic Bisimilarity*. In *FoSSaCS*, volume 7213 of Lecture Notes in Computer Science, pages 437–451. Springer, 2012.
- [15] V. Danos et al. *Bisimulation and Cocongruence for Probabilistic Systems*, Inf. and Comp. 204(4):503–523, 2006.
- [16] E.P. de Vink, J. Rutten, *Bisimulation for probabilistic transition systems: A coalgebraic approach*. TCS 221(1-2):271–293, 1999.
- [17] J. Desharnais et al. *A logical characterization of bisimulation for labeled Markov processes*. In Proc LICS'98, IEEE Press, 1998.
- [18] J. Desharnais, A. Edalat, and P. Panangaden. *Bisimulation for labeled Markov processes*. Information and Computation, 179(2):163–193, Dec 2002.
- [19] J. Desharnais, V. Gupta, R. Jagadeesan, and P. Panangaden. *Metrics for labelled Markov processes*. Theoretical Computer Science, 318(3):323–354, 2004.
- [20] E.-E. Doberkat. *Stochastic Relations. Foundations for Markov Transition Systems*. Chapman and Hall, New York, 2007.
- [21] R. M. Dudley. *Real Analysis and Probability*. Wadsworth and Brookes/Cole, 1989.
- [22] N. Ferns and P. Panangaden and D. Precup. *Metrics for finite Markov Decision Processes*. In proc. of the 20th conference on Uncertainty in Artificial Intelligence, 162–169, 2004.
- [23] S. Givant and P. Halmos. *Introduction to Boolean Algebras*. Undergraduate Texts in Mathematics, Springer-Verlag, 2009.
- [24] R. Goldblatt. *Deduction systems for coalgebras over measurable spaces*. Journal of Logic and Computation, 20(5):1069–1100, 2010.
- [25] R. Goldblatt. *On the role of the Baire category theorem in the foundations of logic*. Journal of Symbolic logic, pages 412–422, 1985.
- [26] J. Harsanyi. *Games with incomplete information played by bayesian players, part one*. Management Sci., 14:159–182, 1967.

- [27] H. Hermans. *Interactive Markov Chains*. LNCS 2428, 2008.
- [28] A. Kurz, *Coalgebras and Modal Logics*. Lecture Notes, ESSLI 2001.
- [29] K. G. Larsen and A. Skou. *Bisimulation through probabilistic testing*. Information and Computation, 94:1–28, 1991.
- [30] Insup Lee, et.al. *High-Confidence Medical Device Software and Systems*. IEEE Computer, Volume 39, Issue 4:33-38, 2006
- [31] R. B. Lyngsø and C. N. Pedersen. *The consensus string problem and the complexity of comparing hidden Markov models*. Journal of Computer and System Sciences, 65(3):545–569, 2002. Special Issue on Computational Biology 2002.
- [32] R. Mardare, M. Cavaliere, S. Sedwards. *A Logical Characterization of Robustness, Mutants and Species in Colonies of Synchronizing Agents*. International Journal of Foundations of Computer Science, IJFCS 19(5):1199-1221, 2008.
- [33] R. Mardare, M. Cavaliere, S. Sedwards. *A Multiset-Based Model of Synchronizing Agents: Computability and Robustness*. Theoretical Computer Science, TCS 391(3):216 - 238, 2008.
- [34] R. Mardare, C. Priami, P. Quaglia, O. Vagin. *Model checking biological systems described using ambient calculus*. In Proc. of the 2nd International Workshop on Computational Methods in Systems Biology, CMSB04, Lecture Notes in Bioinformatics 3082:85-103, Springer, 2005.
- [35] L.S. Moss, I.D. Viglizzo. *Harsanyi type spaces and final coalgebras constructed from satisfied theories*. ENTCS 106:279-295, 2004.
- [36] B. Novak, and J. J. Tyson. *Numerical analysis of a comprehensive model of M-phase control in Xenopus oocyte extracts and intact embryos*. Journal of Cell Science, volume 106: 1153-1168, 1993
- [37] P. Panangaden. *Labelled Markov Processes*. Imperial College Press, 2009.
- [38] C. Priami. *Stochastic  $\pi$ -Calculus*. Computer Journal, 38(7):578-589, 1995.
- [39] H. Rasiowa and R. Sikorski. *A proof of the completeness theorem of Gödel*. Fund. Math, 37:193–200, 1950.
- [40] D. Turi, G.D. Plotkin. *Towards a mathematical operational semantics*, In Proc. LICS'97, IEEE Press, 1997.
- [41] A.M. Turing. *Computing Machinery and Intelligence*, Mind 59, pp 433-460, 1950
- [42] A.M. Turing. *The Chemical Basis of Morphogenesis*, Phil. Trans. R. Soc. London B 237 pp 37-72, 1952

- [43] F. van Breugel, M. Mislove, J. Ouaknine, and J. Worrell. *An intrinsic characterization of approximate probabilistic bisimilarity*. In Proceedings of FOSSACS 03, volume 2620 of Lecture Notes In Computer Science. Springer-Verlag, 2003.
- [44] F. van Breugel, B. Sharma, and J. Worrell. *Approximating a Behavioural Pseudometric without Discount for Probabilistic Systems*. Logical Methods in Computer Science, 4(2), 2008.
- [45] F. van Breugel, J. Worrell. *An algorithm for quantitative verification of probabilistic systems*. LNCS 2154:336-350, 2001.
- [46] C. Zhou. *A complete deductive system for probability logic with application to Harsanyi type spaces*. PhD thesis, Indiana University, 2007.

**Part II**

**Stochastic Process Algebras**



The problem of specifying and analyzing nondeterministic concurrent systems has found a successful solution in the class of *Process Algebras* (PAs) [8]. The compositionality of the processes is reflected by the construction principles of PAs, while their behaviours are transition systems. As a result, one obtains a class of processes with an elegant algebraic-coalgebraic structure, supported by appealing theories and easy to adapt to various modelling requirements.

The same approach has been taken for probabilistic and stochastic concurrent systems. *Probabilistic process algebras* [8], *interactive Markov chain algebra* [11,37] and *stochastic process algebras* (SPA) such as TIPP [32], PEPA [35,36], EMPA [9] and *stochastic Pi-calculus* [47] are extensions of classic PAs. The nondeterminism is replaced by a race policy and this requires important modifications in the semantic format.

Stressed to mimic the pointwise structural operational semantics (SOS) of nondeterministic PAs, the stochastic process algebras find *ad hoc* solutions to the problems introduced by stochasticity, such as the *multi-transition system* approach of PEPA or the *proved SOS* approach of stochastic Pi-calculus. These result in complex constructs that are difficult to extend to a general SOS format for well-behaved stochastic specifications and problematic when recursion or fresh name quantification are considered. As emphasized by Klin and Sassone in [38], for stochastic Pi-calculus of Priami [47] the parallel composition fails to be associative up to bisimulation, while for PEPA, if arbitrary relations between the rates of processes and subprocesses are allowed, stochastic bisimulation ceases to be a congruence. An explanation for these situations is given in [38]: the information carried by the aforementioned SOS frameworks is excessive, while a well-behaved framework should only carry the exact amount of data required for the derivation of the intended semantics.

These problems motivate our research summarized in this part of the monograph, initiated with [17] and extended with [16,18], that aims to reconsider the semantics of SPAs from a perspective faithful to the algebraic-coalgebraic structure of stochastic processes. The key observation is that *structural congruence* induces a  $\sigma$ -algebra on processes and organizes a measurable space of stochastic processes.

We propose a semantics that assigns to each process a set of measures indexed by observable actions. Thus, difficult instance-counting problems that otherwise require complicated versions of SOS can be solved by exploiting the properties of measures (e.g. additivity). Our previous work showed that along this line one obtains an elegant semantics that resembles the one of nondeterministic PAs and provides a well-behaved notion of bisimulation.

Moreover, our approach allow us to propose a natural extension of the equivalence-based behavioural theory, typically centred on concepts such as bisimilarity or trace equivalence, to a metric theory that instead of equating systems with identical behaviours, it quantifies the differences between non-equivalent systems. In this respect, our work relates to [50] where it is proposed a definition of a rule format that guaranties non-expansivity for operators on discrete states probabilistic systems with respect to epsilon-bisimulation metrics.

**Recent related works.** The research on bialgebraic approach for the definition of GSOS formats for probabilistic and stochastic Markov processes with continuous state space, initiated with the research presented in this Chapter, has been further extended in [5, 6]. In [19], it is studied how one can obtain a set of equations that axiomatizes bisimilarity from a set of SOS rules, for discrete state probabilistic systems; this is also extended to the metric setting by proving the equation systems that computes the Kantorovich distance. Other recent research regards the subclasses of the dual of the GSOS format (i.e., lookahead and negative clauses are allowed in the premises) [20, 41]. These works are preliminary attempts to obtain probabilistic extension of the ntyft/ntyxt formats for standard labelled transition systems.

This part of the paper consists in two case studies of stochastic process algebras: the case of Stochastic CCS in Chapter 4 and the case of Stochastic Pi-Calculus with channel-based communication, mobility, fresh name quantification and replication, in Chapter 5. Both these calculi are designed to satisfy the specific requirements of Systems Biology that we have identified in our previous research.

This Part of the Monograph summarizes our work published in the following articles.

- [I] L. Cardelli, R. Mardare. *The Measurable Space of Stochastic Processes*. *Fundamenta Informaticae*, FI volume 131(3-4), pages:351-371, 2014.
- [II] L. Cardelli, R. Mardare. *Stochastic Pi-calculus Revisited*. In Proc. of 10th International Colloquium Theoretical Aspects of Computing, ICTAC 2013, LNCS 8049, pages: 1-21, 2013.
- [III] L. Cardelli, R. Mardare. *The Measurable Space of Stochastic Processes*. In Proc. of Seventh International Conference on the Quantitative Evaluation of Systems, QEST2010, IEEE Computer Society, pages: 171-180. 2010.

# Chapter 4

## Stochastic-CCS

### 4.1 Introduction

The content of this Chapter summarizes the results published in [17, 18]. At the time of the publication of [17], these results opened a new research direction by revealing the necessity to involve measure theory into the research on structural operational semantics for stochastic process algebras.

*Process algebras* (PAs) [8] are formalisms designed to describe the evolution of concurrent communicating systems. For capturing observable behaviors, PAs are conceptualised along two orthogonal axes. From an algebraic point of view, they are endowed with construction principles in the form of algebraic operations that allow composing larger processes from more basic ones; a process is identified by its algebraic term. On the other hand, there exists a notion of nondeterministic evolution, described by a coalgebraic structure, in the form of a transition system. The algebraic and coalgebraic structures are not independent: Structural Operational Semantics (SOS) defines the behavior of a process inductively on its syntactic structure. In this way, classic PAs are supported by an easy and appealing underlying theory that guarantees their success.

In the past decades *probabilistic* and *stochastic* behaviors have also become of central interest due to the applications in performance evaluation and computational systems biology. *Probabilistic process algebras* solve non-determinism by labeling the transitions with probabilities [8, 40, 55]. *Stochastic process algebras* such as TIPP [32], PEPA [35, 36], EMPA [9] and stochastic  $\pi$ -calculus [47] have been defined as extensions of classic PAs, by considering more complex coalgebraic structures. The label of a stochastic transition contains, in addition to the name of the action, the rate of an exponentially distributed random variable that characterizes the duration of the transition. Consequently, SOS associates a non-negative rate value to each tuple  $\langle \text{state}, \text{action}, \text{state} \rangle$ . This additional information imposes important modifications in the SOS format, such as the *multi-transition system* approach of PEPA or the *proved SOS* approach of stochastic  $\pi$ -calculus, mainly because the nondeterminism is replaced by the race policy.

With the intention of developing a stochastic process calculus for applications in Systems Biology, in this chapter we propose a stochastic version of CCS [43] based on the *mass action law* [12] and equipped with an SOS particularly suited to a domain where an equational theory and a measure of similarity of behaviours is important. At the same time we aim to avoid the complicated labeling and counting of previous approaches and to provide an operational semantics that resembles the ones for non-deterministic process algebras, by lifting process-results to measure-results. For doing this, our SOS rules are not given in the pointwise style, but using constructions based on measure theory.

We organise the set of processes as a measurable space and associate to each process an indexed set of measures. Thus, for an action  $a$  and a measurable set  $S$  of processes, the measure  $\mu_a$  associated to a process  $P$  specifies the rate  $\mu_a(S) \in \mathbb{R}^+$  of  $a$ -transitions from  $P$  to (elements of)  $S$ . In this way, difficult instance-counting problems that otherwise require complicated versions of SOS can be solved by exploiting the properties of measures (e.g. additivity). Similar ideas have been proposed for probabilistic automata [39, 49] and Markov processes [11, 37, 46]. Following the *transition-systems-as-coalgebras* paradigm [24, 48], this approach follows naturally in the sequence started by nondeterministic and probabilistic transition systems.

The idea of transitions from states to measures has been previously advocated in the context of probabilistic automata [39, 49] and Markov processes [46]. The *transition-systems-as-coalgebras* paradigm [24, 48] exploits it providing a uniform characterisation of transition systems that covers the sequence nondeterministic, probabilistic and stochastic systems. A general SOS format for SPAs without new name operators or recursion is proposed in [38]. In [23] these ideas are applied to particular SPAs with pointwise semantics. With respect to these works, in our research we consider a different measurable space that not only answers to practical modelling requirements, but also simplifies the semantics and gives us smooth solutions for the complex technical problems, such as fresh name quantification and replication addressed in Chapter 5, without requiring additional constructs.

The novelty of our approach derives firstly from the structure of the measurable space of stochastic processes. This space is organised by *structural congruence*, an equivalence that equates processes that are indistinguishable from a modeling perspective. For instance, if we model the parallel evolution of two processes, say  $Q$  and  $R$ , we expect no difference between  $Q|R$ ,  $R|Q$  and  $R|Q|0$  ( $0$  denotes an inactive process). This relation is required in systems biology where it models chemical mixing: structural congruence was invented in the first place from a chemical analogy [7]. In effect, our  $\sigma$ -algebra is generated by the structural congruence classes and our stochastic transitions are defined from processes to measurable (structural congruence-closed) sets of processes. In this way, if  $P$  can perform an action  $a$  with a rate  $r$  to  $Q|R$ , written  $P \xrightarrow{a,r} Q|R$ , we can also derive

$$P \xrightarrow{a,r} R|Q \text{ and } P \xrightarrow{a,r} R|(Q|0).$$

Otherwise, the alternative approach of considering any set of processes measurable permits to calculate the rate of the  $a$ -transitions from  $P$  to the set

$$\{Q|R, R|Q, R|(Q|0)\}$$

and obtain the undesired result

$$P \xrightarrow{a, 3r} \{Q|R, R|Q, R|(Q|0)\}.$$

The Example 4.5.5 in Section 4.5 presents an anomaly that derives from a pointwise semantics for stochastic SOS and underlines the kind of problems we want to address in this part of the monograph.

Anticipating the development of our theory, consider the stochastic CCS processes

$$A = b.0|c.0 \text{ and } B = b.c.0 + c.b.0.$$

As in the case of the classic CCS, we expect them to be bisimilar and consequently, the processes

$$P = \tau_r.A + \tau_r.B, \quad Q = \tau_r.A + \tau_r.A \text{ and } R = \tau_r.B + \tau_r.B$$

have to be bisimilar as well, where  $\tau_r$  denotes a  $\tau$ -action with rate  $r$ .

However, a pointwise semantics will prove that

$$\begin{aligned} P &\xrightarrow{\tau, r} A, & P &\xrightarrow{\tau, r} B, \\ Q &\xrightarrow{\tau, 2r} A, & Q &\xrightarrow{\tau, 0} B, \\ R &\xrightarrow{\tau, 0} A \text{ and } R &\xrightarrow{\tau, 2r} B, \end{aligned}$$

implying that the three processes are not bisimilar. To avoid such undesirable anomalies, in the literature have been proposed complicated variants of SOS that make the underlying theory heavy and problematic. In particular they propose to handle each structural congruence case by a different dedicated SOS rule. While this might work in "linear" syntaxes such as CCS or Pi-Calculus, in hierarchical calculi, such as Ambient Calculus [15], Brane Calculus [14], etc., this approach recalls an infinite set of SOS rules that does not have a finitary representation.

The theory of GSOS [51] has been extended for the case of stochastic systems in [38], where general congruence formats for stochastic GSOS (SGSOS) are studied. The SGSOS framework, as well as GSOS, focuses on the *monads freely generated* by the algebraic signature of a process calculus. Our case is different: we have an *equational monad* because the structural congruence provides extra structure for the class of processes and thus we get a different type of SOS. In our format, for instance, the algebraic signature of processes is different from the algebraic signature of behaviours. Using a non-discrete  $\sigma$ -algebra makes our approach different, while considering the measurable sets closed to some congruence relation makes it more appropriate for modelling and for extensions

to other *equational theories*. Recent papers followed and generalized the ideas proposed in this part of our monograph [4,5].

Our choice of developing a stochastic process algebra under the restrictions of the equational theory of structural congruence not only that solves the aforementioned problems, but it is sustained by an elegant SOS that supports a smooth development of the basic theory, which in the next chapter will be further generalized for stochastic Pi-calculus.

The structures we work with, simply called in this part of the monograph *Markov processes* (MPs), are the class of the stochastic Markov processes with continuous state-space and continuous-time evolution defined in [28]. In the Preliminaries we denoted the set of these processes over the measurable space  $(M, \Sigma)$  by  $\Delta(M, \Sigma)$  to underline the fact that we use general distributions. These MPs extend the notions of *labelled Markov process* [10,26,27] and *Harsanyi type space* [34,45] on to the stochastic level. However, MPs are more general than the continuous-time Markov chains because only the transitions to measurable sets are permitted (which are never singletons) and cannot be described in a pointwise style.

We also introduce a notion of *stochastic bisimulation* for MPs, along the lines of [26–28, 40]. It generalizes *rate aware bisimulation* of [23], being defined for arbitrary measurable spaces and closed to an equational theory. We prove that stochastic bisimulation is a congruence that extends structural congruence.

Another advantage of our approach consists in the fact that it can be naturally extended to define a class of metrics on stochastic processes which measure the similarity of process behaviours. This result has considerable practical application. The standard notion of bisimulation for probabilistic or stochastic systems cannot distinguish between two processes that are substantially different and two processes that differ by only a small amount in a real valued parameter. It is often more useful to say how similar two processes are than to say whether they are exactly the same. This is precisely what our metrics do: stochastic bisimilar processes are at distance zero, processes that differ by small values of rates are closer than the processes with bigger differences.

The Chapter is organised as follows. Section 4.2 defines the general concept of Markov process (MP) and the stochastic bisimulation of MPs. Section 4.3 introduces the syntax of our process algebra and the axiomatization of structural congruence; we prove that the space of processes can be organised as a Markov kernel and that each process is an MP. These results guide us, in Section 5.2.4, to the definition of a structural operational semantics which induces a notion of behavioural equivalence that coincides with the bisimulation of MPs. In Section 4.5 we show that the bisimulation behaves well with respect to the algebraic structure of processes: stochastic bisimulation is a congruence. This relation is extended in Chapter 6 with a class of metrics on the space of processes that measure how similar two processes are. We also have a section dedicated to related work and a concluding section.

## 4.2 Basic Concepts and Notations

In this section we establish the basic terminology and notations used in this chapter.

The concept of Markov processes used in what follows is slightly different of the same concept introduced in preliminaries. The difference consists in the fact that it involves a set  $A$  of labels. The labels  $\alpha \in A$  represent types of interactions with the environment. If  $m$  is the current state of the system and  $N$  is a measurable set of states,  $\theta(\alpha)(m)$  is a measure on the state space and  $\theta(\alpha)(m)(N) \in \mathbb{R}^+$  represents the *rate* of the exponentially distributed random variable that characterizes the duration of an  $\alpha$ -transition from  $m$  to arbitrary  $n \in N$ . Indeterminacy is solved by races between events executing at different rates.

**Definition 4.2.1** (Labeled Markov kernels and labeled Markov processes). *Let  $(M, \Sigma)$  be an analytic space, where  $\Sigma$  represents the Borel algebra, and  $A$  a countable set disjoint of  $M$ . An  $A$ -Markov kernel is a tuple  $\mathcal{M} = (M, \Sigma, \theta)$ , with*

$$\theta : A \rightarrow \llbracket M \rightarrow \Delta(M, \Sigma) \rrbracket,$$

where  $\llbracket M \rightarrow \Delta(M, \Sigma) \rrbracket$  denotes the set of measurable functions between  $(M, \Sigma)$  and  $\Delta(M, \Sigma)$ . If  $m \in M$  then the tuple  $(M, \Sigma, \theta, m)$  is an  $A$ -Markov process of  $\mathcal{M}$  and  $m$  is its initial state.

Notice that  $\theta(\alpha)$  is defined as a measurable mapping between  $(M, \Sigma)$  and the measurable space  $\Delta(M, \Sigma)$  of the measures on  $(M, \Sigma)$ . This condition is equivalent to the conditions on the two-variable *rate function* used in [28, 46] (see, e.g. Proposition 2.9, of [29]). If  $\mathcal{M} = (M, \Sigma, \theta)$ , we sometimes denote the process  $(M, \Sigma, \theta, m)$  by  $(\mathcal{M}, m)$ .

We define the rate-bisimulation relations on MPs following the similar definitions of [26, 28, 46].

Given a binary relation  $\mathfrak{R} \subseteq M \times M$  on a set  $M$ , we call a subset  $N \subseteq M$   $\mathfrak{R}$ -closed iff

$$\{m \in M \mid \exists n \in N, (n, m) \in \mathfrak{R}\} \subseteq N.$$

If  $(M, \Sigma)$  is a measurable space,  $\Sigma(\mathfrak{R})$  denotes the set of measurable  $\mathfrak{R}$ -closed subsets of  $M$ .

**Definition 4.2.2** (Rate-bisimulation). *Let  $\mathcal{M} = (M, \Sigma, \theta)$  be an  $A$ -Markov kernel. A rate-bisimulation relation is an equivalence relation  $\mathfrak{R} \subseteq M \times M$  such that  $(m, n) \in \mathfrak{R}$  iff for any  $C \in \Sigma(\mathfrak{R})$  and any  $\alpha \in A$ ,*

$$\theta(\alpha)(m)(C) = \theta(\alpha)(n)(C).$$

*Two MPs  $(\mathcal{M}, m)$  and  $(\mathcal{M}, n)$  are stochastic bisimilar, written  $m \sim_{\mathcal{M}} n$ , if  $m$  and  $n$  are related by a rate-bisimulation relation. The stochastic bisimilarity is the largest rate-bisimulation.*

Observe in the previous definition the role of  $\mathfrak{R}$  in the structure of  $C$ . This definition reflects the coinductive nature of this definition. Notice also that, for any  $A$ -MP  $(M, \Sigma, \theta)$ , there exist rate-bisimulation relations: for instance, the identity of the elements of  $M$  is a rate-bisimulation relation. However exist also rate-bisimulations that are non-trivial as we will see in what follows.

### 4.3 A minimal Stochastic Process Algebra

In this section we introduce a stochastic extension of CCS without replication [43].

As usual in stochastic process algebras, each transition  $a$  has associated a *rate* in  $\mathbb{R}^+$  representing the parameter of an exponentially distributed random variable that characterizes the duration of an  $a$ -action. In addition, we also consider synchronizations of actions. As in CCS, the set of actions is equipped with an *involution* that associates with each action  $a$  its paired action  $\bar{a}$ ; the paired actions have the same rates. The synchronization of  $(a, \bar{a})$  counts as an *internal  $\tau$ -action* with the rate satisfying the *mass action law* [12].

Formally, the set of *labels (actions)* is a countable set  $\mathbb{A}$  endowed with

1. an *involution* associating to each  $a \in \mathbb{A}$  an element  $\bar{a} \in \mathbb{A}$  such that  $a \neq \bar{a}$  and  $\overline{\bar{a}} = a$ ;
2. a *weight function*  $\iota : \mathbb{A} \rightarrow \mathbb{Q}^+$ , such that for any  $a \in \mathbb{A}$ ,  $\iota(a) = \iota(\bar{a})$ .

In what follows we use two extensions of  $\mathbb{A}$  defined for the *internal action*  $\tau \notin \mathbb{A}$ . On syntactic level we involve the set

$$\mathbb{A}^* = \mathbb{A} \cup \{\tau_r \mid r \in \mathbb{Q}^+\},$$

where indexed internal actions will be used for modelling delays in a system<sup>1</sup>.

We extend  $\iota$  to  $\mathbb{A}^*$  by  $\iota(\tau_r) = r$ . For operational semantics we use the set

$$\mathbb{A}^+ = \mathbb{A} \cup \{\tau\}$$

of labels.

We use  $a, a', a_i$  to denote arbitrary elements of  $\mathbb{A}$ ,  $\varepsilon, \varepsilon', \varepsilon_i$  to denote arbitrary elements of  $\mathbb{A}^*$  and  $\alpha, \alpha', \alpha_i$  to denote arbitrary elements of  $\mathbb{A}^+$ .

**Definition 4.3.1.**  $\mathbb{A}$ -stochastic processes are defined, for arbitrary  $\varepsilon \in \mathbb{A}^*$ , as follows

$$P ::= 0 \mid \varepsilon.P \mid P \mid P \mid P + P.$$

We denote by  $\mathbb{P}$  the set of stochastic processes.

An essential notion for processes is the *structural congruence relation* which equates processes that, in spite of their different syntactic forms, they represent the same systems.

**Definition 4.3.2** (Structural congruence). *Structural congruence*  $\equiv \subseteq \mathbb{P} \times \mathbb{P}$  is the smallest equivalence relation satisfying, for arbitrary  $P, Q, R \in \mathbb{P}$  and  $\varepsilon \in \mathbb{A}^*$  the following conditions:

1.  $P \mid Q \equiv Q \mid P$

---

<sup>1</sup>In practice we cannot measure nor specify models with irrational rates and for this reason we have chosen  $\iota(\varepsilon) \in \mathbb{Q}^+$  for all  $\varepsilon \in \mathbb{A}^*$ . However, the technical development does not change if  $\iota : \mathbb{A} \rightarrow \mathbb{R}^+$ .

2.  $(P|Q)|R \equiv P|(Q|R)$
3.  $P|0 \equiv P$
4.  $P + Q \equiv Q + P$
5.  $(P + Q) + R \equiv P + (Q + R)$
6.  $P + 0 \equiv P$
7.  $P \equiv Q \implies P|R \equiv Q|R$
8.  $P \equiv Q \implies P + R \equiv Q + R$
9.  $P \equiv Q \implies \varepsilon.P \equiv \varepsilon.Q$

Let  $\mathbb{P}^{\equiv}$  be the set of  $\equiv$ -equivalence classes on  $\mathbb{P}$ . For arbitrary  $P \in \mathbb{P}$ , we denote by  $P^{\equiv}$  the  $\equiv$ -equivalence class of  $P$ .

The set of stochastic processes is organized as a measurable space  $(\mathbb{P}, \Pi)$ , where  $\Pi$  is the  $\sigma$ -algebra generated by  $\mathbb{P}^{\equiv}$ .

Note that  $\mathbb{P}^{\equiv}$  is a base for the  $\sigma$ -algebra  $\Pi$  and the measurable sets are (possibly countable) unions of  $\equiv$ -equivalence classes on  $\mathbb{P}$ .

In what follows we use  $\mathcal{P}, \mathcal{P}_i, \mathcal{R}, \mathcal{Q}$  to denote arbitrary measurable sets of  $\Pi$ .

Consider the following operations on the measurable sets in  $\Pi$ , defined for arbitrary  $\mathcal{P}, \mathcal{Q} \in \Pi$  and  $P \in \mathbb{P}$  as follows.

$$\mathcal{P}|Q = \bigcup_{P \in \mathcal{P}, Q \in \mathcal{Q}} (P|Q)^{\equiv},$$

$$\mathcal{P}_P = \bigcup_{P|R \in \mathcal{P}} R^{\equiv}.$$

Notice that, by construction,  $\mathcal{P}|Q$  and  $\mathcal{P}_P$  are measurable sets.

Now we show that the measurable space  $(\mathbb{P}, \Pi)$  of stochastic processes can be organized as an  $\mathbb{A}^+$ -Markov kernel. This will implicitly provide a structural operational semantics for our process algebra such that the behavioural equivalence coincides with the bisimulation of MPs.

For the beginning, notice that  $(\mathbb{P}, \Pi)$  is a Polish space, hence, analytic space.

Next, we define a function  $\theta: \mathbb{A}^+ \rightarrow \llbracket \mathbb{P} \rightarrow \Delta(\mathbb{P}, \Pi) \rrbracket$  which organizes  $(\mathbb{P}, \Pi, \theta)$  as an  $\mathbb{A}^+$ -Markov kernel. In this interpretation, for arbitrary  $P \in \mathbb{P}$ ,  $\mathcal{P} \in \Pi$  and  $\alpha \in \mathbb{A}^+$ ,  $\theta(\alpha)(P)(\mathcal{P})$  will represent the total rate of the  $\alpha$  actions from  $P$  to (elements of)  $\mathcal{P}$ .

We use  $\omega$  to denote the *null measure* on  $(\mathbb{P}, \Pi)$ ; and for  $P \in \mathbb{P}$  and  $r \in \mathbb{R}^+$ , let  $D(r, P)$  be the *r-Dirac measure at P* defined, for arbitrary  $\mathcal{P} \in \Pi$ , by  $D(r, P)(\mathcal{P}) = r$  if  $P \in \mathcal{P}$  and  $D(r, P)(\mathcal{P}) = 0$  otherwise.

**Definition 4.3.3.** Let  $\theta: \mathbb{A}^+ \rightarrow [\mathbb{P} \rightarrow \Delta(\mathbb{P}, \Pi)]$  be defined, by induction on the structure of  $P \in \mathbb{P}$ , as follows

**The case  $P = 0$ :** For any  $\alpha \in \mathbb{A}^+$ , let  $\theta(\alpha)(0) = \omega$ .

**The case  $P = \varepsilon.Q$  with  $\varepsilon \in \mathbb{A}^*$ :** For arbitrary  $a \in \mathbb{A}$ , let

$$\theta(\tau)(\varepsilon.Q) = \begin{cases} D(\iota(\varepsilon), Q), & \varepsilon \notin \mathbb{A} \\ \omega, & \varepsilon \in \mathbb{A} \end{cases}$$

$$\theta(a)(\varepsilon.Q) = \begin{cases} D(\iota(\varepsilon), Q), & \varepsilon = a \\ \omega, & \varepsilon \neq a \end{cases}$$

**The case  $P = Q + R$ :** For any  $\alpha \in \mathbb{A}^+$  and  $\mathcal{P} \in \Pi$ ,

$$\theta(\alpha)(Q + R)(\mathcal{P}) = \theta(\alpha)(Q)(\mathcal{P}) + \theta(\alpha)(R)(\mathcal{P}).$$

**The case  $P = Q|R$ :** For any  $a \in \mathbb{A}$  and  $\mathcal{P} \in \Pi$ ,

$$\theta(a)(Q|R)(\mathcal{P}) = \theta(a)(R)(\mathcal{P}_Q) + \theta(a)(Q)(\mathcal{P}_R),$$

$$\begin{aligned} \theta(\tau)(Q|R)(\mathcal{P}) &= \theta(\tau)(R)(\mathcal{P}_Q) + \theta(\tau)(Q)(\mathcal{P}_R) + \\ &\quad \sum_{\substack{a \in \mathbb{A}, \iota(a) \neq 0 \\ \mathcal{P}_1 | \mathcal{P}_2 \subseteq \mathcal{P}}} \frac{\theta(a)(Q)(\mathcal{P}_1) \cdot \theta(\bar{a})(R)(\mathcal{P}_2)}{2 \cdot \iota(a)}. \end{aligned}$$

If we define the set of active actions of a process  $P \in \mathbb{P}$  by

$$\begin{aligned} act(0) &= \emptyset, \\ act(a.P) &= \{a\}, \\ act(P + Q) &= act(P) \cup act(Q), \\ act(P|Q) &= act(P) \cup act(Q), \end{aligned}$$

then any process has only a finite set of active actions.

Notice that  $\theta(a)(P) \neq \omega$  iff  $a \in act(P)$ . This means that for any  $a \notin act(P)$  and any  $\mathcal{R} \in \Pi$ ,  $\theta(a)(P)(\mathcal{R}) = 0$ . Consequently, the infinitary sum involved in Definition 4.3.3, the case  $P = Q|R$ , has a finite number of non-zero summands. Regarding this sum, observe that it is divided by 2 because we count the interaction pairs  $(a, \bar{a})$  twice since  $a = \bar{\bar{a}}$ ; and it is divided by  $\iota(a)$  to guarantee that the mass action law is satisfied.

The next theorem states that the space of processes with the function defined above is an  $\mathbb{A}^+$ -Markov kernel. It implicitly states the correctness of the previous definition: for each  $\alpha \in \mathbb{A}^+$  and each  $P \in \mathbb{P}$ ,  $\theta(\alpha)(P) \in \Delta(\mathbb{P}, \Pi)$ . It follows that for each  $\alpha \in \mathbb{A}^+$ ,  $\theta(\alpha) \in [\mathbb{P} \rightarrow \Delta(\mathbb{P}, \Pi)]$ .

**Theorem 4.3.4.**  $(\mathbb{P}, \Pi, \theta)$  is an  $\mathbb{A}^+$ -Markov kernel.

*Proof.* It is sufficient to show that for each  $P \in \mathbb{P}$  and each  $\alpha \in \mathbb{A}^+$ ,  $\theta(\alpha)(P)$  is a measure on  $(\mathbb{P}, \Pi)$ .

The proof follows the inductive steps of Definition 4.3.3.

**The cases**  $P = 0$  and  $P = \varepsilon.Q$  with  $\varepsilon \in \mathbb{A}^*$  are trivial.

**The case**  $P = Q + R$ :

Firstly notice that for any  $\alpha \in \mathbb{A}$ ,  $\theta(\alpha)(Q + R)(\emptyset) = \theta(\alpha)(Q)(\emptyset) + \theta(\alpha)(R)(\emptyset) = 0$ .

Consider now an arbitrary sequence of pairwise disjoint sets  $(\mathcal{R}_i)_{i \in I} \in \Pi$ . Then, for arbitrary  $\alpha \in \mathbb{A}$ ,

$$\theta(\alpha)(Q + R)(\cup_{i \in I} \mathcal{R}_i) = \theta(\alpha)(Q)(\cup_{i \in I} \mathcal{R}_i) + \theta(\alpha)(R)(\cup_{i \in I} \mathcal{R}_i).$$

The inductive hypothesis guarantees that

$$\theta(\alpha)(Q)(\cup_{i \in I} \mathcal{R}_i) = \sum_{i \in I} \theta(\alpha)(Q)(\mathcal{R}_i) \quad \text{and}$$

$$\theta(\alpha)(R)(\cup_{i \in I} \mathcal{R}_i) = \sum_{i \in I} \theta(\alpha)(R)(\mathcal{R}_i).$$

Consequently,

$$\begin{aligned} \theta(\alpha)(Q + R)(\cup_{i \in I} \mathcal{R}_i) &= \sum_{i \in I} \theta(\alpha)(Q)(\mathcal{R}_i) + \sum_{i \in I} \theta(\alpha)(R)(\mathcal{R}_i) \\ &= \sum_{i \in I} (\theta(\alpha)(Q)(\mathcal{R}_i) + \theta(\alpha)(R)(\mathcal{R}_i)) = \sum_{i \in I} \theta(\alpha)(Q + R)(\mathcal{R}_i). \end{aligned}$$

**The case**  $P \equiv Q|R$ :

For  $a \in \mathbb{A}$ ,

$$\theta(a)(Q|R)(\emptyset) = \theta(a)(R)(\emptyset_Q) + \theta(a)(Q)(\emptyset_R) = 0,$$

because using the inductive hypothesis one can get that  $\theta(a)(R)$  and  $\theta(a)(Q)$  are measures and  $\emptyset_Q = \emptyset_R = \emptyset$ . Moreover,

$$\begin{aligned} \theta(\tau)(Q|R)(\emptyset) &= \theta(\tau)(R)(\emptyset_Q) + \theta(\tau)(Q)(\emptyset_R) + \\ &+ \sum_{\mathcal{P}_1 | \mathcal{P}_2 \subseteq \emptyset} \frac{\theta(a)(Q)(\mathcal{P}_1) \cdot \theta(\bar{a})(R)(\mathcal{P}_2)}{2 \cdot \iota(a)} = 0. \end{aligned}$$

This is because  $\emptyset_Q = \emptyset_R = \emptyset$  and  $\mathcal{P}_1 | \mathcal{P}_2 \subseteq \emptyset$  implies  $\mathcal{P}_1 = \mathcal{P}_2 = \emptyset$  and the inductive hypothesis guarantees that

$$\theta(\tau)(R)(\emptyset) = \theta(\tau)(Q)(\emptyset) = \theta(\bar{a})(R)(\emptyset) = \theta(a)(Q)(\emptyset) = 0.$$

Consider now an arbitrary sequence of pairwise disjoint sets  $(\mathcal{R}^i)_{i \in I} \in \Pi$  and let  $\mathcal{P} = \cup_{i \in I} \mathcal{R}^i$ . Observe that  $\mathcal{R}_Q^i$  and  $\mathcal{R}_R^i$  are pairwise disjoint. Consequently,

$$\theta(a)(Q|R)(\mathcal{P}) = \sum_{i \in I} [\theta(a)(R)(\mathcal{R}_Q^i) + \theta(a)(Q)(\mathcal{R}_R^i)] = \sum_{i \in I} \theta(a)(Q|R)(\mathcal{R}^i).$$

$$\begin{aligned} \theta(\tau)(Q|R)(\mathcal{P}) &= \theta(\tau)(R)(\mathcal{P}_Q) + \theta(\tau)(Q)(\mathcal{P}_R) + \sum_{\mathcal{P}_1 | \mathcal{P}_2 \subseteq \mathcal{P}}^{a \in \mathbb{A}} \frac{\theta(a)(Q)(\mathcal{P}_1) \cdot \theta(\bar{a})(R)(\mathcal{P}_2)}{2 \cdot \iota(a)} \\ &= \sum_{i \in I} \theta(\tau)(R)(\mathcal{R}_Q^i) + \sum_{i \in I} \theta(\tau)(Q)(\mathcal{R}_R^i) + \sum_{i \in I} \sum_{\mathcal{P}_1 | \mathcal{P}_2 \in \mathcal{R}_i}^{a \in \mathbb{A}} \frac{\theta(a)(Q)(\mathcal{P}_1) \cdot \theta(\bar{a})(R)(\mathcal{P}_2)}{2 \cdot \iota(a)} = \\ &= \sum_{i \in I} \theta(\tau)(Q|R)(\mathcal{R}_i). \end{aligned}$$

□

A consequence of the previous theorem is that for each  $P \in \mathbb{P}$ ,  $(\mathbb{P}, \Pi, \theta, P)$  is a Markov process. In effect, we can define a stochastic bisimulation for the elements of our process algebra simply as stochastic bisimulation of Markov processes in  $(\mathbb{P}, \Pi, \theta)$ .

## 4.4 Structural Operational Semantics

In this section we introduce the structural operational semantics for the minimal process algebra, with the intention to induce a behavioural equivalence on processes that coincides with their bisimulation as MPs. In this case we do not associate to each tuple  $(process, action, process)$  a rate, as usual in stochastic process algebras, because a transition in our case is not between two processes, but from a process to an infinite measurable set of processes. However, our intention is to maintain “the spirit” of process algebras and for this reason we use “generalised” transitions of type  $P \rightarrow \mu$  where  $\mu: \mathbb{A}^+ \rightarrow \Delta(\mathbb{P}, \Pi)$  is a function defining a class of  $\mathbb{A}^+$ -indexed measures on  $(\mathbb{P}, \Pi)$ .

For simplifying the rules of the operational semantics, we first define some operations on the functions in  $\Delta(\mathbb{P}, \Pi)^{\mathbb{A}^+}$  that are required in the process of lifting the algebraic structure of processes to the level of distribution over the space of processes. We analyze firstly the mathematical structures and properties of these operations.

We say that a function  $\mu \in \Delta(\mathbb{P}, \Pi)^{\mathbb{A}^+}$  has *finite support* if  $\mathbb{A} \setminus \mu^{-1}(\omega)$  is finite or empty.

**Definition 4.4.1** (Operations on indexed-distributions.).

1. Let  $\bar{\omega}: \mathbb{A}^+ \rightarrow \Delta(\mathbb{P}, \Pi)$  be defined by

$$\bar{\omega}(\alpha) = \omega, \text{ for any } \alpha \in \mathbb{A}^+.$$

2. For arbitrary  $\varepsilon \in \mathbb{A}^*$  and  $P \in \mathbb{P}$ , let  $[\varepsilon_P]: \mathbb{A}^+ \rightarrow \Delta(\mathbb{P}, \Pi)$  be defined, for arbitrary  $a \in \mathbb{A}$ , as follows

$$[\varepsilon_P](\tau) = \begin{cases} D(\iota(\varepsilon), P), & \varepsilon \notin \mathbb{A} \\ \omega, & \varepsilon \in \mathbb{A} \end{cases}$$

$$[\varepsilon_P](a) = \begin{cases} D(\iota(\varepsilon), P), & a = \varepsilon \\ \omega, & a \neq \varepsilon \end{cases}$$

3. For arbitrary  $\mu', \mu'' \in \Delta(\mathbb{P}, \Pi)^{\mathbb{A}^+}$ , let  $\mu' \oplus \mu'': \mathbb{A}^+ \rightarrow \Delta(\mathbb{P}, \Pi)$  be defined, for  $\alpha \in \mathbb{A}^+$ , as follows

$$(\mu' \oplus \mu'')(\alpha) = \mu'(\alpha) + \mu''(\alpha).$$

4. For arbitrary  $\mu', \mu'' \in \Delta(\mathbb{P}, \Pi)^{\mathbb{A}^+}$  with finite support and arbitrary  $P, Q \in \mathbb{P}$ , let  $\mu' \text{ }_P \otimes_Q \mu'': \mathbb{A}^+ \rightarrow \Delta(\mathbb{P}, \Pi)$  be defined as follows

$$(\mu' \text{ }_P \otimes_Q \mu'')(a)(\mathcal{R}) = \mu'(a)(\mathcal{R}_Q) + \mu''(a)(\mathcal{R}_P) \text{ for } a \in \mathbb{A}$$

$$(\mu' \text{ }_P \otimes_Q \mu'')(\tau)(\mathcal{R}) = \mu'(\tau)(\mathcal{R}_Q) + \mu''(\tau)(\mathcal{R}_P) + \sum_{\substack{a \in \mathbb{A}, \iota(a) \neq 0 \\ \mathcal{P}_1 | \mathcal{P}_2 \subseteq \mathcal{R}}} \frac{\mu'(a)(\mathcal{P}_1) \cdot \mu''(\bar{a})(\mathcal{P}_2)}{2 \cdot \iota(a)}.$$

Because  $\mu'$  and  $\mu''$  have finite support, the sum involved in the definition of  $\text{ }_P \otimes_Q$  has a finite number of non-zero summands.

The next lemma proves that the definitions of  $\oplus$  and  $\text{ }_P \otimes_Q$  for arbitrary  $P, Q \in \mathbb{P}$  are correct; it also states some basic properties of these operators.

**Lemma 4.4.2.**

1. For arbitrary  $\mu, \mu', \mu'' \in \Delta(\mathbb{P}, \Pi)^{\mathbb{A}^+}$ ,  $\mu \oplus \mu' \in \Delta(\mathbb{P}, \Pi)^{\mathbb{A}^+}$  and
  - (a)  $\mu \oplus \mu' = \mu' \oplus \mu$ ,
  - (b)  $(\mu \oplus \mu') \oplus \mu'' = \mu \oplus (\mu' \oplus \mu'')$ ,
  - (c)  $\mu = \mu \oplus \bar{\omega}$ .
2. For arbitrary  $P, Q, R \in \mathbb{P}$  and arbitrary  $\mu', \mu'', \mu''' \in \Delta(\mathbb{P}, \Pi)^{\mathbb{A}^+}$  with finite support,  $\mu \text{ }_P \otimes_Q \mu' \in \Delta(\mathbb{P}, \Pi)^{\mathbb{A}^+}$  and
  - (a)  $\mu' \text{ }_P \otimes_Q \mu'' = \mu'' \text{ }_Q \otimes_P \mu'$ ,
  - (b)  $(\mu' \text{ }_P \otimes_Q \mu'') \text{ }_{P|Q} \otimes_R \mu''' = \mu' \text{ }_P \otimes_{Q|R} (\mu'' \text{ }_Q \otimes_R \mu''')$ ,
  - (c)  $\mu' \text{ }_P \otimes_Q \bar{\omega} = \mu'$ .
3. For arbitrary  $P, P', Q, Q' \in \mathbb{P}$ , arbitrary  $\varepsilon \in \mathbb{A}^+$  and arbitrary  $\mu', \mu'' \in \Delta(\mathbb{P}, \Pi)^{\mathbb{A}^+}$  with finite support,
  - (a) if  $P \equiv P'$  and  $Q \equiv Q'$ , then  $\mu' \text{ }_P \otimes_Q \mu'' = \mu' \text{ }_{P'} \otimes_{Q'} \mu''$ ,

(b) if  $P \equiv Q$ , then  $[\varepsilon]_P = [\varepsilon]_Q$ .

*Proof.* We only prove 2(b) and 2(c), the other cases being trivial.

**2(b).** Let  $\mu = \mu' \otimes_P \mu''$  and consider arbitrary  $a \in \mathbb{A}$  and  $\mathcal{R} \in \Pi$ .

$$\begin{aligned} ((\mu' \otimes_P \mu'') \otimes_{P|Q} \mu''') (a)(\mathcal{R}) &= (\mu \otimes_{P|Q} \mu''') (a)(\mathcal{R}) = \\ &= \mu(a)(\mathcal{R}_R) + \mu'''(a)(\mathcal{R}_{P|Q}) \end{aligned}$$

However,  $\mu(a)(\mathcal{R}) = \mu'(a)(\mathcal{R}) \otimes_Q \mu''(a)(\mathcal{R}) = \mu'(a)(\mathcal{R}_Q) + \mu''(a)(\mathcal{R}_P)$ .

$$\begin{aligned} ((\mu' \otimes_P \mu'') \otimes_{P|Q} \mu''') (a)(\mathcal{R}) &= \\ &= (\mu'(a)(\mathcal{R}_{R|Q}) + \mu''(a)(\mathcal{R}_{R|P})) + \mu'''(a)(\mathcal{R}_{P|Q}). \end{aligned}$$

Observe that for arbitrary  $P, Q \in \mathbb{P}$  and arbitrary  $\mathcal{R} \in \Pi$ ,  $(\mathcal{R}_P)_Q = \mathcal{R}_{P|Q}$ . Using this, we obtain

$$\begin{aligned} ((\mu' \otimes_P \mu'') \otimes_{P|Q} \mu''') (a)(\mathcal{R}) &= \\ &= \mu'(a)(\mathcal{R}_{Q|R}) + \mu''(a)(\mathcal{R}_{P|R}) + \mu'''(a)(\mathcal{R}_{P|Q}). \end{aligned}$$

In the same way we can prove that

$$\begin{aligned} \mu' \otimes_{P|Q|R} (\mu'' \otimes_Q \mu''') (a)(\mathcal{R}) &= \\ &= \mu'(a)(\mathcal{R}_{Q|R}) + \mu''(a)(\mathcal{R}_{P|R}) + \mu'''(a)(\mathcal{R}_{P|Q}). \end{aligned}$$

Now we prove that

$$((\mu' \otimes_P \mu'') \otimes_{P|Q} \mu''') (\tau)(\mathcal{R}) = (\mu' \otimes_{P|Q|R} (\mu'' \otimes_Q \mu''')) (\tau)(\mathcal{R}).$$

As before, we have

$$\begin{aligned} ((\mu' \otimes_P \mu'') \otimes_{P|Q} \mu''') (\tau)(\mathcal{R}) &= (\mu \otimes_{P|Q} \mu''') (\tau)(\mathcal{R}) = \\ &= \mu(\tau)(\mathcal{R}_R) + \mu'''(\tau)(\mathcal{R}_{P|R}) + \sum_{\mathcal{P}_1 | \mathcal{P}_2 \subseteq \mathcal{R}}^{a \in \mathbb{A}} \frac{\mu(a)(\mathcal{P}_1) \cdot \mu'''(\bar{a})(\mathcal{P}_2)}{2 \cdot \iota(a)} = \\ &= \mu'(\tau)(\mathcal{R}_{R|Q}) + \mu''(\tau)(\mathcal{R}_{R|P}) + \mu'''(\tau)(\mathcal{R}_{P|R}) + \\ &= \sum_{\mathcal{Q}_1 | \mathcal{Q}_2 \subseteq \mathcal{R}}^{b \in \mathbb{A}} \frac{\mu'(b)(\mathcal{Q}_1) \cdot \mu''(\bar{b})(\mathcal{Q}_2)}{2 \cdot \iota(b)} + \sum_{\mathcal{P}_1 | \mathcal{P}_2 \subseteq \mathcal{R}}^{a \in \mathbb{A}} \frac{\mu(a)(\mathcal{P}_1) \cdot \mu'''(\bar{a})(\mathcal{P}_2)}{2 \cdot \iota(a)}. \end{aligned}$$

However,

$$\sum_{\mathcal{P}_1 | \mathcal{P}_2 \subseteq \mathcal{R}}^{a \in \mathbb{A}} \frac{\mu(a)(\mathcal{P}_1) \cdot \mu'''(\bar{a})(\mathcal{P}_2)}{\iota(a) \cdot 2} =$$

$$\begin{aligned} \sum_{\mathcal{P}_1 | \mathcal{P}_2 \subseteq \mathcal{R}}^{a \in \mathbb{A}} \frac{[\mu'(a)((\mathcal{P}_1)_Q) + \mu''(a)((\mathcal{P}_1)_P)] \cdot \mu'''(\bar{a})(\mathcal{P}_2)}{\iota(a) \cdot 2} = \\ \sum_{\mathcal{P}_1 | \mathcal{P}_2 \subseteq \mathcal{R}}^{a \in \mathbb{A}} \frac{\mu'(a)((\mathcal{P}_1)_Q) \cdot \mu'''(\bar{a})(\mathcal{P}_2)}{\iota(a) \cdot 2} + \\ \sum_{\mathcal{P}_1 | \mathcal{P}_2 \subseteq \mathcal{R}}^{a \in \mathbb{A}} \frac{\mu''(a)((\mathcal{P}_1)_P) \cdot \mu'''(\bar{a})(\mathcal{P}_2)}{\iota(a) \cdot 2}. \end{aligned}$$

Observe that, due to the way the sum is defined (and because pairing is an involution) we have that

$$\begin{aligned} \sum_{\mathcal{P}_1 | \mathcal{P}_2 \subseteq \mathcal{R}}^{a \in \mathbb{A}} \frac{\mu'(a)((\mathcal{P}_1)_Q) \cdot \mu'''(\bar{a})(\mathcal{P}_2)}{\iota(a) \cdot 2} = \\ \sum_{\mathcal{P}_1 | \mathcal{P}_2 \subseteq \mathcal{R}_Q}^{a \in \mathbb{A}} \frac{\mu'(a)(\mathcal{P}_1) \cdot \mu'''(\bar{a})(\mathcal{P}_2)}{\iota(a) \cdot 2} \end{aligned}$$

and

$$\begin{aligned} \sum_{\mathcal{P}_1 | \mathcal{P}_2 \subseteq \mathcal{R}}^{a \in \mathbb{A}} \frac{\mu''(a)((\mathcal{P}_1)_P) \cdot \mu'''(\bar{a})(\mathcal{P}_2)}{\iota(a) \cdot 2} = \\ \sum_{\mathcal{P}_1 | \mathcal{P}_2 \subseteq \mathcal{R}_P}^{a \in \mathbb{A}} \frac{\mu''(a)(\mathcal{P}_1) \cdot \mu'''(\bar{a})(\mathcal{P}_2)}{\iota(a) \cdot 2}. \end{aligned}$$

Consequently,

$$\begin{aligned} ((\mu'_{P \otimes_Q} \mu''_{P|Q \otimes_R} \mu''''_{P|R}) (\tau_r)) (\mathcal{R}) = \\ = \mu'(\tau)(\mathcal{R}_{R|Q}) + \mu''(\tau)(\mathcal{R}_{R|P}) + \mu'''(\tau)(\mathcal{R}_{P|R}) + \\ \sum_{\mathcal{P}_1 | \mathcal{P}_2 \subseteq \mathcal{R}_R}^{a \in \mathbb{A}} \frac{\mu'(a)(\mathcal{P}_1) \cdot \mu''(\bar{a})(\mathcal{P}_2)}{\iota(a) \cdot 2} + \\ \sum_{\mathcal{P}_1 | \mathcal{P}_2 \subseteq \mathcal{R}_Q}^{a \in \mathbb{A}} \frac{\mu'(a)(\mathcal{P}_1) \cdot \mu'''(\bar{a})(\mathcal{P}_2)}{\iota(a) \cdot 2} + \\ \sum_{\mathcal{P}_1 | \mathcal{P}_2 \subseteq \mathcal{R}_P}^{a \in \mathbb{A}} \frac{\mu''(a)(\mathcal{P}_1) \cdot \mu'''(\bar{a})(\mathcal{P}_2)}{\iota(a) \cdot 2}. \end{aligned}$$

Similarly it can be proved that

$$\begin{aligned} (\mu'_{P \otimes_Q|R} (\mu''_{Q \otimes_R} \mu''''_{P|R})) (\tau) (\mathcal{R}) = \\ = \mu'(\tau)(\mathcal{R}_{R|Q}) + \mu''(\tau)(\mathcal{R}_{R|P}) + \mu'''(\tau)(\mathcal{R}_{P|R}) + \end{aligned}$$

$$\begin{array}{ll}
(\text{Null}) \quad \overline{0 \rightarrow \bar{\omega}} & (\text{Guard}) \quad \overline{\varepsilon.P \rightarrow [\varepsilon]} \\
(\text{Sum}) \quad \frac{P \rightarrow \mu' \quad Q \rightarrow \mu''}{P + Q \rightarrow \mu' \oplus \mu''} & (\text{Par}) \quad \frac{P \rightarrow \mu' \quad Q \rightarrow \mu''}{P|Q \rightarrow \mu' \otimes_Q \mu''}
\end{array}$$

Table 4.1: Structural Operational Semantics

$$\begin{aligned}
& \sum_{\mathcal{P}_1 | \mathcal{P}_2 \subseteq \mathcal{R}_R} \frac{a \in \mathbb{A} \quad \mu'(a)(\mathcal{P}_1) \cdot \mu''(\bar{a})(\mathcal{P}_2)}{\iota(a) \cdot 2} + \\
& \sum_{\mathcal{P}_1 | \mathcal{P}_2 \subseteq \mathcal{R}_Q} \frac{a \in \mathbb{A} \quad \mu'(a)(\mathcal{P}_1) \cdot \mu'''(\bar{a})(\mathcal{P}_2)}{\iota(a) \cdot 2} + \\
& \sum_{\mathcal{P}_1 | \mathcal{P}_2 \subseteq \mathcal{R}_P} \frac{a \in \mathbb{A} \quad \mu''(a)(\mathcal{P}_1) \cdot \mu'''(\bar{a})(\mathcal{P}_2)}{\iota(a) \cdot 2}.
\end{aligned}$$

3(c). We prove now that  $\mu' \otimes_P \bar{\omega} = \mu'$ . Consider some arbitrary  $a \in \mathbb{A}$  and  $\mathcal{R} \in \Pi$ .

$$(\mu' \otimes_P \bar{\omega})(a)(\mathcal{R}) = \mu'(a)(\mathcal{R}_0) + \bar{\omega}(a)(\mathcal{R}_P).$$

But  $\mathcal{R}_0 = \mathcal{R}$  and  $\bar{\omega}(a)(\mathcal{R}_P) = 0$ . Consequently,

$$(\mu' \otimes_P \bar{\omega})(a)(\mathcal{R}) = \mu'(a)(\mathcal{R}).$$

We also have

$$\begin{aligned}
& (\mu' \otimes_P \bar{\omega})(\tau)(\mathcal{R}) = \mu'(\tau)(\mathcal{R}_0) + \bar{\omega}(\tau)(\mathcal{R}_P) + \\
& \sum_{\mathcal{P}_1 | \mathcal{P}_2 \subseteq \mathcal{R}} \frac{a \in \mathbb{A} \quad \mu'(a)(\mathcal{P}_1) \cdot \bar{\omega}(\bar{a})(\mathcal{P}_2)}{\iota(a) \cdot 2}.
\end{aligned}$$

But  $\bar{\omega}(\bar{a})(\mathcal{P}_1) = \bar{\omega}(\tau)(\mathcal{R}_P) = 0$  and  $\mathcal{R}_0 = \mathcal{R}$ , where from we obtain

$$(\mu' \otimes_P \bar{\omega})(\tau)(\mathcal{R}) = \mu'(\tau)(\mathcal{R}).$$

□

Having these operations available and thanks to their properties, we can introduce the operational semantics for our stochastic-CCS.

The rules of the structural operational semantics, given for arbitrary  $P, Q \in \mathbb{P}$  and  $\varepsilon \in \mathbb{A}^+$ , are listed in Table 4.1.

The *stochastic transition relation* is the smallest relation  $\rightarrow \subseteq \mathbb{P} \times \Delta(\mathbb{P}, \Pi)^{\mathbb{A}^+}$  satisfying the rules in Table 4.1.

The operational semantics associates with each process  $P \in \mathbb{P}$  a mapping  $\mu \in \Delta(\mathbb{P}, \Pi)^{\mathbb{A}^+}$ . For each  $\equiv$ -closed set of processes  $\mathcal{P} \in \Pi$  and each  $\alpha \in \mathbb{A}^+$ ,  $\mu(\alpha)(\mathcal{P}) = r \in \mathbb{R}^+$  represents the total rate of the  $\alpha$ -reductions of  $P$  to some arbitrary element of  $\mathcal{P}$ ; for simplicity we write  $P \xrightarrow{\alpha, r} \mathcal{P}$ .

The next lemma guarantees the consistency of the relation  $\rightarrow$  and of our operational semantics. It can be trivially proved by induction on the structure of a process term.

**Lemma 4.4.3.** *For any  $P \in \mathbb{P}$  there exists a unique  $\mu \in \Delta(\mathbb{P}, \Pi)^{\mathbb{A}^+}$  such that  $P \rightarrow \mu$ ; moreover,  $\mu$  has finite support.*

Our SOS can be further used to define pointwise semantics as, for instance, by

$$P \xrightarrow{\alpha, s} Q \text{ iff } \mu(\alpha)(Q^{\equiv}) = s.$$

The operational semantics proposed above does not differentiate between structural congruent processes. This can be proven inductively for each of the axioms of the structural congruence.

**Lemma 4.4.4.** *If  $P \equiv Q$  and  $P \rightarrow \mu$ , then  $Q \rightarrow \mu$ .*

*Proof.* The proof is done by induction on the structures of  $P$  and  $Q$  following the axioms of the structural congruence.

**The case  $P = R'|S, Q = R''|S$  with  $R' \equiv R''$ :**

Suppose that  $S \rightarrow \mu'$  and  $R' \rightarrow \mu''$  (from the inductive hypothesis,  $R'' \rightarrow \mu''$ ). Then,  $\mu = \mu'' \mathbin{R' \otimes_S} \mu'$ . Using 3(a) of Lemma 4.4.2, we obtain that  $\mu = \mu'' \mathbin{R'' \otimes_S} \mu'$  and, by (Par),  $Q \rightarrow \mu$ .

**The case  $P = R' + S, Q = R'' + S$  with  $R' \equiv R''$ :**

Suppose that  $S \rightarrow \mu'$  and  $R' \rightarrow \mu''$  (from the inductive hypothesis,  $R'' \rightarrow \mu''$ ). Then, by (Sum),  $Q \rightarrow \mu'' \oplus \mu'$ . However,  $\mu = \mu'' \oplus \mu'$ .

**The case  $P = \alpha.R, Q = \alpha.S$  with  $R \equiv S$ :**

We have  $\mu = D(\iota(\alpha), R)$  and  $S \rightarrow D(\iota(\alpha), S)$ . As  $R \equiv S$ , we obtain that  $\mu = D(\iota(\alpha), S)$ , i.e.,  $Q \rightarrow \mu$ .

**The case  $P = R|S, Q = S|R$ :**

Suppose that  $R \rightarrow \mu'$  and  $S \rightarrow \mu''$ . Then  $Q \rightarrow \mu'' \mathbin{S \otimes_R} \mu'$  and  $\mu = \mu' \mathbin{R \otimes_S} \mu''$ . At this point, to prove that  $Q \rightarrow \mu$ , it is sufficient to recall a result proven in Lemma 4.4.2,

$$\mu'' \mathbin{S \otimes_R} \mu' = \mu' \mathbin{R \otimes_S} \mu''.$$

**The case  $P = (R|S)|T, Q = R|(S|T)$ :**

Suppose that  $R \rightarrow \mu', S \rightarrow \mu''$  and  $T \rightarrow \mu'''$ . Then,

$$Q \rightarrow \mu' \mathbin{R \otimes_{S|T}} (\mu'' \mathbin{S \otimes_T} \mu''') \text{ and } \mu = (\mu' \mathbin{R \otimes_S} \mu'') \mathbin{R|S \otimes_T} \mu''.$$

However, Lemma 4.4.2 states the following equality, which completes our proof.

$$\mu' \mathbin{R \otimes_{S|T}} (\mu'' \mathbin{S \otimes_T} \mu''') = (\mu' \mathbin{R \otimes_S} \mu'') \mathbin{R|S \otimes_T} \mu''.$$

**The case  $Q = P|0$ :**

$Q \rightarrow \mu_{P \otimes_0 \bar{0}}$ , and from Lemma 4.4.2 we know that  $\mu_{P \otimes_0 \bar{0}} = \mu$ .

**The case  $P = R + S$  and  $Q = S + R$ :**

derives from the commutativity of  $\oplus$  prove in Lemma 4.4.2.

**The case  $P = (R + S) + T$  and  $Q = R + (S + T)$ :**

derives from the associativity of  $\oplus$  prove in Lemma 4.4.2.

**The case  $Q = P + 0$ :**

derives from the fact that  $0 \rightarrow \bar{0}$  and  $\bar{0}$  is a null with respect to the operation  $\oplus$ , as proven in Lemma 4.4.2.  $\square$

Notice from what we have developed so far that the signature of the algebra  $\mathbb{P}$  of processes does not correspond to the signature of the algebra  $\Delta(\mathbb{P}, \Pi)^{\mathbb{A}^+}$  of indexed distributions. This fact differentiates our approach from the other GSOS [51] or SGSOS [38] formats proposed in the literature.

For instance, to the parallel operator “|”, which is a binary operator in the algebra of processes, corresponds (in the domain of functions) a countable class of binary operators indexed by processes  $-_P \otimes_Q$ . This situation is a consequence of the fact that  $\equiv \subsetneq \approx$ .

To convince ourselves that these indexed composition operators might be different when they have different indexes, consider the processes  $P = a.0|b.0$  and  $Q = a.b.0 + b.a.0$  for  $a, b \in \mathbb{A}$  such that  $\{a, \bar{a}\} \cap \{b, \bar{b}\} = \emptyset$ . Then,

$$P \rightarrow (\mu_1 = \begin{bmatrix} a \\ 0 \end{bmatrix}_{a.0 \otimes_{b.0} \begin{bmatrix} b \\ 0 \end{bmatrix}}) \text{ and } Q \rightarrow (\mu_2 = \begin{bmatrix} a \\ b.0 \end{bmatrix} \oplus \begin{bmatrix} b \\ a.0 \end{bmatrix}).$$

One can verify that  $\mu_1 = \mu_2$ , however  $P \not\equiv Q$ . This shows that for some  $R \rightarrow \nu$ , we can have  $\mu_1 \otimes_R \nu \neq \mu_2 \otimes_R \nu$ .

Notice also that the “generalised” transition system induced by our SOS is image-finite. The importance of this property was motivated from the perspective of GSOS in [51] where it is observed that image-finite GSOS are in one-to-one correspondence with the *distributive laws* that ensure the cooperation between the algebraic and the coalgebraic structures of the class of processes. The next lemma shows that our system has a similar property.

We write  $P \Longrightarrow Q$  if there exists  $\alpha \in \mathbb{A}^+$  and  $r \neq 0$  such that  $P \xrightarrow{\alpha, r} Q^{\equiv}$  and let  $\Longrightarrow^*$  be the transitive closure of  $\Longrightarrow$ .

**Lemma 4.4.5.** *For an arbitrary process  $P \in \mathbb{P}$  the sets  $S_1, S_2$  and  $S_3$  defined below are finite*

$$S_1 = \{\alpha \in \mathbb{A}^+ \mid P \xrightarrow{\alpha, r} \mathbb{P}, r \neq 0\},$$

$$S_2 = \{Q^{\equiv} \in \Pi \mid P \Longrightarrow Q\},$$

$$S_3 = \{Q^{\equiv} \in \Pi \mid P \Longrightarrow^* Q\}.$$

*Proof.* We prove it by induction on the syntactic structure of  $P$ . The nontrivial cases are:

**The case  $P = P_1 + P_2$ :** suppose that  $P \rightarrow \mu$  and  $P_i \rightarrow \mu_i$  for  $i = 1, 2$ . Then, for arbitrary  $a \in \mathbb{A}$  and  $Q \in \mathbb{P}$ ,

$$\mu(a)(Q^\equiv) = \mu_1(a)(Q^\equiv) + \mu_2(a)(Q^\equiv).$$

As there are a finite number of  $a \in \mathbb{A}$  and a finite number of  $Q \in \mathbb{P}$  such that  $\mu_1(a)(Q^\equiv) \neq 0$  and a finite number of  $a \in \mathbb{A}$  and a finite number of  $Q \in \mathbb{P}$  such that  $\mu_2(a)(Q^\equiv) \neq 0$ , we deduce that there exist a finite number of  $a \in \mathbb{A}$  and a finite number of  $Q \in \mathbb{P}$  such that  $\mu(a)(Q^\equiv) \neq 0$ .

Similarly, for arbitrary  $r \in \mathbb{Q}_+$ ,

$$\mu(\tau_r)(Q^\equiv) = \sum_{r=s+t} \mu_1(\tau_s)(Q^\equiv) + \mu_2(\tau_t)(Q^\equiv).$$

There exist a finite number of  $s \in \mathbb{Q}_+$  and a finite number of  $Q \in \mathbb{P}$  such that  $\mu_1(\tau_s)(Q^\equiv) \neq 0$  and there exist a finite number of  $t \in \mathbb{Q}_+$  and a finite number of  $Q \in \mathbb{P}$  such that  $\mu_2(\tau_t)(Q^\equiv) \neq 0$ . These imply that there exist a finite number of  $r \in \mathbb{Q}_+$  and a finite number of  $Q \in \mathbb{P}$  such that  $\mu(\tau_r)(Q^\equiv) \neq 0$ .

**The case  $P = P_1 | P_2$ :** suppose that  $P \rightarrow \mu$  and  $P_i \rightarrow \mu_i$  for  $i = 1, 2$ . Then, for arbitrary  $a \in \mathbb{A}$  and  $Q \in \mathbb{P}$ ,  $\mu(a)(Q^\equiv) = \mu_1(a)(Q_{P_2}^\equiv) + \mu_2(a)(Q_{P_1}^\equiv)$ . As there are a finite number of  $a \in \mathbb{A}$  and a finite number of  $R \in \mathbb{P}$  such that  $\mu_1(a)(R^\equiv) \neq 0$  and a finite number of  $a \in \mathbb{A}$  and a finite number of  $R \in \mathbb{P}$  such that  $\mu_2(a)(R^\equiv) \neq 0$ , we deduce that there exist a finite number of  $a \in \mathbb{A}$  and a finite number of  $Q \in \mathbb{P}$  such that  $\mu(a)(Q^\equiv) \neq 0$ . Similarly, for arbitrary  $r \in \mathbb{Q}_+$ ,

$$\mu(\tau_r)(Q^\equiv) = \mu_1(\tau_r)(Q_{P_2}^\equiv) + \mu_2(\tau_r)(Q_{P_1}^\equiv) +$$

There exist a finite number of  $s \in \mathbb{Q}_+$  and a finite number of  $R \in \mathcal{P}$  such that  $\mu_1(\tau_s)(R^\equiv) \neq 0$ , there exist a finite number of  $a \in \mathbb{A}_r$  and a finite number of  $\mathcal{R} \in \Pi$  such that  $\mu_1(a)(\mathcal{R}) \neq 0$ , there exist a finite number of  $s \in \mathbb{Q}_+$  and a finite number of  $R \in \mathbb{P}$  such that  $\mu_2(\tau_s)(R^\equiv) \neq 0$  and there exist a finite number of  $\bar{a} \in \mathbb{A}_r$  and a finite number of  $\mathcal{R} \in \Pi$  such that  $\mu_2(\bar{a})(\mathcal{R}) \neq 0$ . These imply that there exist a finite number of  $r \in \mathbb{Q}_+$  and a finite number of  $Q \in \mathbb{P}$  such that  $\mu(\tau_r)(Q^\equiv) \neq 0$ .  $\square$

## 4.5 Stochastic bisimulation is a congruence

This section is dedicated to the study of stochastic bisimulation for the minimal stochastic process algebra. In the pointwise approach, since the operational semantics requires various mathematical artifacts such as the multi-transition systems [35,36] or the proved SOS [47], the problem of stochastic bisimulation is difficult to trace. Recently, an elegant solution was proposed in [38] for the case when there are no equational restrictions on the algebraic level. As argued before, our algebra is endowed with an equational

theory of structural congruence that organizes the measurable space of processes and consequently, stochastic bisimulation requires a different treatment.

We introduce the stochastic bisimulation for the minimal process algebra as the stochastic bisimulation on the Markov kernel  $(\mathbb{P}, \Pi, \theta)$ . We show that it behaves well both on coalgebraic and on algebraic levels: processes that have associated the same functions by our SOS are bisimilar and the bisimulation is a congruence that extends the structural congruence.

Lemma 4.4.3 shows that the operational semantics induces a function  $\vartheta: \mathbb{P} \rightarrow \Delta(\mathbb{P}, \Pi)^{\mathbb{A}^+}$  defined by

$$\vartheta(P) = \mu \quad \text{iff} \quad P \rightarrow \mu.$$

There exists an obvious relation between  $\vartheta$  and the function  $\theta$  that organises  $\mathbb{P}$  as a Markov kernel. It reflects the similarity between Definitions 4.3.3 and 5.2.7.

**Lemma 4.5.1.** *If  $(\mathbb{P}, \Pi, \theta)$  is the Markov kernel of processes and  $\vartheta: \mathbb{P} \rightarrow \Delta(\mathbb{P}, \Pi)^{\mathbb{A}^+}$  is the function induced by SOS, then for arbitrary  $P \in \mathbb{P}$ ,  $\alpha \in \mathbb{A}^+$  and  $\mathcal{P} \in \Pi$ ,*

$$\theta(\alpha)(P)(\mathcal{P}) = \vartheta(P)(\alpha)(\mathcal{P}).$$

Recall that for a Markov kernel  $(M, \Sigma, \theta)$ ,  $\sim_{(M, \Sigma, \theta)}$  denotes the stochastic bisimulation on it. The next result is a direct consequence of the previous lemma stating that  $\sim_{(\mathbb{P}, \Pi, \theta)}$  is an extension of the inseparability relation induced by  $\vartheta$ .

**Corollary 4.5.2.** *For arbitrary  $P, Q \in \mathbb{P}$ , if  $P \rightarrow \mu$  and  $Q \rightarrow \mu$ , then  $P \sim_{(\mathbb{P}, \Pi, \theta)} Q$ .*

This result guarantees that we can safely define the concepts of *stochastic bisimulation/bisimilarity* for our process algebra as the stochastic bisimulation/bisimilarity on the Markov kernel  $(\mathbb{P}, \Pi, \theta)$ .

The next theorem provides a characterization of stochastic bisimilarity.

**Theorem 4.5.3.** *The stochastic bisimilarity  $\sim$  is the largest equivalence relation on  $\mathbb{P}$  such that for arbitrary  $P, Q \in \mathbb{P}$  with  $P \rightarrow \mu$  and  $Q \rightarrow \mu'$ ,*

$$P \sim Q \quad \text{iff} \quad [\text{for any } C \in \Pi(\sim) \text{ and any } \alpha \in \mathbb{A}^+, \mu(\alpha)(C) = \mu'(\alpha)(C)].$$

*Proof.* Before proceeding with the proof observe that for arbitrary equivalence relations  $\mathcal{R}_1, \mathcal{R}_2$  on a set  $M$ , there exists an equivalence relation  $\mathcal{R}$  on  $M$  such that  $\mathcal{R}_1 \cup \mathcal{R}_2 \subseteq \mathcal{R}$ . Moreover, each  $\mathcal{R}$ -equivalence class is the union of  $\mathcal{R}_1$ -equivalence classes as well as the union of  $\mathcal{R}_2$ -equivalence classes. The same result holds for a countable set of equivalence relations.

We prove that  $\sim$  is an equivalence relation. Reflexivity and symmetry are trivial.

Transitivity: suppose that  $P \sim Q$  and  $Q \sim R$ ,  $P \rightarrow \mu$ ,  $Q \rightarrow \mu'$  and  $R \rightarrow \mu''$ . Then, there exist two stochastic bisimulation relations  $\mathcal{R}_1, \mathcal{R}_2$  such that  $(P, Q) \in \mathcal{R}_1$

and  $(Q, R) \in \mathcal{R}_2$ . Let  $\mathcal{R}$  be the smallest equivalence relation such that  $\mathcal{R}_1 \cup \mathcal{R}_2 \subseteq \mathcal{R}$ . Consider arbitrary  $\alpha \in \mathbb{A}^+$  and  $C \in \Pi(\mathcal{R})$ . Observe that

$$\Pi(\mathcal{R}) = \Pi \cap \mathbb{P}^{\mathcal{R}},$$

where  $\mathbb{P}^{\mathcal{R}}$  is the set of  $\mathcal{R}$ -equivalence classes. Hence,  $C \in \mathbb{P}^{\mathcal{R}}$  and there exist  $(C_1^i)_{i \in I} \subseteq \mathbb{P}^{\mathcal{R}_1}$  and  $(C_2^j)_{j \in J} \subseteq \mathbb{P}^{\mathcal{R}_2}$  such that

$$C = \bigcup_{i \in I} C_1^i = \bigcup_{j \in J} C_2^j.$$

Assume that the elements of  $(C_k^i)_{i \in I}$  are pairwise distinct hence, pairwise disjoint for  $k = 1, 2$ .

Since  $(P, Q) \in \mathcal{R}_1$ , for each  $C_i \in \Pi(\mathcal{R}_1) = \Pi \cap \mathbb{P}^{\mathcal{R}_1}$  and  $\alpha \in \mathbb{A}^+$ ,

$$\mu(\alpha)(C_i) = \mu'(\alpha)(C_i).$$

Since  $(Q, R) \in \mathcal{R}_2$ , for each  $C_j \in \Pi(\mathcal{R}_2) = \Pi \cap \mathbb{P}^{\mathcal{R}_2}$  and  $\alpha \in \mathbb{A}^+$ ,

$$\mu'(\alpha)(C_j) = \mu''(\alpha)(C_j).$$

We show that for each  $C \in \Pi(\mathcal{R})$  and each  $\alpha \in \mathbb{A}^+$ ,  $\mu(\alpha)(C) = \mu''(\alpha)(C)$ .

Because  $\mu(\alpha)$ ,  $\mu'(\alpha)$  and  $\mu''(\alpha)$  are measures, we obtain

$$\mu(\alpha)(C) = \sum_{i \in I} \mu(\alpha)(C_i) = \sum_{i \in I} \mu'(\alpha)(C_i) = \mu'(\alpha)(C).$$

Similarly,

$$\mu'(\alpha)(C) = \sum_{j \in J} \mu'(\alpha)(C_j) = \sum_{j \in J} \mu''(\alpha)(C_j) = \mu''(\alpha)(C).$$

Hence,  $\mu(\alpha)(C) = \mu''(\alpha)(C)$  proving that  $\mathcal{R}$  is a stochastic bisimulation and concluding the transitivity proof.

For showing that

$$P \sim Q \text{ iff [for any } C \in \Pi(\sim) \text{ and } \alpha \in \mathbb{A}, \mu(\alpha)(C) = \mu'(\alpha)(C)],$$

we proceed as before, observing that  $P \sim Q$  implies the existence of a bisimulation relation  $\mathcal{R}$  such that  $(P, Q) \in \mathcal{R}$ . Each  $C \in \Pi(\sim)$  is a countable union of pairwise disjoint measurable  $\mathcal{R}$ -equivalence classes and since  $\mu(\alpha)$ ,  $\mu'(\alpha)$  are measures, we can prove that  $\mu(\alpha)(C) = \mu'(\alpha)(C)$ .  $\square$

Denote by  $\mathbb{P}^{\sim}$  the set of  $\sim$ -equivalence classes on  $\mathbb{P}$ , and by  $P^{\sim}$  the  $\sim$ -equivalence class of  $P \in \mathbb{P}$ . If  $P$  and  $Q$  are not stochastic bisimilar, we write  $P \not\sim Q$ .

In what follows we show some bisimilar processes. The first example is a general rule for concurrent Markovian processes (see Section 4.1.2 of [37]).

**Example 4.5.4.** If  $a, b \in A$  such that  $\bar{a} \neq b$ , then for any  $P, Q \in \mathbb{P}$ ,

$$a.P|b.Q \sim a.(P|b.Q) + b.(a.P|Q).$$

Indeed,

$$\begin{aligned} a.P|b.Q &\rightarrow [^a_P]_{a.P \otimes b.Q} [^b_Q], \\ a.(P|b.Q) + b.(a.P|Q) &\rightarrow [^a_{P|b.Q}] \oplus [^b_{a.P|Q}] \end{aligned}$$

and for arbitrary  $C \in \mathbb{P}^\sim$ ,

$$[^a_P]_{a.P \otimes b.Q} [^b_Q](x)(C) = [^a_{P|b.Q}] \oplus [^b_{a.P|Q}](x)(C) = \begin{cases} \iota(a) & \text{if } x = a \text{ and } P|b.Q \in C, \\ \iota(b) & \text{if } x = b \text{ and } a.P|Q \in C, \\ 0 & \text{otherwise.} \end{cases}$$

**Example 4.5.5.** Let  $b, c \in A$  be such that  $\bar{b} \neq c$ . In Example 4.5.4 we have seen that

$$b.0|c.0 \sim b.c.0 + c.b.0.$$

Consider now the processes

$$\begin{aligned} P &= \tau_r.(b.0|c.0) + \tau_r.(b.c.0 + c.b.0), \\ Q &= \tau_r.(b.0|c.0) + \tau_r.(b.0|c.0) \text{ and} \\ R &= \tau_r.(b.c.0 + c.b.0) + \tau_r.(b.c.0 + c.b.0). \end{aligned}$$

If  $C$  is the  $\sim$ -equivalence class that contains  $b.0|c.0$  and  $b.c.0 + c.b.0$ , then

$$P \xrightarrow{\tau, 2r} C, \quad Q \xrightarrow{\tau, 2r} C, \quad R \xrightarrow{\tau, 2r} C$$

and for any other  $\sim$ -equivalence class  $C'$ ,

$$P \xrightarrow{\tau, 0} C', \quad Q \xrightarrow{\tau, 0} C' \text{ and } R \xrightarrow{\tau, 0} C'.$$

Consequently,  $P \sim Q \sim R$ .

On the other hand, if we consider the pointwise semantics, we obtain

$$\begin{array}{lll} P \xrightarrow{\tau, r} b.0|c.0 & Q \xrightarrow{\tau, 2r} b.0|c.0 & R \xrightarrow{\tau, 0} b.0|c.0 \\ P \xrightarrow{\tau, r} b.c.0 + c.b.0 & Q \xrightarrow{\tau, 0} b.c.0 + c.b.0 & R \xrightarrow{\tau, 2r} b.c.0 + c.b.0. \end{array}$$

Notice that they are not agreeing on any ‘‘pointwise’’ transition. This emphasizes the difficulties with pointwise semantics.

The relation  $\sim$  on  $\mathbb{P}$  can be lifted to  $\Delta(\mathbb{P})^{\mathbb{A}^+}$  by defining, for arbitrary  $\mu, \mu' \in \Delta(\mathbb{P})^{\mathbb{A}^+}$ ,  
 $\mu \sim \mu'$  iff for any  $C \in \mathbb{P}^{\sim}$  and any  $\alpha \in \mathbb{A}^+$ ,  $\mu(\alpha)(C) = \mu'(\alpha)(C)$ .

Notice that  $\sim \subseteq \Delta(\mathbb{P})^{\mathbb{A}^+} \times \Delta(\mathbb{P})^{\mathbb{A}^+}$  is an equivalence relation. We denote by  $(\Delta(\mathbb{P})^{\mathbb{A}^+})^{\sim}$  the set of  $\sim$ -equivalence classes on  $\Delta(\mathbb{P})^{\mathbb{A}^+}$  and for an arbitrary  $\mu \in \Delta(\mathbb{P})^{\mathbb{A}^+}$  we denote by  $\mu^{\sim}$  the  $\sim$ -equivalence class of  $\mu$ .

With this notation, from Theorem 4.5.3 we derive the next corollary.

**Corollary 4.5.6.** *Given  $P, Q \in \mathbb{P}$ , if  $P \rightarrow \mu$  and  $Q \rightarrow \mu'$ , then  $P \sim Q$  iff  $\mu \sim \mu'$ .*

A consequence of  $\sim$  being an equivalence on  $\Delta(\mathbb{P})^{\mathbb{A}^+}$  is the next theorem that shows that our processes behave “correctly” with respect to structural congruence.

**Theorem 4.5.7.** *Given  $P, Q \in \mathbb{P}$ , if  $P \equiv Q$ , then  $P \sim Q$ .*

*Proof.* Suppose that  $P \rightarrow \mu$ .  $P \equiv Q$  implies (Lemma 4.4.4) that  $Q \rightarrow \mu$ . As for any  $\sim$ -equivalence class  $C$  and any  $\alpha \in \mathbb{A}^+$ ,  $\mu(\alpha)(C) = \mu(\alpha)(C)$ , we obtain  $P \sim Q$ .  $\square$

In addition, notice that  $\sim$  is strictly larger than  $\equiv$ , because for arbitrary  $a, b \in \mathbb{A}$  with  $\bar{a} \neq b$ , we have

$$a.0|b.0 \sim a.b.0 + b.a.0 \text{ and } a.0|b.0 \not\equiv a.b.0 + b.a.0.$$

We now state the main theorem of this section.

**Theorem 4.5.8 (Congruence).** *Stochastic bisimulation on  $\mathbb{P}$  is a congruence, i.e. for arbitrary  $P, P', Q, Q' \in \mathbb{P}$  and  $\varepsilon \in \mathbb{A}^*$ , if  $P \sim P'$  and  $Q \sim Q'$ , then*

1.  $\varepsilon.P \sim \varepsilon.P'$ ;
2.  $P + Q \sim P' + Q'$ ;
3.  $P|Q \sim P'|Q'$ .

*Proof.* The only non-trivial case is

$$[P \sim P' \text{ and } Q \sim Q' \text{ implies } P|Q \sim P'|Q'].$$

To prove this it is sufficient to prove that

$$[P \sim Q \text{ implies } P|R \sim Q|R].$$

We prove this last implication inductively on the *complexity* of the processes involved defined by:

$$cx(0) = 0, \quad cx(\alpha.P) = cx(P) + 1 \text{ and } cx(P|Q) = cx(P + Q) = cx(P) + cx(Q).$$

Observe that the complexity of a process is strictly related to the behavior of the process. Indeed, if for some  $r \neq 0$ ,  $P \xrightarrow{\alpha, r} Q$ , then  $cx(P) > cx(Q)$ .

For  $(x_1, x_2), (y_1, y_2) \in \mathbb{N}^2$  we write  $(x_1, x_2) < (y_1, y_2)$  iff for each  $i = 1, 2$ ,  $x_i \leq y_i$  and for some  $j = 1, 2$ ,  $x_j < y_j$ . With this notation, we will prove the result inductively on  $(\max(cx(P), cx(Q)), cx(R))$ .

The base case is trivial, so we prove, in what follows, the inductive step. Suppose that for any  $P', Q', R' \in \mathbb{P}$  with

$$(\max(cx(P'), cx(Q')), cx(R')) < (\max(cx(P), cx(Q)), cx(R))$$

we have that  $[P' \sim Q' \text{ implies } P'|R' \sim Q'|R']$ . And we prove that

$$[P \sim Q \text{ implies } P|R \sim Q|R].$$

Suppose that  $P \rightarrow \mu$ ,  $Q \rightarrow \eta$  and  $R \rightarrow \rho$ . Then,  $P|R \rightarrow \mu \otimes_R \rho$  and  $Q|R \rightarrow \eta \otimes_R \rho$ . For showing  $P|R \sim Q|R$ , it is sufficient to show that for arbitrary  $\alpha \in \mathbb{A}^+$  and  $C \in \mathbb{P}^\sim$ ,

$$(\mu \otimes_R \rho)(\alpha)(C) = (\eta \otimes_R \rho)(\alpha)(C).$$

**The case  $\alpha = a \in \mathbb{A}$ :** Due to Lemma 4.4.5 we can assume that:

- there exists a finite set of processes  $\mathcal{P} = \{P_1^1, \dots, P_1^{n_1}, \dots, P_k^1, \dots, P_k^{n_k}\}$ , pairwise non-structural congruent, such that  $P \xrightarrow{a,0} \mathbb{P} \setminus \mathcal{P}$  and  $P \xrightarrow{a,p_i^j} P_i^j$  for some  $p_i^j \neq 0$ ; in addition, for each  $i = 1..k$  and each  $j, j' \in \{1, ..n_i\}$ ,  $P_i^j \sim P_i^{j'}$  and for  $i \neq i'$ ,  $x = 1..n_i$ ,  $x' = 1..n_{i'}$ ,  $P_i^x \not\sim P_{i'}^{x'}$ ; let  $p_i = \sum_{j=1..n_i} p_i^j$ ;
- there exists a finite set of processes  $\mathcal{Q} = \{Q_1^1, \dots, Q_1^{m_1}, \dots, Q_l^1, \dots, Q_l^{m_l}\}$ , pairwise non-structural congruent, such that  $Q \xrightarrow{a,0} \mathbb{P} \setminus \mathcal{Q}$  and  $Q \xrightarrow{a,q_i^j} Q_i^j$  for some  $q_i^j \neq 0$ ; in addition, for each  $i = 1..l$  and each  $j, j' \in \{1, ..m_i\}$ ,  $Q_i^j \sim Q_i^{j'}$  and for  $i \neq i'$ ,  $x = 1..m_i$ ,  $x' = 1..m_{i'}$ ,  $Q_i^x \not\sim Q_{i'}^{x'}$ ; let  $q_i = \sum_{j=1..m_i} q_i^j$ ;
- there exists a finite set of processes  $\mathcal{R} = \{R_1^1, \dots, R_1^{u_1}, \dots, R_v^1, \dots, R_v^{u_v}\}$ , pairwise non-structural congruent, such that  $R \xrightarrow{a,0} \mathbb{P} \setminus \mathcal{R}$  and  $R \xrightarrow{a,r_i^j} R_i^j$  for some  $r_i^j \neq 0$ ; in addition, for each  $i = 1..v$  and each  $j, j' \in \{1, ..u_i\}$ ,  $R_i^j \sim R_i^{j'}$  and for  $i \neq i'$ ,  $x = 1..u_i$ ,  $x' = 1..u_{i'}$ ,  $R_i^x \not\sim R_{i'}^{x'}$ ; let  $r_i = \sum_{j=1..u_i} r_i^j$ ;

$P \sim Q$  implies  $k = l$ ; suppose that  $P_i^j \sim Q_i^{j'}$  and for each  $i = 1..k$ ,  $p_i = q_i$ . For arbitrary  $C \in \mathbb{P}^\sim$ ,

$$(\mu \otimes_R \rho)(a)(C) = \mu(a)(C_R) + \rho(a)(C_P) = \sum_{(P_1|R)^\equiv \subseteq C} \mu(a)(P_1^\equiv) + \sum_{(R_1|P)^\equiv \subseteq C} \rho(a)(R_1^\equiv)$$

and

$$(\eta \otimes_R \rho)(a)(C) = \eta(a)(C_R) + \rho(a)(C_Q) = \sum_{(Q_1|R)^\equiv \subseteq C} \eta(a)(Q_1^\equiv) + \sum_{(R_1|Q)^\equiv \subseteq C} \rho(a)(R_1^\equiv).$$

If there exist  $i_1, ..i_t$  such that for each  $i \in \{i_1, ..i_t\}$  and only for them there exist  $j \in \{1..n_i\}$  with  $P_i^j|R \in C$ , then, from the inductive hypothesis we have that for each  $j' =$

$1..n_i, P_i^j | R \in C$ . Moreover, if  $P' | R \in C$  such that  $P | R \xrightarrow{a,s} P' | R$  for  $s \neq 0$ , then there exist  $i, j$  such that  $P' \equiv P_i^j$ . Consequently,

$$\sum_{(P_1 | R) \equiv \subseteq C} \mu(a)(P_1^{\equiv}) = \sum_{s=1..t} p_s.$$

Since  $P_i^j \sim Q_i^j$ , the inductive hypothesis provides  $Q_i^j | R \in C$ . A similar argument gives

$$\sum_{(Q_1 | R) \equiv \subseteq C} \eta(a)(Q_1^{\equiv}) = \sum_{s=1..t} p_s.$$

On the other hand, if there exist no  $i$  and  $j$  such that  $P_i^j | R \in C$ , from  $P \sim Q$  we obtain that there is no  $i, j$  such that  $Q_i^j | R \in C$ ; hence,

$$\sum_{(Q_1 | R) \equiv \subseteq C} \eta(a)(Q_1^{\equiv}) = \sum_{(P_1 | R) \equiv \subseteq C} \mu(a)(P_1^{\equiv}) = 0.$$

Observe now that using the inductive hypothesis,  $P \sim Q$  implies  $P | R_i^j \sim Q | R_i^j$ , i.e.,  $[R_i^j | P \in C \text{ iff } Q | R_i^j \in C]$ . Hence, if  $R_i^j | P \in C$ , we obtain

$$\sum_{(R' | P) \equiv \subseteq C} \rho(a)(R') = \sum_{(R' | Q) \equiv \subseteq C} \rho(a)(R') = r_i$$

and otherwise,

$$\sum_{(R' | P) \equiv \subseteq C} \rho(a)(R') = \sum_{(R' | Q) \equiv \subseteq C} \rho(a)(R') = 0.$$

**The case  $\alpha = \tau$**  Due to Lemma 4.4.5 we can assume that:

- there exists a finite set of processes  $\mathcal{P} = \{P_0^1, \dots, P_0^{n_0}\}$ , pairwise non structural congruent, such that  $P \xrightarrow{\tau, p_0^j} P_0^j$  for some  $p_0^j \neq 0$  and  $P \xrightarrow{\tau, 0} \mathbb{P} \setminus \mathcal{P}$ ; in addition, there exists a finite set of actions  $a \in \mathbb{A}$  with  $P \xrightarrow{a,s} \mathbb{P}$  for some  $s \neq 0$  and for each such  $a$  there exists a set  $\{P_1^1, \dots, P_1^{n_1}, \dots, P_k^1, \dots, P_k^{n_k}\}$  of processes, pairwise non structural congruent, such that  $P \xrightarrow{a, p_i^j} P_i^j$  for some  $p_i^j \neq 0$ ; moreover, for each  $i = 0..k$  and  $j, j' \in \{1..n_i\}$ ,  $P_i^j \sim P_i^{j'}$  and for  $i \neq i'$ ,  $x = 1..n_i$ ,  $x' = 1..n_{i'}$ ,  $P_i^x \not\sim P_{i'}^{x'}$ ; let  $p_i = \sum_{j=1..n_i} p_i^j$  for each  $i = 0..k$ ;
- there exists a finite set of processes  $\mathcal{Q} = \{Q_0^1, \dots, Q_0^{m_0}\}$ , pairwise non structural congruent, such that  $Q \xrightarrow{\tau, q_0^j} Q_0^j$  for some  $q_0^j \neq 0$  and  $Q \xrightarrow{\tau, 0} \mathbb{P} \setminus \mathcal{Q}$ ; in addition, there exists a finite set of actions  $a \in \mathbb{A}$  with  $Q \xrightarrow{a,s} \mathbb{P}$  for some  $s \neq 0$  and for each such  $a$  there exists a set  $\{Q_1^1, \dots, Q_1^{m_1}, \dots, P_1^1, \dots, P_1^{m_1}\}$  of processes, pairwise non

structural congruent, such that  $Q \xrightarrow{a, q_i^j} Q_i^j$  for some  $q_i^j \neq 0$ ; moreover, for each  $i = 0..l$  and  $j, j' \in \{1, ..m_i\}$ ,  $Q_i^j \sim Q_i^{j'}$  and for  $i \neq i'$ ,  $x = 1..m_i$ ,  $x' = 1..m_{i'}$ ,  $Q_i^x \not\sim Q_{i'}^{x'}$ ; let  $q_i = \sum_{j=1..n_i} q_i^j$  for each  $i = 0, ..l$ ;

- there exists a finite set of processes  $\mathcal{R} = \{R_0^1, \dots, R_0^{u_0}\}$ , pairwise non structural congruent, such that  $R \xrightarrow{\tau, r_0^j} R_0^j$  for some  $r_0^j \neq 0$  and  $R \xrightarrow{\tau, 0} \mathbb{P} \setminus \mathcal{R}$ ; in addition, there exists a finite set of actions  $a \in \mathbb{A}$  with  $R \xrightarrow{a, s} \mathbb{P}$  for some  $s \neq 0$  and for each such  $a$  there exists a set  $\{R_1^1, \dots, R_1^{n_1}, \dots, R_k^1, \dots, R_k^{n_k}\}$  of processes, pairwise non structural congruent, such that  $R \xrightarrow{a, r_i^j} R_i^j$  for some  $r_i^j \neq 0$ ; moreover, for each  $i = 0..v$  and  $j, j' \in \{1, ..u_i\}$ ,  $R_i^j \sim R_i^{j'}$  and for  $i \neq i'$ ,  $x = 1..u_i$ ,  $x' = 1..u_{i'}$ ,  $R_i^x \not\sim R_{i'}^{x'}$ ; let  $r_i = \sum_{j=1..n_i} p_i^j$  for each  $i = 0, ..v$ ;

Observe that  $P \sim Q$  implies, for each  $a$  having the mentioned properties, that  $k = l$ ; we can suppose, without loosing generality, that  $P_i^j \sim Q_i^{j'}$  and for each  $i = 0..k$ ,  $p_i = q_i$ . For arbitrary  $C \in \mathbb{P}^\sim$ ,

$$\begin{aligned} (\mu_{P \otimes_R \rho})(\tau)(C) &= \mu(\tau)(C_R) + \rho(\tau)(C_P) + \\ &\sum_{(P_1|P_2) \equiv \subseteq C} \frac{\sum_{a \in \mathbb{A}} \mu(a)(P_1^{\equiv}) \cdot \rho(\bar{a})(P_2^{\equiv})}{\iota(a) \cdot 2} = \\ &\sum_{(P_1|R) \equiv \subseteq C} \mu(\tau)(P_1^{\equiv}) + \sum_{(R_1|P) \equiv \subseteq C} \rho(\tau)(R_1^{\equiv}) + \\ &\sum_{(P_1|P_2) \equiv \subseteq C} \frac{\sum_{a \in \mathbb{A}} \mu(a)(P_1^{\equiv}) \cdot \rho(\bar{a})(P_2^{\equiv})}{\iota(a) \cdot 2}, \end{aligned}$$

and

$$\begin{aligned} (\eta_{Q \otimes_R \rho})(a)(C) &= \eta(\tau)(C_R) + \rho(\tau)(C_Q) + \\ &\sum_{(Q_1|Q_2) \equiv \subseteq C} \frac{\sum_{a \in \mathbb{A}} \eta(a)(Q_1^{\equiv}) \cdot \rho(\bar{a})(Q_2^{\equiv})}{\iota(a) \cdot 2} = \\ &\sum_{(Q_1|R) \equiv \subseteq C} \eta(\tau)(Q_1^{\equiv}) + \sum_{(R_1|Q) \equiv \subseteq C} \rho(\tau)(R_1^{\equiv}) + \\ &\sum_{(Q_1|Q_2) \equiv \subseteq C} \frac{\sum_{a \in \mathbb{A}} \eta(a)(Q_1^{\equiv}) \cdot \rho(\bar{a})(Q_2^{\equiv})}{\iota(a) \cdot 2}. \end{aligned}$$

At this level we can demonstrate, using the same strategy as in the case  $\alpha = a$ , that

•

$$\sum_{(P_1|R) \equiv \subseteq C} \mu(\tau)(P_1^{\equiv}) = \sum_{(Q_1|R) \equiv \subseteq C} \eta(\tau)(Q_1 \equiv) = \sum_{i=1..t} p_i,$$

where  $i_1, \dots, i_t$  are such that for each  $i \in \{i_1, \dots, i_t\}$ , there exists some  $j$  such that (hence, for all  $j$ )  $P_i^j|R \in C$  and, from the inductive hypothesis, there exists  $j'$  such that (hence, for all  $j'$ )  $Q_i^{j'}|R \in C$ ;

• because  $P|R_i^j \sim Q|R_i^{j'}$ ,

$$\sum_{(R_1|P) \equiv \subseteq C} \rho(\tau)(R_1^{\equiv}) = \sum_{(R_1|Q) \equiv \subseteq C} \rho(\tau)(R_1^{\equiv}) = r_i,$$

where  $i$  is (the unique index) such that for some (hence, for all)  $j, j', P|R_i^j, Q|R_i^{j'} \in C$ .

• for each  $a$  as before we also have

$$\begin{aligned} \sum_{(P_1|P_2) \equiv \subseteq C} \frac{\eta(a)(P_1^{\equiv}) \cdot \rho(\bar{a})(P_2^{\equiv})}{\iota(a) \cdot 2} &= \sum_{(i,j) \in I} p_i \cdot q_j = \\ &= \sum_{(P_1|P_2) \equiv \subseteq C} \frac{\eta(a)(P_1^{\equiv}) \cdot \rho(\bar{a})(P_2^{\equiv})}{\iota(a) \cdot 2}, \end{aligned}$$

where  $I$  is the set of pairs of indexes  $(i, j)$  such that for some  $x, y$  (hence, for all),  $P_i^x|Q_j^y \in C$ .

We can proceed now with the proof of Theorem 4.5.8.

1. If  $P \sim P'$ , we have that for any  $C \in \mathbb{P}^{\sim}$ ,  $P \in C$  iff  $P' \in C$ . From here, we derive that  $[\varepsilon_p](\alpha)(C) = [\varepsilon_{p'}](\alpha)(C)$ .

2. If  $P \sim P'$  and  $Q \sim Q'$ , then  $P + Q \sim P' + Q'$ .

Suppose that  $P \rightarrow \mu$ ,  $P' \rightarrow \mu'$ ,  $Q \rightarrow \eta$  and  $Q' \rightarrow \eta'$ . Consider an arbitrary  $C \in \mathbb{P}^{\sim}$ .

For  $\alpha \in \mathbb{A}^+$ ,  $(\mu \oplus \eta)(\alpha)(C) = \mu(\alpha)(C) + \eta(\alpha)(C)$ .

But  $P \sim P'$  and  $Q \sim Q'$ , i.e. for any  $C \in \mathbb{P}^{\sim}$ ,

$$\mu(\alpha)(C) = \mu'(\alpha)(C) \text{ and } \eta(\alpha)(C) = \eta'(\alpha)(C).$$

Hence,

$$\mu(\alpha)(C) + \eta(\alpha)(C) = \mu'(\alpha)(C) + \eta'(\alpha)(C) = (\mu' \oplus \eta')(\alpha)(C).$$

3. If  $P \sim P'$  and  $Q \sim Q'$ , then  $P|Q \sim P'|Q'$ .

Since  $P \sim P'$  implies  $P|P' \sim Q|P'$  and  $Q \sim Q'$  implies  $Q|P' \sim Q'|P'$ , the transitivity of  $\sim$  proves  $P|Q \sim P'|Q'$ .  $\square$

Because the processes that agree on all instantiations of  $\vartheta$  are related by  $\sim$  and  $\sim$  is a congruence on the class of processes, we deduce that

if  $P \sim P', Q \sim Q', P \rightarrow \mu, P' \rightarrow \mu', Q \rightarrow \nu$  and  $Q' \rightarrow \nu'$ , then

$$\mu_{P \otimes Q} \nu \sim \mu'_{P' \otimes Q'} \nu'$$

and

for any  $\varepsilon \in \mathbb{A}^*$ ,  $[\varepsilon_P] \sim [\varepsilon_{P'}]$ .

These show that the quotients of  $\sim$  on processes and functions produce identical signatures for both domains and an SOS format in the style of [38,51].

## 4.6 Concluding remarks

In this chapter we developed a stochastic extension of CCS. We proposed a structural operational semantics based on measure theory and particularly suited for the extension that we will introduce in Chapter 5 where we consider a measure of similarity of behaviours for processes.

For organizing the set of processes as a measurable space, we have chosen the  $\sigma$ -algebra generated by the structural congruence classes of processes and we base the theory on top of it. This choice is motivated by practical modelling reasons: the calculus was meant to be used for applications in computational systems biology. In this context, the structural congruence and the distributions over the space of congruence classes play a key role. The congruence classes represent chemical “soups” and the various syntactic representations of the same soup need to be identified. In fact, structural congruence was inspired by a chemical analogy [7].

The stochastic behaviour is defined using a general concept of Markov process that encapsulates most of the Markovian models, including continuous ones, as well as other models of probabilistic systems, e.g., Harsanyi type spaces. This concept is based on unspecified analytic (hence, measurable) spaces and generalizes rate transition systems [23,38]. Consequently, we obtain a general definition of stochastic bisimulation similar to the one used in [28].

The novelty of this work consists in the fact that the measurable space of processes is axiomatized by structural congruence and the operational semantics reflects the interrelation between this space and the space of distributions on it. Our technology is appropriate for practical modelling purposes where various congruences can be relevant. It will help design (more complex) stochastic process algebras in a uniform way, possibly involving different equational axiomatizations, while avoiding the heavy techniques for counting of reductions.

# Chapter 5

## Stochastic Pi-Calculus

### 5.1 Introduction

In the previous chapter, dedicated to stochastic CCS, we have proposed an operational semantics that assigns to each process a set of measures indexed by observable actions. Thus, difficult instance-counting problems that otherwise require complicated versions of SOS can be solved by exploiting the properties of measures. We have shown that along this line one obtains an elegant semantics that resembles the one of nondeterministic PAs and provides a well-behaved notion of bisimulation. In this chapter we extend the work to stochastic Pi-calculus with channel-based communication, mobility, fresh name quantification and replication. Also this calculus, as the one introduced in the previous chapter, is designed to satisfy the specific requirements of Systems Biology.

There are a couple of novel ideas in this chapter that help us solving the specific issues related to name passing, new name operators and replication.

The processes in this chapter are interpreted in *stochastic environments* that associate *basic rates* to channels. In a rate environment  $E$ , a process  $P$  has associated a class of measures  $\mu$ , written  $E \vdash P \rightarrow \mu$ . For each action  $\alpha$ ,  $\mu(\alpha)$  is a measure over the space of processes;  $\mu(\alpha)(S) \in \mathbb{Q}^+$  is the rate of an exponentially distributed random variable that characterizes the  $\alpha$ -transitions from  $P$  to (elements of) a measurable set  $S$ .

As in the previous chapter, only the structural congruence-closed sets are measurable. This is essential for modelling in systems biology, where such sets represent chemical soups. This choice provides simple solutions to the problems of replications and bound outputs which otherwise, as with Milner's Abstraction-Concretion method [44], require complicated high-order reasoning.

Also novel is our *stochastic bisimulation* that extends other similar ones [17, 23, 38, 40, 46] by making explicit the role of the rate environments. Also in this case we show that bisimulation is a congruence that extends structural congruence.

The use of name environments has been considered in [30, 31] where it involves the machinery of nominal sets. We have tried to avoid this together with any coalgebraic description of the lifting from processes to measures, as our intention is to make these

ideas accessible also for the readers less familiar with this specific jargon.

**Relation to Nondeterministic Pi-Calculus.** There is no trivial relation between non-deterministic Pi-calculus and our stochastic Pi-calculus, in the sense that one cannot simply recover the semantics of the other by simple mathematical transformations. This is because the measure-based semantics of stochastic-Pi calculus require important modification of the SOS rules. One example regards the replication: while in classic Pi-calculus

$$!a(b) \equiv a(b)|!a(b),$$

in stochastic-Pi this is illegal since the rate of the input on channel  $a$  in the process  $a(b)|!a(b)$  is strictly bigger than the rate of the same input in the process  $!a(b)$ . For this reason in stochastic-Pi there exist no structural congruence rules of type

$$!P \equiv P|!P \text{ or } !!P \equiv !P$$

since such rules would generate processes with infinite rates; instead, there are dedicated SOS rules that establish the correct behaviours.

The chapter is organised as follows. Section 4.2 defines the general concept of Markov process (MP) and the stochastic bisimulation of MPs. Section 4.3 introduces the syntax of our process algebra and the axiomatization of structural congruence; we prove that the space of processes can be organised as a Markov kernel and that each process is an MP. These results guide us, in Section 5.2.4, to the definition of a structural operational semantics which induces a notion of behavioural equivalence that coincides with the bisimulation of MPs. In Section 4.5 we show that the bisimulation behaves well with respect to the algebraic structure of processes: stochastic bisimulation is a congruence.

## 5.2 Stochastic Pi-Calculus

In this section we introduce a version of stochastic Pi-calculus equipped with an *early semantics* [8] expressed in terms of measure theory. Being developed mainly for applications in Systems Biology, this calculus is designed to respect the *chemical kinetics* (the *Chemical Master Equation*) [13] which provides the mathematical principles for calculating the rates of the channel-based communications.

The class  $\mathbb{P}$  of processes is endowed with structural congruence which generates a  $\sigma$ -algebra  $\Pi$  on  $\mathbb{P}$ . In addition, rate environments assign base rates to channel names. The behaviour of a process  $P$  in a rate environment  $E$  is defined by an indexed set of measures  $\mu : \mathbb{A}^+ \rightarrow \Delta(\mathbb{P}, \Pi)$ , where  $\mathbb{A}^+$  is the set of actions.

The SOS is a proof system with statements of type  $E \vdash P \rightarrow \mu$ . For an arbitrary  $\alpha \in \mathbb{A}^+$ ,  $\mu(\alpha)$  is a measure on  $(\mathbb{P}, \Pi)$  and for a measurable set of processes  $S \in \Pi$ ,  $\mu(\alpha)(S) \in \mathbb{R}^+$  represents the rate of an exponentially distributed random variable that characterizes the duration of the  $\alpha$ -transitions from  $P$  to (some element of)  $S$ .

### 5.2.1 Syntax

**Definition 5.2.1** (Processes). Let  $\mathcal{N}$  be a countable set of channel names. The stochastic processes are defined, on top of 0, for arbitrary  $r \in \mathbb{Q}^+$  and  $a, b, c \in \mathcal{N}$ , as follows.

$$P := 0 \mid x.P \mid (a@r)P \mid P \mid P \mid P + P \mid !P,$$

$$x := a(b) \mid a[b].$$

Let  $\mathbb{P}$  be the set of stochastic processes.

0 stands for the inactive process.

An *input* “ $a(b)$ ” is the capability of the process  $a(b).P$  to receive a name on channel  $a$  that replaces  $b$  in all its occurrences inside  $P$ .

An *output* prefix “ $a[b]$ ” represents the action of sending a name  $b$  on channel  $a$ . “ $(a@r)$ ” is the *fresh name operator* that, unlike in nondeterministic PAs, also specifies the rate  $r$  of the fresh name<sup>1</sup>.

As usual in Pi-calculus, we have the *parallel composition* “ $\mid$ ”, the *choice operator* “ $+$ ” and the *replication operator* “ $!$ ”.

For arbitrary  $P \in \mathbb{P}$ , we define the set  $fn(P)$  of the *free names* of  $P$  inductively by

$$\begin{aligned} fn(0) &= \emptyset, \\ fn(a(b).P) &= (fn(P) \setminus \{b\}) \cup \{a\}, \\ fn(a[b].P) &= fn(P) \cup \{a, b\}, \\ fn(P \mid Q) &= fn(P + Q) = fn(P) \cup fn(Q), \\ fn((a@r)P) &= fn(P) \setminus \{a\} \text{ and} \\ fn(!P) &= fn(P). \end{aligned}$$

As usual in process algebras, for arbitrary  $a, b \in \mathcal{N}$ , we write  $P_{\{a/b\}}$  for the process term obtained from  $P$  by substituting all the free occurrences of  $b$  with  $a$ , renaming as necessary to avoid capture.

**Definition 5.2.2** (Structural congruence). *Structural congruence is the smallest equivalence relation  $\equiv \subseteq \mathbb{P} \times \mathbb{P}$  satisfying the following conditions.*

1.  $(\mathbb{P}, \mid, 0)$  is a commutative monoid for  $\equiv$ , i.e.,

- (a)  $P \mid Q \equiv Q \mid P$ ;
- (b)  $(P \mid Q) \mid R \equiv P \mid (Q \mid R)$ ;
- (c)  $P \mid 0 \equiv P$ .

2.  $(\mathbb{P}, +, 0)$  is a commutative monoid for  $\equiv$ , i.e.,

- (a)  $P + Q \equiv Q + P$ ;
- (b)  $(P + Q) + R \equiv P + (Q + R)$ ;

---

<sup>1</sup>Because in practice we cannot measure nor specify models with irrational rates, we have chosen  $r \in \mathbb{Q}^+$ . However, the technical development does not change if  $r \in \mathbb{R}^+$ .

$$(c) P + 0 \equiv P.$$

3.  $\equiv$  is a congruence for the algebraic structure of  $\mathbb{P}$ , i.e., if  $P \equiv Q$ , then

$$(a) P|R \equiv Q|R;$$

$$(b) P + R \equiv Q + R;$$

$$(c) !P \equiv !Q;$$

$$(d) a[b].P \equiv a[b].Q$$

$$(e) (a@r)P \equiv (a@r)Q;$$

$$(f) a(b).P \equiv a(b).Q.$$

4. the fresh name quantifiers satisfy the following conditions

$$(a) \text{ if } a \neq b, \text{ then } (a@r)(b@s)P \equiv (b@s)(a@r)P;$$

$$(b) (a@r)0 \equiv 0;$$

$$(c) \text{ if } a \notin \text{fn}(P), \text{ then } (a@r)(P|Q) \equiv P|(a@r)Q \text{ and } (a@r)(P + Q) \equiv P + (a@r)Q.$$

5. the replication satisfies the following conditions

$$(a) !0 \equiv 0;$$

$$(b) !(P|Q) \equiv !P|!Q.$$

6.  $\equiv$  satisfies the alpha-conversion rules

$$(a) (a@r)P \equiv (b@r)P_{\{b/a\}};$$

$$(b) a(b)P \equiv a(c)P_{\{c/b\}}.$$

If  $Q$  is obtained from  $P$  by alpha-conversion (6)a-b, we write  $P \equiv^* Q$ .

Let  $\Pi$  be the set of the  $\equiv$ -closed subsets of  $\mathbb{P}$ . Note that  $\mathbb{P}^\equiv$  is a countable partition of  $\mathbb{P}$  and  $\Pi$  is the  $\sigma$ -algebra generated by  $\mathbb{P}^\equiv$ .

Notice that, unlike in the nondeterministic case, we do not have

$$!!P \equiv !P \text{ nor } !P \equiv P|!P.$$

These are not sound due to the rate competition which else will generate processes with infinite rates.

**Theorem 5.2.3** (Measurable space).  $(\mathbb{P}, \Pi)$  is a measurable space of processes.

The measurable sets of  $\mathbb{P}$  are the unions of  $\equiv$ -equivalence classes on  $\mathbb{P}$ . In what follows  $\mathcal{P}, \mathcal{R}, \mathcal{Q}$  range over  $\Pi$ .

We lift some functions and algebraic operations from processes to measurable sets, for arbitrary  $a, b \in \mathcal{N}$  and  $r \in \mathbb{Q}^+$ , as follows.

$$\begin{aligned}
fn(\mathcal{P}) &= \bigcup_{P \in \mathcal{P}} fn(P), \\
\mathcal{P}_{\{a/b\}} &= \bigcup_{P \in \mathcal{P}} P_{\{a/b\}}, \\
\mathcal{P}|Q &= \bigcup_{P \in \mathcal{P}} \bigcup_{Q \in \mathcal{Q}} (P|Q)^{\equiv}, \\
\mathcal{P}_Q &= \bigcup_{R|Q \in \mathcal{P}} R^{\equiv}, \\
(a@r)\mathcal{P} &= \bigcup_{P \in \mathcal{P}} (a@r)P^{\equiv}.
\end{aligned}$$

It is not difficult to see that these operations are internal operations on  $\Pi$ .

**Lemma 5.2.4.** *If  $\mathcal{P}, Q \in \Pi$ ,  $a, b \in \mathcal{N}$ ,  $r \in \mathcal{Q}^+$ , then*

$$\mathcal{P}_{\{a/b\}}, \mathcal{P}|Q, \mathcal{P}_Q, (a@r)\mathcal{P} \in \Pi.$$

## 5.2.2 Rate environments

Now we introduce *rate environments* used to interpret stochastic processes. They are partial functions on  $\mathcal{N}$  assigning rates to channels. We chose to introduce them in the “process algebra style” instead of defining a type systems for environment correctness, which would complicate the semantics.

**Definition 5.2.5** (Rate Environment). *The rate environments associated to  $\mathcal{N}$  are defined, on top of a constant  $\varepsilon$ , for arbitrary  $a \in \mathcal{N}$  and  $r \in \mathcal{Q}^+$ , by*

$$E := \varepsilon : E, a@r.$$

A suffix  $a@r$  is called a *rate declaration*. If  $a@r$  appears in  $E$ , we write  $a@r \in E$ .  $\varepsilon$  stands for the *empty environment*. We treat “,” as concatenation symbol for rate environments and use “ $E, E'$ ” to denote the concatenation of  $E$  and  $E'$ ;  $\varepsilon$  is the empty symbol for concatenation.

Let  $\mathbb{E}$  be the set of rate environments.

For  $E = E_1, \dots, E_n \in \mathbb{E}$  and  $\{1, \dots, n\} = \{i_1, \dots, i_k\} \cup \{j_1, \dots, j_{n-k}\}$  with  $i_1 < \dots < i_k$ ,  $j_1 < \dots < j_{n-k}$ , if  $E' = E_{i_1}, \dots, E_{i_k}$  and  $E'' = E_{j_1}, \dots, E_{j_{n-k}}$ , we write  $E' \subset E$  and  $E'' = E \setminus E'$ . Notice that  $\varepsilon \subset E$ ,  $E \subset E, E = E \setminus \varepsilon$  and  $\varepsilon = E \setminus E$ .

The *domain of a rate environment* is the partial function on  $\mathbb{E}$  defined as follows.

1.  $dom(\varepsilon) = \emptyset$ ;
2. if  $dom(E)$  is defined and  $a \notin dom(E)$ , then  $dom(E, a@r) = dom(E) \cup \{a\}$ ;
3. undefined otherwise.

In what follows, whenever we use  $dom(E)$  we implicitly assume that  $dom$  is defined in  $E$ . Observe that, if  $a \in dom(E)$ , then there exists a rate declaration  $a@r \in E$  and for no  $s \neq r, a@s \in E$ ; for this reason we also write  $r = E(a)$ .

When  $dom(E)$  is defined, let  $dom(E)^* = \{a \in dom(E) \text{ s.t. } E(a) \neq 0\}$ .

### 5.2.3 The class of indexed measures

To provide the preliminary definitions needed to introduce the operational semantics, in this section we focus on a special class of mappings over the class of measures on  $(\mathbb{P}, \Pi)$ .

The semantics will involve terms of type  $E \vdash P \rightarrow \mu$ , where  $E$  is a rate environment,  $P$  is a process and  $\mu : \mathbb{A}^+ \rightarrow \Delta(\mathbb{P}, \Pi)$  is a mapping that defines a set of labeled measures.

The labels are the observable actions collected in the set  $\mathbb{A}^+$  defined below.

$$\mathbb{A} = \{a[b], a[@r], ab, \text{ for } a, b \in \mathcal{N}, r \in \mathbb{Q}^+\} \text{ and } \mathbb{A}^+ = \mathbb{A} \cup \{\tau\}.$$

We denote by  $\mathfrak{M}$  the set  $\Delta(\mathbb{P}, \Pi)^{\mathbb{A}^+}$  of labeled measures.

The observable actions consist of four classes:

1. *free outputs* of type  $a[b]$  denoting the action of sending a free name  $b$  over the channel  $a$ ,
2. *bound outputs* of type  $a[@r]$  denoting the action of sending a fresh unspecified name, with base-rate  $r$ , on channel  $a$ ,
3. *input actions* of type  $ab$  representing the fact that channel  $a$  has received a name  $b$  (as the result of an output action on  $a$ ),
4. *internal action*  $\tau$  – communications.

In what follows we use  $\alpha, \alpha_i$  to represent arbitrary elements of  $\mathbb{A}^+$ .

Notice the relation between the syntactic prefixes of the calculus and the observable actions.

The output prefixes, as in Pi-calculus, represent observable output actions.

The input prefix of the calculus, such as  $a(b)$  in the process  $a(b).P$ , does not represent an authentic action, but the capability of  $P$  to receive a name on channel  $a$ ; consequently we adopt an *early semantics* [8]: if a name  $c$  is sent on  $a$ , the input action is  $ac$  and it labels the transitions to  $P_{\{c/b\}}$ . In this way, to a single prefix  $a(b)$  correspond as many

input actions  $ac$  as names  $c$  can be sent on  $a$  in the given rate-environment. Unlike the nondeterministic case, for stochastic Pi-calculus we cannot define a *late semantics* [8] because only the input actions of type  $ac$  correspond to a measure on the space of processes, while  $a(b)$  represents a set of measures, one for each name received. Because our semantics aims to associate a measure to each process and action label, we need to refuse the inputs of type  $a(b)$  in the set of labels and chose an early semantics.

The bound output  $a[@r]$  in the form that ignores the argument of communication is novel. It labels a bound output of type  $(b@r)a[b].P$ . The example below explains its action; anticipating the semantics,  $E \vdash P \xrightarrow{\alpha, r} Q^\equiv$  means that in the environment  $E$ ,  $P$  can do an  $\alpha$ -transition with rate  $r$  to the elements of  $Q^\equiv$ .

**Example 5.2.6.** *The processes*

$$Q = (b@r)a[b].P \text{ and } R = (c@r)a[c].P_{\{c/b\}}$$

are structural congruent and we want them bisimilar in our semantics. If we consider that the (only) observable transition in which  $Q$  can be involved is  $a[b@r]$ , as it is done in other PAs, then the transition is

$$E \vdash (b@r)a[b].P \xrightarrow{a[b@r], E(a)} (b@r)P^\equiv,$$

while for  $R$  the transition is

$$E \vdash (c@r)a[c].P_{\{c/b\}} \xrightarrow{a[c@r], E(a)} (c@r)P_{\{c/b\}}^\equiv.$$

Obviously,

$$(b@r)P^\equiv = (c@r)P_{\{c/b\}}^\equiv,$$

but if  $b \neq c$ , then  $a[b@r] \neq a[c@r]$  and in effect,  $Q$  and  $R$  are not bisimilar in this interpretation.

Similarly, we expect that the processes

$$S = (b@r)a[b].P|(b@r)a[b].P \text{ and } T = (b@r)a[b].P|(c@r)a[c].P_{\{c/b\}}$$

are bisimilar and this is equally impossible if we consider the observable actions  $a[b@r]$  and  $a[c@r]$ ; because in this case the transitions are

$$E \vdash S \xrightarrow{a[b@r], 2E(a)} (b@r)P|(b@r)a[b].P^\equiv,$$

$$E \vdash T \xrightarrow{a[b@r], E(a)} (b@r)P|(c@r)a[c].P_{\{c/b\}}^\equiv \text{ and}$$

$$E \vdash T \xrightarrow{a[c@r], E(a)} (b@r)a[b].P|(c@r).P_{\{c/b\}}^\equiv.$$

For obtaining the expected bisimulations, we need that for any  $b, c \in \mathcal{N}$ ,  $a[b@r] = a[c@r]$ ; and this is equivalent with accepting that an external observer can only see that a private name with the base rate  $r$  has been sent on channel  $a$  without seeing the name. Hence, the real observable action has to be  $a[@r]$ .

Thus, the previous pairs of processes are now bisimilar, because

$$(b@r)P^{\equiv} = (c@r)P_{\{c/b\}}^{\equiv},$$

$$(b@r)a[b].P|(b@r)P^{\equiv} = (b@r)a[b].P|(c@r)P_{\{c/b\}}^{\equiv} = (b@r).P|(c@r)a[c].P_{\{c/b\}}^{\equiv},$$

$$E : Q \xrightarrow{a[@r], E(a)} (b@r)P^{\equiv},$$

$$E : R \xrightarrow{a[@r], E(a)} (c@r)P_{\{c/b\}}^{\equiv},$$

$$E : S \xrightarrow{a[@r], 2E(a)} (b@r)a[b].P|(b@r)P^{\equiv},$$

$$E : T \xrightarrow{a[@r], 2E(a)} (b@r)a[b].P|(c@r)P_{\{c/b\}}^{\equiv}.$$

Our solution is similar to the *Abstraction-Concretion method* proposed in [44] for non-deterministic Pi-calculus.  $a[@r]$  does "the job" of Abstraction, as our measurable sets of processes are Milner's *abstracted processes*. Only that in our case, because the transitions are not between processes but from processes to structural-congruence classes, we need no Concretions. So, the main advantage of our approach is that it solves the problem of bound outputs without using higher order syntax as in the classic Pi-calculus.

Before proceeding with the operational semantics, we need to define a set of operations on  $\mathfrak{M}$  that lift the process constructors of stochastic Pi-calculus to the level of the labeled distributions over the space of processes. These operations reflect the complexity of the normal forms of the  $\tau$ -reductions for stochastic Pi and for this reason the reader is invited to study Definition 5.2.7 in the context of the operational semantics presented in the next section. The SOS rules clarify and prove the correctness of these operations.

Let

$$\mathbb{A}_{@} = \{a[@r], \text{ for } a \in \mathcal{N}, r \in \mathbb{Q}^{+}\}$$

denote the set of bound output actions and

$$\mathbb{A}_a = \{a[b], ab, a[@r], \text{ for } b \in \mathcal{N}, r \in \mathbb{Q}^{+}\}$$

be the set of actions on channel  $a$ .

A labeled measure  $\mu \in \mathfrak{M}$  has *finite support* if the set of output actions  $\alpha \in \mathbb{A}^{+}$  with  $\mu(\alpha) \neq \omega$  is finite or empty. Recall that  $\omega$  denotes the null measure and  $D_{P^{\equiv}}$  the  $P^{\equiv}$ -Dirac measure defined for arbitrary  $Q \in \Pi$  by  $D_{P^{\equiv}}(Q) = 1$  if  $Q \subseteq P^{\equiv}$  and  $D_{P^{\equiv}}(Q) = 0$  otherwise.

**Definition 5.2.7.** Consider the following operations on  $\mathfrak{M}$  defined for arbitrary  $\mu, \eta \in \mathfrak{M}$ ,  $E \in \mathbb{E}$ ,  $\alpha \in \mathbb{A}^{+}$ ,  $a, b, c \in \mathcal{N}$ ,  $P \in \mathbb{P}$  and  $\mathcal{P}, \mathcal{Q}, \mathcal{R} \in \Pi$ .

### 1. Operations of arity 0.

- (a) Let  $\bar{\omega} \in \mathfrak{M}$  defined by  $\bar{\omega}(\alpha) = \omega$  for any  $\alpha \in \mathbb{A}^+$ ;
- (b) Let  $E_{P \equiv}^{a[b]}, E_{P \equiv}^{a(b)} \in \mathfrak{M}$  defined whenever  $\text{fn}(P) \subseteq \text{dom}(E)$ , by  
 $E_{P \equiv}^{a[b]}(a[b]) = E(a)D_{P \equiv}$  and  $E_{P \equiv}^{a[b]}(\alpha) = \omega$ , for  $\alpha \neq a[b]$ ;  
 $E_{P \equiv}^{a(b)}(ac) = E(a)D_{P \equiv_{\{c/b\}}}$  and  $E_{P \equiv}^{a(b)}(\alpha) = \omega$ , for  $\alpha \neq ac$ .

## 2. Operations of arity 1.

- (a) Let  $\mu_{\mathcal{P}} \in \mathcal{M}$  defined by  $\mu_{\mathcal{P}}(\alpha)(\mathcal{R}) = \mu(\alpha)(\mathcal{R}_{\mathcal{P}})$ .
- (b) Let  $(a@r)\mu \in \mathcal{M}$  defined as follows
- if  $\alpha \notin \mathbb{A}_a \cup \mathbb{A}_{@}$ , then  $(a@r)\mu(\alpha)((a@r)\mathcal{P}) = \mu(\alpha)(\mathcal{P})$
  - $(a@r)\mu(b[@r])((a@r)\mathcal{P}) = \mu(b[a])(\mathcal{P}) + \mu(b[@r])(\mathcal{P})$
  - otherwise,  $(a@r)\mu(\alpha)(\mathcal{R}) = 0$

## 3. Operations of arity 2.

- (a) Let  $\mu \oplus \eta \in \mathcal{M}$ , defined by

$$(\mu \oplus \eta)(\alpha) = \mu(\alpha) + \eta(\alpha).$$

- (b) For  $\mu, \eta \in \mathfrak{M}$  with finite support, let  $\mu \mathcal{P} \otimes_Q^E \eta \in \mathcal{M}$  defined as follows

- for  $\alpha \in \mathbb{A}$ ,

$$(\mu \mathcal{P} \otimes_Q^E \eta)(\alpha)(\mathcal{R}) = \mu_Q(\alpha)(\mathcal{R}) + \eta_{\mathcal{P}}(\alpha)(\mathcal{R});$$

- for  $\tau$ ,

$$(\mu \mathcal{P} \otimes_Q^E \eta)(\tau)(\mathcal{R}) = \mu_Q(\tau)(\mathcal{R}) + \eta_{\mathcal{P}}(\tau)(\mathcal{R}) +$$

$$\sum_{\substack{a \in \text{dom}(E)^* \\ b \in \mathcal{N} \\ \mathcal{P}_1 | \mathcal{P}_2 \subseteq \mathcal{R}}} \frac{\mu(a[b])(\mathcal{P}_1) \cdot \eta(ab)(\mathcal{P}_2) + \eta(a[b])(\mathcal{P}_1) \cdot \mu(ab)(\mathcal{P}_2)}{E(a)} +$$

$$\sum_{\substack{((x@r)y[x].P'|P'') + P''' \equiv \subseteq \mathcal{P} \\ (y(z).Q'|Q'') + Q''' \equiv \subseteq \mathcal{Q} \\ (x@r)(P'|Q'_{\{x/z\}})|P''|Q''' \equiv \subseteq \mathcal{R}}} \frac{\mu(y[@r])((x@r)P'|P'' \equiv) \cdot \eta(yx)(Q'_{\{x/z\}}|Q''' \equiv)}{E(a)} +$$

$$\sum_{\substack{(y(z).P'|P'') + P''' \equiv \subseteq \mathcal{P} \\ ((x@r)y[x].Q'|Q'') + Q''' \equiv \subseteq \mathcal{Q} \\ (x@r)(P'_{\{x/z\}}|Q')|P''|Q''' \equiv \subseteq \mathcal{R}}} \frac{\mu(yx)(P'_{\{x/z\}}|P'' \equiv) \cdot \eta(y[@r])((x@r)Q'|Q'' \equiv)}{E(a)}$$

Observe that because we work with functions with finite support and because  $\text{dom}(E)$  is defined and finite, the sums involved in the definition of  $\mu \mathcal{P} \otimes_Q^E \eta$  have finite numbers of non-zero summands.

These operations are the building blocks for the lifting of the algebraic structure of processes to the level of functions: the operations of arity 0 encode the process 0 and the prefixing, the operations of arity 1 encode the quotient and the fresh name quantification and the operations of arity 2 correspond to the choice and parallel composition.

We postpone the discussion of these operations to the next section, where their meaning will be revealed by the SOS rules. Until then, we prove that these operations are correctly defined and we state some of their basic properties.

**Lemma 5.2.8.** 1. For arbitrary  $\mu, \eta, \rho \in \mathfrak{M}$ ,  $\mu \oplus \eta \in \mathfrak{M}$  and the following equalities are satisfied

- (a)  $\mu \oplus \eta = \eta \oplus \mu$ ;
- (b)  $(\mu \oplus \eta) \oplus \rho = \mu \oplus (\eta \oplus \rho)$ ;
- (c)  $\mu = \mu \oplus \bar{\omega}$ .

2. For arbitrary  $\mu, \eta, \rho \in \mathfrak{M}$  with finite support,  $\mu \mathcal{P} \otimes_Q^E \eta \in \mathfrak{M}$  and the following equalities are satisfied

- (a)  $\mu \mathcal{P} \otimes_Q^E \eta = \eta \mathcal{Q} \otimes_P^E \mu$ ;
- (b)  $(\mu \mathcal{P} \otimes_Q^E \eta) \mathcal{P} | \mathcal{Q} \otimes_R^E \rho = \mu \mathcal{P} \otimes_{\mathcal{Q} | \mathcal{R}}^E (\eta \mathcal{Q} \otimes_R^E \rho)$ ;
- (c)  $\mu \mathcal{P} \otimes_{0=}^E \bar{\omega} = \mu$ .

## 5.2.4 Semantics

The *stochastic transition relation* is the smallest relation  $\mathfrak{T} \subseteq \mathbb{E} \times \mathbb{P} \times \mathfrak{M}$  satisfying the semantics rules listed in Table 5.1, where  $E \vdash P \rightarrow \mu$  denotes  $(E, P, \mu) \in \mathfrak{T}$ . This relation states that the behaviour of  $P$  in the environment  $E$  is defined by the mapping  $\mu \in \mathfrak{M}$ . For each  $\equiv$ -closed set of processes  $\mathcal{P} \in \Pi$  and each  $\alpha \in \mathbb{A}^+$ ,  $\mu(\alpha)(\mathcal{P}) \in \mathbb{Q}^+$  represents the total rate of the  $\alpha$ -reductions of  $P$  to the elements of  $\mathcal{P}$ .

The rules also involve predicates of type  $E \vdash \text{ok}$  that encode the correctness of  $E$ , i.e. that the environment associates base rates to a finite number of channels only, and that no channel appears in more than one rate declaration in that environment.

Recall that  $\equiv^*$  is used to denote alpha-conversion.

The rule (*Null*) guarantees that in any correct environment the behaviour of the process 0 is described by  $\bar{\omega}$ , which associates the rate 0 to any transition.

The rules (*Out*) and (*Imp*) have similar actions. They associate to any prefixed process  $x.P$ , where  $x \in \{a(b), a[b] \mid a, b \in \mathcal{N}\}$ , the mapping  $E_{P=}^x$  which, as described in Definition 5.2.7, associates the base-rate of the channel of  $x$  to the  $x$ -transitions from  $x.P$  to  $P^{\equiv}$  and rate 0 to all the other transitions<sup>2</sup>.

<sup>2</sup>The set  $\{P^{\equiv}, P \in \mathbb{P}\}$  is a base for  $\Pi$ , hence  $E_{P=}^x$  is well defined.

(Env $\epsilon$ ).	$\epsilon \vdash ok$
(Env $@$ ).	$\frac{E \vdash ok \quad a \notin \text{dom}(E)}{E, a@r \vdash ok}$
(Null).	$\frac{E \vdash ok}{E \vdash 0 \rightarrow \bar{\omega}}$
(Out).	$\frac{E \vdash ok \quad \text{fn}(a[b].P) \subseteq \text{dom}(E)}{E \vdash a[b].P \rightarrow E_{P \equiv}^{a[b]}}$
(Imp).	$\frac{E \vdash ok \quad \text{fn}(a(b).P) \subseteq \text{dom}(E)}{E \vdash a(b).P \rightarrow E_{P \equiv}^{a(b)}}$
(Sum).	$\frac{E \vdash P \rightarrow \mu \quad E \vdash Q \rightarrow \eta}{E \vdash P + Q \rightarrow \mu \oplus \eta}$
(Par).	$\frac{E \vdash P \rightarrow \mu \quad E \vdash Q \rightarrow \eta}{E \vdash P Q \rightarrow \mu_{P \equiv \otimes Q \equiv}^E \eta}$
(New).	$\frac{E, a@r \vdash P \rightarrow \mu}{E \vdash (a@r)P \rightarrow (a@r)\mu}$
(Alpha).	$\frac{E \vdash P \rightarrow \mu \quad P \equiv^* Q}{E \vdash Q \rightarrow \mu}$
(Rep).	$\frac{E \vdash P \rightarrow \mu}{E \vdash !P \rightarrow \mu_{!P \equiv}}$

Table 5.1: Semantics of Stochastic-Pi Calculus

The rule (*Sum*) computes the rate of the  $\alpha$ -transitions from  $P + Q$  to  $\mathcal{R} \in \Pi$ , as the sum of the rates of the  $\alpha$ -transitions from  $P$  and  $Q$  to  $\mathcal{R}$  respectively.

The rule (*Par*) describes the possible interactions between the processes. If

$$\rho = \mu_{P \equiv \otimes Q \equiv}^E \eta,$$

the rate  $\rho(\alpha)(\mathcal{R})$  of the  $\alpha$ -transitions from  $P|Q$  to  $\mathcal{R}$  for  $\alpha \neq \tau$ , is the sum of the rates  $\mu(\alpha)(\mathcal{R}_{Q \equiv})$  and  $\eta(\alpha)(\mathcal{R}_{P \equiv})$  of the  $\alpha$ -transitions from  $P$  to  $\mathcal{R}_Q$  and from  $Q$  to  $\mathcal{R}_P$  respectively; the rate of the  $\tau$ -transitions from  $P|Q$  to  $\mathcal{R}$  is the sum of the rates of the  $\tau$ -transitions that  $P$  or  $Q$  can do independently plus the rate of all communications between  $P$  and  $Q$  (bound communications represented by the first sum in Definition 5.2.7 3.(b) and unbound communications represented by the last two sums). Because we use

the base rate of the channel  $a$  when we calculate the rates of both inputs and outputs on  $a$ , the sums in Definition 5.2.7 3.(b) are normalised by  $E(a)$ .

The rule (*New*) establishes that the rate of the transitions from  $(a@r)P$  to  $(a@r)\mathcal{R} \in \Pi$  in the environment  $E$  is the rate of the corresponding transitions from  $P$  to  $\mathcal{R}$  in the environment  $E, a@r$ . The only thing one needs to take care of (see Definition 5.2.7) is when an output becomes bound while (*New*) is used. Consider, for instance, the process

$$Q = b[a].P + (c@r)b[c].P_{\{c/a\}}.$$

Observe that

$$E, a@r \vdash Q \xrightarrow{b[a], E(b)} P^{\equiv} \quad \text{and} \quad E, a@r \vdash Q \xrightarrow{b[@r], E(b)} (c@r)P_{\{c/a\}}^{\equiv}.$$

If we consider

$$(a@r)Q \equiv (a@r)b[a].P + (c@r)b[c].P_{\{c/a\}},$$

since  $(a@r)P \equiv (c@r)P_{\{c/a\}}$ , the rates of the transitions in the environment  $E$  should be

$$E \vdash (a@r)Q \xrightarrow{b[a], 0} (a@r)P^{\equiv} \quad \text{and} \quad E \vdash (a@r)Q \xrightarrow{b[@r], 2E(b)} (a@r)P^{\equiv}.$$

Notice that the rate of  $b[a]$ -transition of  $Q$  contributes to the rate of  $b[@r]$ -transition of  $(a@r)Q$  and this is how Definition 5.2.7 introduces  $(a@r)\mu$ .

The rule (*Rep*) encodes the intuition that in the case of stochastic systems, if

$$E \vdash P \xrightarrow{\alpha, r} Q^{\equiv}, \quad \text{then} \quad E \vdash !P \xrightarrow{\alpha, r} !P|Q^{\equiv}.$$

This replaces the axioms  $!P \equiv P|!P$  and  $!!P \equiv !P$  that are sound in PAs but not sound in SPAs.

And finally, the rule (*Alpha*) proves properties by alpha-conversion: it guarantees that the behaviour of a process does not change if the bound variables are renamed. The standard presentations of PAs with unlabelled reduction mixes structural congruence with reductions by rules of type (*Struct*). Because our reductions are labelled (the labels are hidden into the mappings), alpha conversion needs to be separately incorporated both in the algebra and coalgebra.

The next example illustrates some transitions in our framework.

**Example 5.2.9. I.** *We prove that*

$$E \vdash (b@r)(a[b].P)|a(c).Q \xrightarrow{\tau, E(a)} (b@r)(P|Q_{\{b/c\}})^{\equiv}.$$

*Applying (Out) we derive*

$$E, b@r \vdash a[b].P \xrightarrow{a[b], E(a)} P^{\equiv}.$$

*(New) gives us further that*

$$E \vdash (b@r)a[b].P \xrightarrow{a[@r], E(a)} (b@r)P^{\equiv}$$

and this is the only transition with non-zero rate. Observe that the definition of  $E_{Q \equiv}^{a(c)}$  implies

$$E \vdash a(c).Q \xrightarrow{ab, E(a)} Q_{\{b/c\}}^{\equiv}.$$

Applying the definition of  $(b@r)(a[b].P) \equiv \otimes_{a(c).Q}^E$ , we obtain

$$E \vdash (b@r)(a[b].P) | a(c).Q \xrightarrow{\tau, E(a)} (b@r)(P | Q_{\{b/c\}})^{\equiv}.$$

A consequence of this result is the well-known case of communication of a private name used for a private communication:

$$E \vdash (b@r)(a[b].b(e).P) | a(c).c[d].0 \xrightarrow{\tau, E(a)} (b@r)(b(e).P | b[d].0)^{\equiv} \xrightarrow{\tau, r} (b@r)P_{\{d/e\}}^{\equiv}.$$

The first transition is a particular case of the example. For the second transition, we apply the case 3(b) of Definition 5.2.7.

**II.** We can also prove that

$$E \vdash (b@r)(a[b].P) | (b@r)(a[b].P) | a(c).Q \xrightarrow{\tau, 2E(a)} (b@r)(a[b].P) | (b@r)(P | Q_{\{b/c\}})^{\equiv}.$$

The proof is similar to the previous case with the only difference that

$$E \vdash (b@r)a[b].P | (b@r)a[b].P \xrightarrow{a[@r], 2E(a)} (b@r)P | (b@r)a[b].P^{\equiv},$$

which induces the result.

Observe also that the case 3 (b) of Definition 5.2.7 guarantees that

$$E \vdash (b@r)a[b].P | (c@r)a[c].P_{\{b/c\}} \xrightarrow{a[@r], 2E(a)} (b@r)P | (b@r)a[b].P^{\equiv}.$$

**Remark 5.2.10.** In stochastic Pi calculus it is not possible to define a binary operator on  $\mathfrak{M}$  that reflects, for a fixed environment  $E$ , the parallel composition of processes.

To prove this, assume that there exists an operator  $\otimes^E$  such that if

$$E \vdash P \rightarrow \mu \quad \text{and} \quad E \vdash Q \rightarrow \eta,$$

then

$$E \vdash P | Q \rightarrow \mu \otimes^E \eta.$$

The processes

$$P = a[b].0 | c[d].0 \quad \text{and} \quad Q = a[b].c[d].0 + c[d].a[b].0$$

have associated, in any correct environment  $E$ , the same mapping  $\mu \in \mathfrak{M}$ .

Suppose that  $E \vdash R \rightarrow \eta$ , where  $R = e[f].0$ .

If, indeed, the operator  $\otimes^E$  is well defined, then

$$E \vdash P|R \rightarrow \mu \otimes^E \eta \quad \text{and} \quad E \vdash Q|R \rightarrow \mu \otimes^E \eta,$$

i.e.  $P|R$  and  $Q|R$  have associated the same mapping. But this is not the case, because  $P^\equiv \neq Q^\equiv$  and

$$E \vdash P|R \xrightarrow{e[f],E(e)} P^\equiv \quad \text{and} \quad E \vdash P|R \xrightarrow{e[f],0} Q^\equiv,$$

while

$$E \vdash Q|R \xrightarrow{e[f],0} P^\equiv \quad \text{and} \quad E \vdash Q|R \xrightarrow{e[f],E(e)} Q^\equiv.$$

This explains why we need to index  $\otimes^E$  with  $P^\equiv$  and  $Q^\equiv$  and why the algebraic signature is changed when the structure of processes is lifted to indexed measures.

The next theorem states that  $\mathfrak{T}$  is well defined and characterizes the correctness of an environment.

**Theorem 5.2.11.** 1. If  $E \vdash ok$  and  $fn(P) \subseteq dom(E)$ , then there exists a unique  $\mu \in \mathfrak{M}$  such that  $E \vdash P \rightarrow \mu$ .

2. If  $E \vdash P \rightarrow \mu$ , then  $E \vdash ok$ . Moreover,  $E \vdash ok$  iff  $E \vdash 0 \rightarrow \bar{w}$ .

*Proof.* 1. The existential part is proved by induction on the structure of  $P$  and the uniqueness by induction on derivations.

Firstly we prove the existential part.

**For  $P = 0$  and  $P = x.Q$ :** (Null), (Imp) and (Out) guarantee the existence of  $\mu$ .

**For  $P = Q + R$ :** because  $fn(P) = fn(Q) \cup fn(R)$ ,  $fn(Q) \subseteq dom(E)$  and  $fn(R) \subseteq dom(E)$ . We use the inductive hypothesis and obtain that exist two functions  $\eta, \rho$  such that  $E \vdash Q \rightarrow \eta$  and  $E \vdash R \rightarrow \rho$ . From (Sum) we obtain that exists  $\mu = \eta \oplus \rho$  such that  $E \vdash P \rightarrow \mu$ .

**For  $P = Q|R$ :** because  $fn(P) = fn(Q) \cup fn(R)$ ,  $fn(Q) \subseteq dom(E)$  and  $fn(R) \subseteq dom(E)$ . We use the inductive hypothesis and obtain that exist two functions  $\eta, \rho$  such that  $E \vdash Q \rightarrow \eta$  and  $E \vdash R \rightarrow \rho$ . From (Par) we obtain that exists  $\mu = \eta \otimes_{Q^\equiv R^\equiv}^E \rho$  such that  $E \vdash P \rightarrow \mu$ .

**For  $P = (a@r)Q$ :** if  $a \notin dom(E)$ , then  $E, a@r \vdash ok$  and the inductive hypothesis guarantees the existence of  $\eta$  such that  $E, a@r \vdash Q \rightarrow \eta$ . Further, applying (New), we get  $E \vdash P \rightarrow (a@r)\eta$ . If  $a \in dom(E)$ , let  $b \in \mathcal{N} \setminus dom(E)$ . Then  $E, b@r \vdash ok$  and the inductive hypothesis guarantees the existence of  $\eta$  such that  $E, b@r \vdash Q_{\{b/a\}} \rightarrow \eta$ . Further, applying (New), we get

$$E \vdash (b@r)Q_{\{b/a\}} \rightarrow (b@r)\eta$$

and (Alpha) gives

$$E \vdash (a@r)Q \rightarrow (b@r)\eta.$$

Consequently, in all the cases there exists  $\mu$  such that  $E \vdash (a@r)Q \rightarrow \mu$ .

**For  $P = !Q$ :** because  $fn(Q) = fn(P)$ , the inductive hypothesis guarantees the existence of a unique  $\eta$  such that  $E \vdash Q \rightarrow \eta$ . Further, applying (Rep), we get  $E \vdash P \rightarrow \eta!Q$ .

The uniqueness part is done by induction on derivations.

The rules (Env $\epsilon$ ) and (Env@) are only proving the correctness of environments and consequently will not interfere with our proof.

Observe that all the derivations involving only the rules (Sum), (Par), (New) and (Rep), called in what follows *basic proofs*, demonstrate properties about processes with a more complex syntax than the processes involved in the hypotheses.

Consequently, taking (Null), (Imp) and (Out) as basic cases, an induction on the structures of the processes involved in the derivations shows the uniqueness of  $\mu$  for the situation of the basic proofs.

Notice, however, that due to (New), a basic proof proves properties of type  $E \vdash P \rightarrow \mu$  only for cases when  $new(P) \cap dom(E) = \emptyset$ , where  $new(P)$  is the set of names of  $P$  bound by fresh name quantifiers. To conclude the proof, we need to show that if  $Q = P_{\{a/b\}}$  with  $a, b \notin fn(P)$  and if  $E \vdash P \rightarrow \mu$  and  $E \vdash Q \rightarrow \eta$  can be proved with basic proofs, then  $\mu = \eta$ .

We do this by induction on  $P$ .

If  $P = 0$ , then  $Q = 0$  and  $\eta = \mu = \bar{\omega}$ .

If  $P = c[d].R$ , then  $Q = c[d].R_{\{a/b\}}$  and  $a, b \notin fn(R)$ . Moreover,  $\mu = E_{R=}^{c[d]}$  and  $\eta = E_{R_{\{a/b\}}}^{c[d]}$ . But because  $a, b \notin fn(R)$ , we have that  $R \equiv R_{\{a/b\}}$  implying further  $\mu = \eta$ .

If  $P = c(d).R$ , then if  $d \neq b$  the proof goes as in the previous case. If  $P = c(b).R$ , then  $Q = c(a).R_{\{a/b\}}$ ,  $\mu = E_R^{c(b)}$  and  $\eta = E_{R_{\{a/b\}}}^{c(a)}$ . It is trivial to verify that  $\mu = \eta$ .

If  $P = S + T$ , then  $Q = S_{\{a/b\}} + T_{\{a/b\}}$ . Suppose that  $E \vdash S \rightarrow \rho$  and  $E \vdash T \rightarrow \nu$ , then from the inductive hypothesis,  $E \vdash S_{\{a/b\}} \rightarrow \rho$  and  $E \vdash T_{\{a/b\}} \rightarrow \nu$ . Hence,  $\mu = \eta = \rho \oplus \nu$ .

If  $P = S|T$  the proof goes as in the previous case.

If  $P = !R$ ,  $Q = !R_{\{a/b\}}$ . Suppose that  $E \vdash R \rightarrow \rho$ . From the inductive hypothesis we also obtain that  $E \vdash R_{\{a/b\}} \rightarrow \rho$ . The conclusion derives further from the fact that  $!R \equiv !R_{\{a/b\}}$  because  $a, b \notin fn(R)$ .

If  $P = (c@r)R$  with  $c \neq b$ , then  $Q = (c@r)R_{\{a/b\}}$ . Because we are in the case of a basic proof,  $c \notin dom(E)$ . Suppose that  $E, c@r \vdash R \rightarrow \rho$ . This is the unique hypothesis that proves  $E \vdash P \rightarrow \mu$ . Then,  $\mu = (c@r)\rho$  and the inductive hypothesis implies that  $E, c@r \vdash R_{\{a/b\}} \rightarrow \rho$  is the unique hypothesis that proves  $E \vdash Q \rightarrow \eta$ . Further, we get  $E \vdash (c@r)R_{\{a/b\}} \rightarrow (c@r)\rho$ . Hence, in this case,  $\mu = \eta$ .

If  $P = (b@r)R$ , then  $Q = (a@r)R_{\{a/b\}}$ . Because we work with basic proofs, we have  $a, b \notin dom(E)$ . A simple induction proves that

$$\text{if } E, b@r \vdash R \rightarrow \rho, \text{ then } E, a@r \vdash R_{\{a/b\}} \rightarrow \rho',$$

where for any  $\alpha \in \mathbb{A}^+$  and any  $\mathcal{R} \in \Pi$ ,  $\rho(\alpha)(\mathcal{R}) = \rho'(\alpha_{\{a/b\}})(\mathcal{R}_{\{a/b\}})$ .

From here we get  $(b@r)\rho = (a@r)\rho'$ . Observe that  $E, b@r \vdash R \rightarrow \rho$  is the unique

hypothesis that can be used in a basic proof to derive  $E \vdash (b@r)R \rightarrow \mu$  and  $\mu = (b@r)\rho$ . Similarly,  $E, a@r \vdash R_{\{a/b\}} \rightarrow \rho'$  is the unique hypothesis to prove  $E \vdash (a@r)R_{\{a/b\}} \rightarrow \eta$  and  $\eta = (a@r)\rho'$ . Hence, also in this case,  $\mu = \eta$ .

In this way we have proved that any couple of alpha-converted processes have associated the same mapping by basic proofs. In addition, (Alpha) guarantees that any kind of proofs will associate to alpha-converted processes the same mapping and this concludes our proof.

2. We prove the first part by induction on derivations. The second part is a consequence of the first part and (Null).

If  $E \vdash P \rightarrow \mu$  is proved by (Null), (Imp) or (Out),  $E \vdash ok$  is required as hypothesis.

If  $E \vdash P \rightarrow \mu$  is proved by (Sum),  $P = Q + R$ ,  $\mu = \eta \oplus \rho$  and  $E \vdash Q \rightarrow \eta$  and  $E \vdash R \rightarrow \rho$  are the hypothesis and we can use the inductive hypothesis.

If  $E \vdash P \rightarrow \mu$  is proved by (Par), the argument goes as in the previous case.

If  $E \vdash P \rightarrow \mu$  is proved by (New), then  $P = (a@r)Q$  and the hypothesis is of type  $E, a@r \vdash Q \rightarrow \eta$ . The inductive hypothesis gives  $E, a@r \vdash ok$  and this can only be proved by (Env@) from  $E \vdash ok$ .

If  $E \vdash P \rightarrow \mu$  is proved by (Rep), then  $P = !Q$ ,  $\mu = \mu'_{!Q}$  and  $E \vdash Q \rightarrow \mu'$  is the hypothesis and we can apply the inductive step.

If  $E \vdash P \rightarrow \mu$  is proved by (Alpha), we can use the inductive hypothesis again.  $\square$

Unlike in other process algebras, our semantics does not contain a (Struct) rule stating that structural congruent processes behave identically. However, such a result can be proved.

**Theorem 5.2.12.** *If  $E \vdash P' \rightarrow \mu$  and  $P' \equiv P''$ , then  $E \vdash P'' \rightarrow \mu$ .*

The next lemma describes how the environments can vary without influencing the mapping associated to a process.

**Lemma 5.2.13.** 1. *If for any  $a \in \mathcal{N}$  and  $r \in \mathcal{Q}$ ,  $[a@r \in E \text{ iff } a@r \in E']$ , then  $[E \vdash P \rightarrow \mu \text{ iff } E' \vdash P \rightarrow \mu]$ .*

2. *If  $E' \vdash ok$ ,  $E \subset E'$  and  $E \vdash P \rightarrow \mu$ , then  $E' \vdash P \rightarrow \mu$ .*

3. *If  $E \subset E'$ ,  $E \vdash P \rightarrow \mu$  and  $dom(E' \setminus E) \cap fn(P) = \emptyset$ , then  $E' \vdash P \rightarrow \mu$ .*

*Proof.* 1. A simple induction on derivations that involve only (Env $\varepsilon$ ) and (Env@) proves that  $E \vdash ok$  iff  $E' \vdash ok$ . For proving our lemma we will proceed with an induction on the derivation of  $E \vdash P \rightarrow \mu$ .

If  $E \vdash P \rightarrow \mu$  is proved by (Null), we have that  $P = 0$  and due to Theorem 5.2.11,  $\mu = \bar{\omega}$ . Applying (Null) we obtain  $E' \vdash P \rightarrow \mu$ .

If  $E \vdash P \rightarrow \mu$  is proved by (Imp) or (Out), we have that  $P = x.Q$  and due to Theorem 5.2.11,  $\mu = E_Q^x$ . Because  $E_Q^x = E_Q'^x$  and  $dom(E) = dom(E')$ , we obtain  $E' \vdash P \rightarrow \mu$ .

If  $E \vdash P \rightarrow \mu$  is proved by (Sum), we have that  $P = Q + R$ ,  $\mu = \eta \oplus \rho$  and the hypothesis are  $E \vdash Q \rightarrow \eta$  and  $E \vdash R \rightarrow \rho$ . From the inductive hypothesis we obtain  $E' \vdash Q \rightarrow \eta$  and  $E' \vdash R \rightarrow \rho$ . Further, applying (Sum) we get  $E' \vdash P \rightarrow \mu$ .

If  $E \vdash P \rightarrow \mu$  is proved by (Par) we have that  $P = Q|R$ ,  $\mu = \eta \otimes_{R=}^E \rho$  and the hypothesis are  $E \vdash Q \rightarrow \eta$  and  $E \vdash R \rightarrow \rho$ . From the inductive hypothesis we obtain  $E' \vdash Q \rightarrow \eta$  and  $E' \vdash R \rightarrow \rho$ . Further, applying (Par) we get  $E' \vdash P \rightarrow \eta \otimes_{R=}^{E'} \rho$ . But  $\eta \otimes_{R=}^E \rho = \eta \otimes_{R=}^{E'} \rho$ .

If  $E \vdash P \rightarrow \mu$  is proved by (Rep), we have that  $P = !Q$ ,  $\mu = \eta!_Q$  and the hypothesis is  $E \vdash Q \rightarrow \eta$ . Applying the inductive step we get  $E' \vdash Q \rightarrow \eta$  and (Rep) guarantees that  $E' \vdash P \rightarrow \mu$ .

If  $E \vdash P \rightarrow \mu$  is proved by (New), we have that  $P = (a@r)Q$ ,  $\mu = (a@r)\eta$  and the hypothesis is  $E, a@r \vdash Q \rightarrow \eta$ . Hence,  $a \notin \text{dom}(E) = \text{dom}(E')$  and we can apply the inductive hypothesis because  $b@s \in E, a@r$  iff  $b@s \in E', a@r$  and obtain  $E', a@r \vdash Q \rightarrow \eta$  where from we get  $E' \vdash P \rightarrow \mu$ .

If  $E \vdash P \rightarrow \mu$  is proved by (Alpha), we have that  $P = Q_{\{a/b\}}$  with  $a, b \notin \text{fn}(P) = \text{fn}(Q)$  and the hypothesis is  $E \vdash Q \rightarrow \mu$ . As before, the inductive hypothesis guarantees that  $E' \vdash Q \rightarrow \mu$  and because  $a, b \notin \text{fn}(Q)$ , (Alpha) proves  $E' \vdash P \rightarrow \mu$ .

## 2. Induction on the derivation of $E \vdash P \rightarrow \mu$ .

If  $E \vdash P \rightarrow \mu$  is proved by (Null), we have that  $P = 0$  and due to Theorem 5.2.11,  $\mu = \bar{w}$ . Applying (Null) we obtain  $E' \vdash P \rightarrow \mu$ .

If  $E \vdash P \rightarrow \mu$  is proved by (Imp) or (Out), we have that  $P = x.Q$  and due to Theorem 5.2.11,  $\mu = E_{Q=}^x$ . Because  $\text{fn}(P) \subseteq \text{dom}(E) \subseteq \text{dom}(E')$  and  $E_{Q=}^x = E_{Q=}^{x'}$ , we obtain  $E' \vdash P \rightarrow \mu$ .

If  $E \vdash P \rightarrow \mu$  is proved by (Sum), we have that  $P = Q + R$ ,  $\mu = \eta \oplus \rho$  and the hypothesis are  $E \vdash Q \rightarrow \eta$  and  $E \vdash R \rightarrow \rho$ . From the inductive hypothesis we obtain  $E' \vdash Q \rightarrow \eta$  and  $E' \vdash R \rightarrow \rho$ . Further, applying (Sum) we get  $E' \vdash P \rightarrow \mu$ .

If  $E \vdash P \rightarrow \mu$  is proved by (Par) we have that  $P = Q|R$ ,  $\mu = \eta \otimes_{R=}^E \rho$  and the hypothesis are  $E \vdash Q \rightarrow \eta$  and  $E \vdash R \rightarrow \rho$ . From the inductive hypothesis we obtain  $E' \vdash Q \rightarrow \eta$  and  $E' \vdash R \rightarrow \rho$ . Further, applying (Par) we get  $E' \vdash P \rightarrow \eta \otimes_{R=}^{E'} \rho$ . But  $\eta \otimes_{R=}^E \rho = \eta \otimes_{R=}^{E'} \rho$ .

If  $E \vdash P \rightarrow \mu$  is proved by (Rep), we have that  $P = !Q$ ,  $\mu = \eta!_Q$  and the hypothesis is  $E \vdash Q \rightarrow \eta$ . Applying the inductive step we get  $E' \vdash Q \rightarrow \eta$  and (Rep) guarantees that  $E' \vdash P \rightarrow \mu$ .

If  $E \vdash P \rightarrow \mu$  is proved by (Alpha), we have that  $P = Q_{\{a/b\}}$  with  $a, b \notin \text{fn}(P) = \text{fn}(Q)$  and the hypothesis is  $E \vdash Q \rightarrow \mu$ . As before, the inductive hypothesis guarantees that  $E' \vdash Q \rightarrow \mu$  and because  $a, b \notin \text{fn}(Q)$ , (Alpha) proves that  $E' \vdash P \rightarrow \mu$ .

If  $E \vdash P \rightarrow \mu$  is proved by (New), we have that  $P = (a@r)Q$ ,  $\mu = (a@r)\eta$  and the hypothesis is  $E, a@r \vdash Q \rightarrow \eta$ . Hence,  $a \notin \text{dom}(E)$ . If  $a \notin \text{dom}(E')$ , the inductive hypothesis guarantees that  $E', a@r \vdash Q \rightarrow \eta$  where from we get  $E' \vdash P \rightarrow \mu$ . If  $a \in \text{dom}(E')$ , let  $b \notin \text{dom}(E') \cup \text{fn}(P)$ . Because  $E, a@r \vdash Q \rightarrow \eta$  is provable, also

$$E, b@r \vdash Q_{\{b/a\}} \rightarrow \eta_{\{b/a\}}$$

is provable, where  $\eta_{\{b/a\}}$  is the mapping obtained from  $\eta$  replacing all the occurrences of  $a$  in the definition of  $\eta$  (in processes and labels) with  $b$ . Moreover, to each proof of  $E, a@r \vdash Q \rightarrow \eta$  corresponds a proof of  $E, b@r \vdash Q_{\{b/a\}} \rightarrow \eta_{\{b/a\}}$  that is, from the point of view of our induction, at the same level with the proof of  $E, a@r \vdash Q \rightarrow \eta$ . Consequently, we can apply the inductive hypothesis to  $E, b@r \vdash Q_{\{b/a\}} \rightarrow \eta_{\{b/a\}}$  and obtain  $E', b@r \vdash Q_{\{b/a\}} \rightarrow \eta_{\{b/a\}}$ . (New) implies  $E' \vdash (b@r)Q_{\{b/a\}} \rightarrow (b@r)\eta_{\{b/a\}}$  and (Alpha)  $E' \vdash (a@r)Q \rightarrow (b@r)\eta_{\{b/a\}}$ . To conclude, it is sufficient to verify that  $(a@r)\eta = (b@r)\eta_{\{b/a\}}$ .

3. The proof goes similarly with the proof of the previous case. We use an induction on the derivation of  $E \vdash P \rightarrow \mu$ .

If  $E \vdash P \rightarrow \mu$  is proved by (Null), we have that  $P = 0$  and  $\mu = \bar{\omega}$ . Applying (Null) we obtain  $E' \vdash P \rightarrow \mu$ .

If  $E \vdash P \rightarrow \mu$  is proved by (Imp) or (Out), we have that  $P = x.Q$  and  $\mu = E'_{Q=}^x$ . Because  $fn(P) \subseteq dom(E)$ ,  $fn(P) \cap dom(E \setminus E') = \emptyset$  and  $E_{Q=}^x = E'_{Q=}^x$ , we obtain  $E' \vdash P \rightarrow \mu$ .

If  $E \vdash P \rightarrow \mu$  is proved by (Sum), we have that  $P = Q + R$ ,  $\mu = \eta \oplus \rho$  and the hypothesis are  $E \vdash Q \rightarrow \eta$  and  $E \vdash R \rightarrow \rho$ . From the inductive hypothesis we obtain  $E' \vdash Q \rightarrow \eta$  and  $E' \vdash R \rightarrow \rho$ . Further, applying (Sum) we get  $E' \vdash P \rightarrow \mu$ .

If  $E \vdash P \rightarrow \mu$  is proved by (Par) we have that  $P = Q|R$ ,  $\mu = \eta_{Q=\otimes_R^E} \rho$  and the hypothesis are  $E \vdash Q \rightarrow \eta$  and  $E \vdash R \rightarrow \rho$ . From the inductive hypothesis we obtain  $E' \vdash Q \rightarrow \eta$  and  $E' \vdash R \rightarrow \rho$ . Further, applying (Par) we get  $E' \vdash P \rightarrow \eta_{Q=\otimes_R^{E'}} \rho$ . But  $\eta_{Q=\otimes_R^E} \rho = \eta_{Q=\otimes_R^{E'}} \rho$ .

If  $E \vdash P \rightarrow \mu$  is proved by (Rep), we have that  $P = !Q$ ,  $\mu = \eta_{!Q}$  and the hypothesis is  $E \vdash Q \rightarrow \eta$ . Applying the inductive step we get  $E' \vdash Q \rightarrow \eta$  and (Rep) guarantees that  $E' \vdash P \rightarrow \mu$ .

If  $E \vdash P \rightarrow \mu$  is proved by (Alpha), we have that  $P = Q_{\{a/b\}}$  with  $a, b \notin fn(P) = fn(Q)$  and the hypothesis is  $E \vdash Q \rightarrow \mu$ . As before, the inductive hypothesis guarantees that  $E' \vdash Q \rightarrow \mu$  and because  $a, b \notin fn(Q)$ , (Alpha) proves that  $E' \vdash P \rightarrow \mu$ .

If  $E \vdash P \rightarrow \mu$  is proved by (New), we have that  $P = (a@r)Q$ ,  $\mu = (a@r)\eta$  and the hypothesis is  $E, a@r \vdash Q \rightarrow \eta$ . Hence,  $a \notin dom(E)$  and because  $dom(E') \subseteq dom(E)$ , we obtain that  $a \notin dom(E')$ . Because  $E, a@r \subset E', a@r$  and

$$dom((E', a@r) \setminus (E, a@r)) = dom(E' \setminus E),$$

we can apply the inductive hypothesis and from  $E, a@r \vdash Q \rightarrow \eta$  we obtain  $E', a@r \vdash Q \rightarrow \eta$  where from we get  $E' \vdash P \rightarrow \mu$ .  $\square$

### 5.3 Stochastic bisimulation

In this section we focus on stochastic bisimulation that reproduces, at the stochastic level, Larsen-Skou probabilistic bisimulation [40]. We have introduced a similar concept

in previous chapter for the case of stochastic CCS [17]. The novelty with the present definition consists in the role of the rate environments:

*two processes are stochastic bisimilar if they have similar stochastic behaviours in any rate environment.*

**Definition 5.3.1** (Stochastic Bisimulation). *A rate-bisimulation on  $\mathbb{P}$  is an equivalence relation  $\mathfrak{R} \subseteq \mathbb{P} \times \mathbb{P}$  such that  $(P, Q) \in \mathfrak{R}$  iff for any  $E \in \mathbb{E}$ ,*

- *if  $E \vdash P \rightarrow \mu$ , then there exists  $\eta \in \mathfrak{M}$  such that  $E \vdash Q \rightarrow \eta$  and for any  $C \in \Pi(\mathfrak{R})$  and  $\alpha \in \mathbb{A}^+$ ,  $\mu(\alpha)(C) = \eta(\alpha)(C)$ .*
- *if  $E \vdash Q \rightarrow \eta$ , then there exists  $\mu \in \mathfrak{M}$  such that  $E \vdash P \rightarrow \mu$  and for any  $C \in \Pi(\mathfrak{R})$  and  $\alpha \in \mathbb{A}^+$ ,  $\eta(\alpha)(C) = \mu(\alpha)(C)$ .*

*Two processes  $P, Q \in \mathbb{P}$  are stochastic bisimilar, denoted  $P \sim Q$ , if there exists a rate-bisimulation connecting them.*

Observe that stochastic bisimulation is the largest rate-bisimulation on  $\mathbb{P}$ .

**Example 5.3.2.** *If  $a, b, x, y \in \mathcal{N}$ ,  $a \neq b$  and  $x \notin \text{fn}(b[y].Q)$ , then*

$$a(x).P|b[y].Q \sim a(x).(P|b[y].Q) + b[y].(a(x).P|Q).$$

*Indeed, for any compatible rate environment  $E$ ,*

$$E \vdash a(x).P|b[y].Q \rightarrow (E_{P \equiv}^{a(x)})_{(a(x).P) \equiv \otimes (b[y].Q) \equiv}^E (E_{Q \equiv}^{b[y]}),$$

$$E \vdash a(x).(P|b[y].Q) + b[y].(a(x).P|Q) \rightarrow E_{P|b[y].Q}^{a(x)} \oplus E_{a(x).P|Q}^{b[y]}$$

*and for arbitrary  $C \in \Pi(\sim)$ ,*

$$[(E_{P \equiv}^{a(x)})_{(a(x).P) \equiv \otimes (b[y].Q) \equiv}^E (E_{Q \equiv}^{b[y]})](C) = [E_{P|b[y].Q}^{a(x)} \oplus E_{a(x).P|Q}^{b[y]}](\alpha)(C).$$

The previous example shows bisimilar processes which are not structurally congruent. However, the reverse affirmation is not true, as one can simply demonstrate inductively on the structural congruence rules by applying the properties stated in Theorem 5.2.8.

**Theorem 5.3.3.** *If  $P' \equiv P''$ , then  $P' \sim P''$ .*

*Proof.* From  $P' \equiv P''$  we obtain that  $\text{fn}(P') = \text{fn}(P'')$  and Theorem 5.2.11 ensures that  $E \vdash P' \rightarrow \mu$  implies that there exists a unique  $\mu'$  such that  $E \vdash P'' \rightarrow \mu'$ .

We prove now that  $E \vdash P' \rightarrow \mu$  implies  $E \vdash P'' \rightarrow \mu$ . The proof is an induction following the rules of structural congruence presented in Definition 5.2.2.

**Rule 1.a:** if  $P' = P|Q$  and  $P'' = Q|P$ . Suppose that  $E \vdash P \rightarrow \eta$  and  $E \vdash Q \rightarrow \rho$ . Then  $\mu = \eta \otimes_Q^E \rho$  and Lemma 4.4.2 guarantees that  $E \vdash P'' \rightarrow \mu$ .

Similarly we can treat all the rules of group 1.

**Rules of group 2:** As previously, the results derive from the properties of  $\oplus$  stated in Lemma 5.2.8.

**Rules of group 3:** If  $(P' = P|R$  and  $P'' = Q|R)$ , or  $(P' = P + R$  and  $P'' = Q + R)$ , or  $(P' = x.P$  and  $P'' = x.Q)$ , or  $(P' = !P$  and  $P'' = !Q)$  for  $P \equiv Q$ , we can apply the inductive hypothesis that guarantees that  $E \vdash P \rightarrow \eta$  iff  $E \vdash Q \rightarrow \eta$ . Further, if  $E \vdash R \rightarrow \rho$ , we obtain the desired results because  $\eta \stackrel{E}{P \equiv Q} \rho = \eta \stackrel{E}{Q \equiv P} \rho$ ,  $\eta \oplus \rho = \eta \oplus \rho$ ,  $E_{P \equiv}^x = E_{Q \equiv}^x$  and  $\mu_{!P} = \mu_{!Q}$ .

If  $P' = (a@r)P$  and  $P'' = (a@r)Q$ , we have two subcases.

**Subcase 1:**  $a \notin \text{dom}(E)$ . Suppose that  $E, a@r \vdash P \rightarrow \eta$ . From the inductive hypothesis we obtain that  $E, a@r \vdash Q \rightarrow \eta$ . Further, rule (New) proves that  $\mu = (a@r)\eta$  and  $E \vdash (a@r)Q \rightarrow \mu$ .

**Subcase 2:**  $a \in \text{dom}(E)$ . Let  $b \in \mathcal{N} \setminus \text{dom}(E)$ . Suppose that  $E, b@r \vdash P_{\{b/a\}} \rightarrow \eta$ . Then, (New) implies

$$E \vdash (b@r)P_{\{b/a\}} \rightarrow (b@r)\eta$$

and (Alpha) proves

$$E \vdash (a@r)P \rightarrow (b@r)\eta.$$

Hence,  $\mu = (b@r)\eta$ . On the other hand, the inductive hypothesis implies

$$E, b@r \vdash Q_{\{b/a\}} \rightarrow \eta,$$

(New) proves

$$E \vdash (b@r)Q_{\{b/a\}} \rightarrow (b@r)\eta$$

and (Alpha) implies

$$E \vdash (a@r)Q \rightarrow (b@r)\eta.$$

**Rule 4.a:** If  $P' = (a@r)(b@s)P$  and  $P'' = (b@s)(a@r)P$ . Let  $c, d \in \mathcal{N} \setminus \text{dom}(E)$ . Suppose that

$$E; c@r; d@s \vdash P_{\{c/a, d/b\}} \rightarrow \eta.$$

Applying twice (New) we obtain

$$E \vdash (c@r)(d@s)P_{\{c/a, d/b\}} \rightarrow (c@r)(d@s)\eta$$

and applying twice (Alpha) we get

$$E \vdash (a@r)(b@s)P \rightarrow (c@r)(d@s)\eta.$$

Hence,  $\mu = (c@r)(d@s)\eta$ . On the other hand, Lemma 5.2.13.1 guarantees that

$$E; c@r; d@s \vdash P_{\{c/a, d/b\}} \rightarrow \eta$$

implies

$$E; d@s; c@r \vdash P_{\{c/a, d/b\}} \rightarrow \eta$$

and, as before, we eventually obtain

$$E \vdash (b@s)(a@r)P \rightarrow (d@s)(c@r)\eta.$$

Now it is sufficient to verify that  $(d@s)(c@r)\eta = (c@r)(d@s)\eta$ .

**Rule 4.b:** If  $P' = (a@r)0$  and  $P'' = 0$ . In this case it is sufficient to notice that

$$(a@r)\bar{\omega} = \bar{\omega}.$$

**Rule 4.c:** If  $P' = (a@r)(P|Q)$  and  $P'' = P|(a@r)Q$ , where  $a \notin fn(P)$ .

Let  $b \in \mathcal{N} \setminus (dom(E) \cup fn(P))$ . Suppose that

$$E, b@r \vdash P \rightarrow \eta \text{ and } E, b@r \vdash Q_{\{b/a\}} \rightarrow \rho.$$

Observe that because  $a \notin fn(P)$ , we also have  $E, b@r \vdash P_{\{b/a\}} \rightarrow \eta$ . Further we obtain

$$E, b@r \vdash (P|Q)_{\{b/a\}} \rightarrow \eta \text{ }_{(P_{\{b/a\}}) \equiv \otimes_{(Q_{\{b/a\}}) \equiv}^{(E, b@r)}} \rho \text{ and}$$

$$E \vdash (b@r)((P|Q)_{\{b/a\}}) \rightarrow (b@r)(\eta \text{ }_{(P_{\{b/a\}}) \equiv \otimes_{(Q_{\{b/a\}}) \equiv}^{(E, b@r)}} \rho).$$

Now we apply (Alpha) and obtain

$$E \vdash (a@r)((P|Q)) \rightarrow (b@r)(\eta \text{ }_{(P_{\{b/a\}}) \equiv \otimes_{(Q_{\{b/a\}}) \equiv}^{(E, b@r)}} \rho).$$

On the other hand, because  $b \notin fn(P)$ , from  $E, b@r \vdash P \rightarrow \eta$  Lemma 5.2.13.2 proves

$$E \vdash P \rightarrow \eta$$

and from  $E, b@r \vdash Q_{\{b/a\}} \rightarrow \rho$  we obtain, applying (New),

$$E \vdash (b@r)Q_{\{b/a\}} \rightarrow (b@r)\rho.$$

And further,

$$E \vdash P|(b@r)Q_{\{b/a\}} \rightarrow \eta \text{ }_{P \equiv \otimes_{((b@r)Q_{\{b/a\}}) \equiv}^E} (b@r)\rho.$$

Applying (Alpha) we obtain

$$E \vdash P|(a@r)Q \rightarrow \eta \text{ }_{P \equiv \otimes_{((b@r)Q_{\{b/a\}}) \equiv}^E} (b@r)\rho.$$

A simple verification based on the observation that (if for all  $R \in \mathcal{R}$ ,  $b \notin fn(R)$ , then  $(b@r)\mathcal{R} = \mathcal{R}$ ) proves that

$$(b@r)(\eta \text{ }_{P \equiv \otimes_{(Q_{\{b/a\}}) \equiv}^{E, b@r}} \rho) = \eta \text{ }_{P \equiv \otimes_{((b@r)Q_{\{b/a\}}) \equiv}^E} (b@r)\rho.$$

Similarly, one can prove that case  $P' = (a@r)(P + Q)$  and  $P'' = P + (a@r)Q$ , where  $a \notin fn(P)$ .

**Rules of group 5:** By a simple verification one can prove that  $\bar{w}_{!0} = \bar{w}$ . For the second rule, observe that if  $E \vdash P \rightarrow \eta$  and  $E \vdash Q \rightarrow \rho$ , then

$$E \vdash !(P|Q) \rightarrow (\eta \otimes_Q^E \rho)_{!(P|Q)}$$

and

$$E \vdash !P|!Q \rightarrow \eta \otimes_{!P|!Q}^E \rho.$$

And a simple verification proves that

$$(\eta \otimes_Q^E \rho)_{!(P|Q)} = \eta \otimes_{!P|!Q}^E \rho.$$

**Rules of group 6:** These rules are a direct consequence of (Alpha). □

The next theorem, stating that stochastic bisimulation is a congruence, proves that we have identified a well-behaved semantics.

**Theorem 5.3.4 (Congruence).** *If  $P \sim Q$ , then*

1. for any  $a, b \in \mathcal{N}$ ,  $a(b).P \sim a(b).Q$  and  $a[b].P \sim a[b].Q$ ;
2. for any  $R \in \mathbb{P}$ ,  $P + R \sim Q + R$ ;
3. for any  $a \in \mathcal{N}$  and  $r \in \mathbb{Q}^+$ ,  $(a@r)P \sim (a@r)Q$ ;
4. for any  $R \in \mathbb{P}$ ,  $P|R \sim Q|R$ ;
5.  $!P \sim !Q$ .

*Proof.* **1. Prefix:** For any  $C \in \Pi(\sim)$ ,  $P \in C$  iff  $Q \in C$ . This entails that for any  $E \in \mathbb{E}$  with  $fn(x.P) \cup fn(x.Q) \subseteq dom(E)$  and any  $\alpha \in \mathbb{A}^+$ ,

$$E_{P \sim}^x(\alpha)(C) = E_{Q \sim}^x(\alpha)(C).$$

**2. Choice:** We can suppose, without loosing generality, that

$$E \vdash P \rightarrow \mu, E \vdash Q \rightarrow \eta \text{ and } E \vdash R \rightarrow \rho,$$

since the other cases are trivially true. Then,

$$E \vdash P + R \rightarrow \mu \oplus \rho \text{ and } E \vdash Q + R \rightarrow \eta \oplus \rho.$$

Let  $C \in \Pi(\sim)$  and  $\alpha \in \mathbb{A}^+$ . Because  $P \sim Q$ ,  $\mu(\alpha)(C) = \eta(\alpha)(C)$  implying

$$\mu(\alpha)(C) + \rho(\alpha)(C) = \eta(\alpha)(C) + \rho(\alpha)(C).$$

This means that

$$(\mu \oplus \rho)(\alpha)(C) = (\eta \oplus \rho)(\alpha)(C).$$

**3. Fresh name quantification:** Let  $E \in \mathbb{E}$  and  $b \notin dom(E) \cup fn(P) \cup fn(Q)$ . Observe that from  $P \sim Q$ , following an observation that we used also in the proof of Lemma

5.2.13 concerning the relation between a mapping  $\eta$  its correspondent  $\eta_{\{b/a\}}$ , we derive  $P_{\{b/a\}} \sim Q_{\{b/a\}}$ . Suppose that

$$E, b@r \vdash P_{\{b/a\}} \rightarrow \mu \text{ and } E, b@r \vdash Q_{\{b/a\}} \rightarrow \eta.$$

Applying (New) we obtain  $E \vdash (b@r)P_{\{b/a\}} \rightarrow (b@r)\mu$  and  $E \vdash (b@r)Q_{\{b/a\}} \rightarrow (b@r)\eta$ .

(Alpha) implies  $E \vdash (a@r)P \rightarrow (b@r)\mu$  and  $E \vdash (a@r)Q \rightarrow (b@r)\eta$ .

From  $P_{\{b/a\}} \sim Q_{\{b/a\}}$  we obtain that for any  $\alpha \in \mathbb{A}^+$  and any  $C \in \Pi(\sim)$ ,  $\mu(\alpha)(C) = \eta(\alpha)(C)$ . to conclude the proof it is sufficient to verify that  $(b@r)\mu(\alpha)(C) = (b@r)\eta(\alpha)(C)$ .

**4. Parallel composition:** For the beginning we consider the processes that, to all syntactic levels, contain no subprocess from the class  $0^{\equiv}$  in a parallel composition. Let's call them *processes with non-trivial forms*.

We will firstly prove the case of the processes with non-trivial forms.

For arbitrary  $n \in \mathbb{N}$ , let  $S^n$  be the set of process terms with non-trivial forms and no more than  $n$  occurrences of the operator “|”.

Let  $\sim^n \subseteq S^n \times S^n$  be the largest rate-bisimulation defined on  $S^n$ .

We define  $\approx^n \in S^n \times S^n$  by

$$\approx^n = \sim^{n-1} \cup$$

$$\{(P_1 | \dots | P_k, Q_1 | \dots | Q_k), (P_1 + \dots + P_k, Q_1 + \dots + Q_k) \text{ for } P_i \sim^{n-1} Q_i, i = 1..k, k \leq n\}.$$

We show, by induction on  $n$ , that  $\approx^n$  is a rate-bisimulation, i.e. that  $\approx^n \subseteq \sim^n$ .

Suppose that  $P \approx^n Q$ . We need to prove that if  $E \vdash P \rightarrow \mu$  and  $E \vdash Q \rightarrow \eta$ , then for any  $\alpha \in \mathbb{A}^+$  and any  $C \in \Pi(\approx^n)$ ,  $\mu(\alpha)(C) = \eta(\alpha)(C)$ .

Observe that, from the way we construct  $\approx^n$ , there are three possibilities: either  $[P \sim^{n-1} Q]$ , or  $[P = P_1 + \dots + P_k \text{ and } Q = Q_1 + \dots + Q_k]$ , or  $[P = P_1 | \dots | P_k \text{ and } Q = Q_1 | \dots | Q_k]$ , for  $k \leq n$ , with  $P_i \sim^{n-1} Q_i$  for each  $i = 1..k$ .

In the first two cases, using also the case of choice operator that we have already proved, it is trivial to verify that  $\mu(\alpha)(C) = \eta(\alpha)(C)$ .

To prove the last case observe for the beginning that because  $\sim^{n-1} \subseteq \sim^n$ , the inductive hypothesis guarantees that for each  $i = 1..k$ ,

$$P_1 | \dots | P_{i-1} | P_{i+1} | \dots | P_k \approx^{n-1} Q_1 | \dots | Q_{i-1} | Q_{i+1} | \dots | Q_k$$

and consequently that

$$P_1 | \dots | P_{i-1} | P_{i+1} | \dots | P_k \sim^{n-1} Q_1 | \dots | Q_{i-1} | Q_{i+1} | \dots | Q_k.$$

Suppose that  $E \vdash P_i \rightarrow \mu_i$  and  $E \vdash Q_i \rightarrow \eta_i$  for all  $i = 1..k$ . Then,

$$\mu = \mu_1 \otimes_{P_1} \otimes_{P_2} \dots \otimes_{P_k}^E (\mu_2 \otimes_{P_2} \otimes_{P_3} \dots \otimes_{P_k}^E (\dots (\mu_{k-1} \otimes_{P_{k-1}} \otimes_{P_k}^E \mu_k) \dots)),$$

$$\eta = \eta_1 \otimes_{Q_1} \otimes_{Q_2} \dots \otimes_{Q_k}^E (\eta_2 \otimes_{Q_2} \otimes_{Q_3} \dots \otimes_{Q_k}^E (\dots (\eta_{k-1} \otimes_{Q_{k-1}} \otimes_{Q_k}^E \eta_k) \dots)),$$

Consider an arbitrary  $\alpha \in \mathbb{A}$ . Then,

$$\mu(\alpha)(C) = \sum_{i=1..k} \mu_i(\alpha)(C_{P_1 | \dots | P_{i-1} | P_{i+1} | \dots | P_k}),$$

$$\eta(\alpha)(C) = \sum_{i=1..k} \eta_i(\alpha)(C_{Q_1|\dots|Q_{i-1}|Q_{i+1}|\dots|Q_k}).$$

Because  $C \in \Pi(\approx^n)$ ,  $C_{P_1|\dots|P_{i-1}|P_{i+1}|\dots|P_k}$  and  $C_{Q_1|\dots|Q_{i-1}|Q_{i+1}|\dots|Q_k}$  contain only processes with at most  $n - 1$  occurrences of  $|$ , for any  $i$ . And because  $P_1|\dots|P_{i-1}|P_{i+1}|\dots|P_k \sim^{n-1} Q_1|\dots|Q_{i-1}|Q_{i+1}|\dots|Q_k$ , we obtain

$$C_{P_1|\dots|P_{i-1}|P_{i+1}|\dots|P_k} = C_{Q_1|\dots|Q_{i-1}|Q_{i+1}|\dots|Q_k} \in \Pi(\sim^{n-1}).$$

Further, using the fact that  $\sim^{n-1}$  is a rate bisimulation, we obtain

$$\mu(\alpha)(C_{P_1|\dots|P_{i-1}|P_{i+1}|\dots|P_k}) = \eta(\alpha)(C_{Q_1|\dots|Q_{i-1}|Q_{i+1}|\dots|Q_k})$$

that implies  $\mu(\alpha)(C) = \eta(\alpha)(C)$ .

A similar argument proves the case  $\alpha = \tau$ . Consequently,  $\approx^n$  is a rate-bisimulation.

Returning to our theorem, suppose that  $P$  and  $Q$  are two processes with non-trivial forms such that  $P \sim Q$ . Then, there exists  $n \in \mathbb{N}$  such that  $P \sim^n Q$ .

Suppose that  $R \in \mathbb{S}^m$  for some  $m \in \mathbb{N}$ . Then  $P \sim^{m+n-1} Q$  and  $R \sim^{m+n-1} R$  implying  $P|R \approx^{m+n} Q|R$ . Because  $\approx^{m+n}$  is a rate-bisimulation, we obtain that  $P|R \sim Q|R$ .

If  $P$ ,  $Q$  or  $R$  (or some of them) have “trivial forms”, then there exist  $P' \equiv P$ ,  $Q' \equiv Q$  and  $R' \equiv R$  with non-trivial forms. And because the bisimulation is an equivalence that extends the structural congruence, we obtain the desired result also for the general case.

**5. Replication:** We use the same proof strategy as for the parallel composition. We say that a process is in canonic form if it contains no parallel composition of replicated subprocesses and no replicated process from the class  $0^\equiv$ . In other words,  $!(P|Q)$  is in canonic form while  $!P|!Q$  and  $!(P|Q)|!0$  are not; using the structural congruence rules, we can associate to each process  $P$  a structural congruent process with a canonic form called a canonic representative for  $P$ . Notice also that all the canonic representatives of a given process have the same number of occurrences of the operator “!”. Let  $\mathbb{S}_*$  be the set of process terms with canonic form. Observe that because structural congruence is a subset of bisimulation, it is sufficient to prove our lemma only for processes in  $\mathbb{S}_*$ .

As before, let  $\mathbb{S}_*^n$  be the set of processes (in canonic form) with no more than  $n$  occurrences of the operator “!”. Let  $\sim^n$  be the stochastic bisimulation on  $\mathbb{S}_*^n$  and  $\approx^n \subseteq \mathbb{S}_*^n \times \mathbb{S}_*^n$  defined by

$$\approx^n = \sim^{n-1} \cup \{(!P, !Q) \mid P \sim^{n-1} Q\}.$$

We firstly show, inductively on  $n$ , that  $\approx^n$  is a rate-bisimulation. Consider two arbitrary processes  $P$  and  $Q$  such that  $P \approx^n Q$ . We prove that if  $E \vdash P \rightarrow \mu$  and  $E \vdash Q \rightarrow \eta$ , then for arbitrary  $\alpha \in \mathbb{A}^+$  and  $C \in \Pi(\approx^n)$ ,  $\mu(\alpha)(C) = \eta(\alpha)(C)$ .

Observe that if  $P \approx^n Q$ , then either  $P \sim^{n-1} Q$ , or  $P \equiv !R$  and  $Q \equiv !S$  with  $R \sim^{n-1} S$ . In the first case the equality is trivially true. In the other case, suppose that  $E \vdash R \rightarrow \mu'$  and  $E \vdash S \rightarrow \eta'$ . Then,  $\mu = \mu'_{!R}$  and  $\eta = \eta'_{!S}$ . We have

$$\mu(\alpha)(C) = \mu'(\alpha)(C_{!R}), \quad \eta(\alpha)(C) = \eta'(\alpha)(C_{!S}).$$

We prove that  $C_{!R} = C_{!S}$ . Let  $U \in C_{!R}$ . Then,  $U|!R \in C$  and from the construction of  $C \in \Pi(\approx^n)$ , we obtain that there exists  $T \in \mathbb{S}_*^{n-1}$  such that  $U = !T$ . Because  $!R|!T \in C$ ,  $!(R|T) \in C$ . Now, from  $R \sim^{n-1} S$  we obtain  $R \sim S$  and because  $T \sim T$ , the case of parallel operator that we have proved guarantees that  $R|T \sim S|T$ . But the canonic representatives  $V, W$  of  $R|T$  and  $S|T$  respectively are in  $\mathbb{S}_*^{n-1}$  meaning that  $V \sim^{n-1} W$ . The construction of  $\approx^n$  guarantees further that  $!V \approx^n !W$  and because  $W \equiv S|T$  we obtain  $!(S|T) \in C$  and  $U \equiv !T \in C_{!S}$ .

Because  $C_{!R} = C_{!S}$  and  $\mu'(\alpha)(C_{!R}) = \eta'(\alpha)(C_{!S})$  (this is implied by  $R \sim^{n-1} S$ ), then  $\mu(\alpha)(C) = \eta(\alpha)(C)$ .  $\square$

## 5.4 Concluding remarks

In this chapter we have proposed a way of introducing stochastic Pi-calculus that is faithful to the algebraic-coalgebraic structures of the concurrent Markovian processes. This work extends the one of the previous chapter where we proposed a measure-theoretical operational semantics for a finite fragment of stochastic CCS. In this chapter we have demonstrated that the semantics can be successfully extended to include name passing, new name operators and replication.

As before, the semantics is given in terms of measure theory and describes the lifting of the algebraic structure of processes to the level of measures on the measurable space of processes. Instead of the discrete measurable space of processes, we consider the measurable space induced by structural congruence and this idea has important advantages. Firstly, it matches practical modelling requirements: the identity of a system is not given by the stochastic process used to model it, but by its structural-congruence class (for systems biology this represents a chemical soup). Secondly, by working with measures on this space, we get important advantages on the level of the underlying theory such as a simple and elegant semantics, simple solutions for the problems related to bound output and replication (that otherwise require complicate transition labeling and higher order reasoning) and a well-behaved notion of stochastic bisimulation. Other advantages derive from the use of the rate environments that guarantees a certain robustness in modelling: a model can be easily refined by modifying its rate environment.

Our approach opens some future research directions. One is the study of the GSOS format where the main challenges are to understand the underlying category and the equational monad induced by structural congruence.



# Chapter 6

## Metrics for Stochastic Process Algebras

### 6.1 Introduction

In the case of stochastic and probabilistic systems, bisimulation is a strict concept: it verifies whether two processes have identical behaviours. In applications we need more. For instance, we want to know whether two processes that may differ by only a small amount in real-valued parameters (rates or probabilities) are behaving in a similar way.

To address this problem we define in this chapter some pseudometrics on the set of stochastic process algebra terms that will measure how much two processes are alike in terms of their behaviours. In this sense, two processes are at distance zero iff they are bisimilar. Thus, the pseudometrics will be quantitative extensions of the notion of bisimulation.

We will argue that the operational semantics based on measure theory is particularly appropriate for this type of development. This chapter is, in this sense, a “proof of concept” and for this reason the entire development is made only for the class of stochastic CCS processes introduced in Chapter [43]. However, a similar development can be adapted to stochastic Pi-calculus.

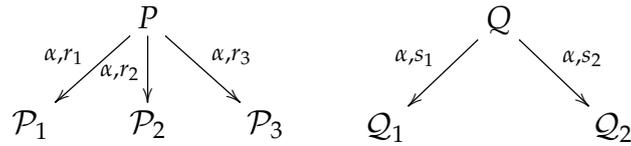
Metrics for measuring the similarity of probabilistic systems in terms of behaviours have been proposed in [22, 42, 54]. In [25, 46] such metrics are introduced using a set of functional expressions, that generalise formulas of Hennessy-Milner logic, in a similar way to which Kantorovich metrics are defined by Lipschitz functions. These metrics are designed for general measurable spaces but the transitions are in discrete time. However, they have been extended in [33] for continuous-time systems (generalized semi-Markov processes).

The metrics developed in this chapter are similar to Desharnais-Panangaden metrics, for instance in the way they explore the transition systems, but they are simpler, being particularly designed for our process algebra. We do not consider any functional expressions for calculating the distance, but we propose a direct approach. In the last part of this monograph we will return to these metrics and study their computational aspects in a more extended settings.

## 6.2 Distances Between Stochastic Process Algebra Terms

The behaviours of stochastic processes can be compared from two main points of view: the immediate transition rates and their future behaviour. The metrics that we propose in this chapter take both aspects into account. For this reason, our metrics  $d^c : \mathbb{P} \times \mathbb{P} \rightarrow \mathbb{R}^+$  are indexed with the parameter  $c \in [0, 1]$ , where the notation is the one proposed in Chapter 4 and  $\mathbb{P}$  represents the set of stochastic CCS processes. The pseudometric  $d^1$  captures only the differences between the transition rates of processes, giving equal importance to the differences between the immediate transitions and the differences that arise deeper in the evolution of the processes. On the other hand, a pseudometric  $d^c$  with  $c \in (0, 1)$  gives more weight to the rate differences that arise earlier in the evolution of the processes; as  $c$  approaches 0, the future gets discounted more, being completely ignored for  $c = 0$ .

The intuition behind these definitions is as follows. Assume we want to measure the distance between the processes  $P$  and  $Q$  that have the immediate transitions as represented below, where  $\mathcal{P}_i, \mathcal{Q}_i \in \Pi(\sim)$  for  $i = 1..3$  are bisimulation classes and the transitions are all labelled by  $\alpha \in \mathbb{A}^+$ .



For calculating the distance  $d^c(P, Q)$ , we firstly pair classes  $\mathcal{P}_i$  and  $\mathcal{Q}_j$  and then sum the differences between the rates of going from  $P$  and  $Q$  to  $\mathcal{P}_i$  and  $\mathcal{Q}_j$ , respectively, and the weighted distance between arbitrary processes  $P_i \in \mathcal{P}_i$  and  $Q_j \in \mathcal{Q}_j$ . We thus obtain, for the pair  $(\mathcal{P}_i, \mathcal{Q}_j)$ , the value

$$|r_i - s_j| + c \cdot d^c(P_i, Q_j).$$

There are various ways in which one can take these pairs:  $d^c$  is the infimum of the values one can get taking all possible pairings of bisimulation classes.

However, these pairings have to be one-to-one and onto on  $\mathbb{P}^\sim$  and for this reason we will use the possible bijections on  $\mathbb{P}^\sim$ .

Another observation is that we only need to consider the pairs  $(\mathcal{P}_i, \mathcal{Q}_j)$ , such that either  $P$  can do an  $\alpha$ -transition to  $\mathcal{P}_i$  with non-zero rate, or  $Q$  can do an  $\alpha$ -transition to  $\mathcal{Q}_j$  with non-zero rate.

In what follows we formalize these intuitions for the class of stochastic CCS-processes; the notations in this chapter are the ones introduced in Chapter 4.

We propose two families of pseudometrics on  $\mathbb{P}$ :  $\mathbb{D}_\alpha$  for  $\alpha \in \mathbb{A}^+$  and  $\mathbb{D}$ . The first family contains measures that concern only  $\alpha$ -transitions, while the second considers all the transitions.

As before, for arbitrary  $P, Q \in \mathbb{P}$ , we write  $P \Longrightarrow Q$  if there exists  $\alpha \in \mathbb{A}^+$  and  $r \neq 0$  such that  $P \xrightarrow{\alpha, r} Q^\equiv$ . Let

$$\mathcal{D}(P) = \bigcup \{Q^\sim \mid P \Longrightarrow Q\}$$

be the set of derivatives of  $P$ .

Let  $\mathcal{B}$  be the set of bijections  $\sigma : \mathbb{P}^\sim \rightarrow \mathbb{P}^\sim$ .

For arbitrary  $P, Q, R, S \in \mathbb{P}$  and  $\sigma \in \mathcal{B}$  we write  $R[\sigma_Q^P]S$  if  $R \in \mathcal{D}(P) \cup \sigma^{-1}(\mathcal{D}(Q))$  and  $S^\sim = \sigma(R^\sim)$ .

**Definition 6.2.1.** For arbitrary  $\alpha \in \mathbb{A}^+$  consider the family  $\mathbb{D}_\alpha$  of functions

$$d_\alpha^c : \mathbb{P} \times \mathbb{P} \rightarrow \mathbb{R}^+, \quad c \in [0, 1],$$

defined for  $P, Q \in \mathbb{P}$  with  $P \rightarrow \mu$  and  $Q \rightarrow \eta$ , by

$$d_\alpha^c(P, Q) = \inf_{\sigma \in \mathcal{B}} \{\sigma_\alpha(P, Q)\},$$

where<sup>1</sup>

$$\sigma_\alpha(P, Q) = \sum_{(R^\sim, S^\sim)}^{R[\sigma_Q^P]S} (|\mu(\alpha)(R^\sim) - \eta(\alpha)(S^\sim)| + c \cdot d_\alpha^c(R, S)).$$

The correctness of this definition derives from Lemma 4.4.5 and from the fact that the transition tree of a derivative of a process  $P$  is strictly less complex than the transition tree of  $P$ . The same arguments guarantee that the infimum considered before is well defined.

The parameter  $c \in [0, 1]$  is used to associate a weight with each transition step. For instance if  $a \in \mathbb{A}$ , then

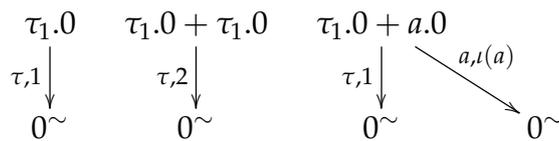
$$d_\tau^c(\tau_1.0, \tau_1.0 + \tau_1.0) = |1 - 2| = 1$$

because the first process is doing a  $\tau$ -transition with rate 1 and the second with rate 2. Similarly,

$$d_\tau^c(\tau_1.0, \tau_1.0 + a.0) = |1 - 1| = 0$$

because both processes are doing  $\tau$ -transitions with rate 1 and for similar reasons,

$$d_\tau^c(\tau_1.0 + \tau_1.0, \tau_1.0 + a.0) = |2 - 1| = 1.$$



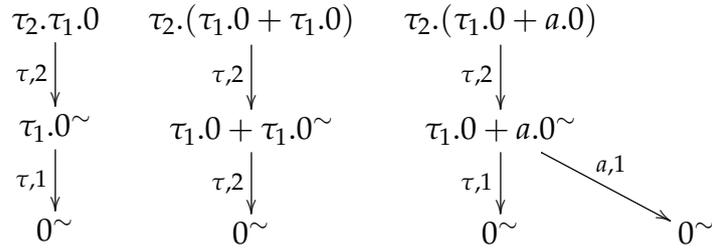
<sup>1</sup>The sum is for all pairs  $(R^\sim, S^\sim)$  such that  $R[\sigma_Q^P]S$ .

If we prefix these three processes, we will see their difference only at the second level transitions and this will influence the measure. Thus,

$$d_\tau^c(\tau_2.\tau_1.0, \tau_2.(\tau_1.0 + \tau_1.0)) = |2 - 2| + c \cdot |2 - 1| = c,$$

$$d_\tau^c(\tau_2.\tau_1.0, \tau_2.(\tau_1.0 + a.0)) = |2 - 2| + c \cdot |1 - 1| = 0,$$

$$d_\tau^c(\tau_2.(\tau_1.0 + \tau_1.0), \tau_2.(\tau_1.0 + a.0)) = |2 - 2| + c \cdot |2 - 1| = c.$$



We can use various values of  $c \in [0, 1]$  to give a certain weight to each transition step. Thus  $d_a^1$  gives equal importance to the differences at each transition step, while  $d_a^0$  is the measure that only looks to the immediate transitions.

Consider now the processes  $\tau_3.0$ ,  $\tau_2.\tau_1.0$ ,  $\tau_1.\tau_2.0$  and  $\tau_1.\tau_1.\tau_1.0$  represented below. Their relative distances are:

$$d_\tau^c(\tau_3.0, \tau_2.\tau_1.0) = |3 - 2| + c \cdot |1 - 0| = 1 + c,$$

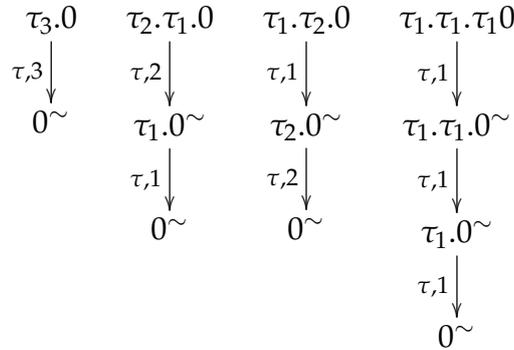
$$d_\tau^c(\tau_3.0, \tau_1.\tau_2.0) = |3 - 1| + c \cdot |2 - 0| = 2 + 2c,$$

$$d_\tau^c(\tau_1.\tau_2.0, \tau_2.\tau_1.0) = |1 - 2| + c \cdot |2 - 1| = 1 + c,$$

$$d_\tau^c(\tau_3.0, \tau_1.\tau_1.\tau_1.0) = |3 - 1| + c \cdot |1 - 0| + c^2 \cdot |1 - 0| = 2 + c + c^2,$$

$$d_\tau^c(\tau_2.\tau_1.0, \tau_1.\tau_1.\tau_1.0) = |2 - 1| + c \cdot |1 - 1| + c^2 \cdot |1 - 0| = 1 + c^2,$$

$$d_\tau^c(\tau_1.\tau_2.0, \tau_1.\tau_1.\tau_1.0) = |1 - 1| + c \cdot |2 - 1| + c^2 \cdot |1 - 0| = c + c^2$$



Notice also that for  $a \in \mathbb{A}$  the values of  $d_a^c$  are of type  $k_0 + k_1 \cdot c + k_2 \cdot c^2 + \dots$  where  $k_i$  are multiples of  $\iota(a)$ . This is not particularly significant, as our main issue is not the absolute value of the metric, but the significance of zero distance or the relative distance of processes (i.e., their induced topology).

The next lemma states that, indeed, our functions are pseudometrics.

**Lemma 6.2.2.** *For any  $c \in [0, 1]$  and any  $\alpha \in \mathbb{A}^+$ ,  $d_\alpha^c$  is a pseudometric on  $\mathbb{P}$ .*

*Proof.* The only non-trivial axiom to verify is the triangle inequality:

$$d_\alpha^c(P, R) \leq d_\alpha^c(P, Q) + d_\alpha^c(Q, R).$$

We prove it by induction on the structures of processes.

We can assume, without loosing generality, that there exist the processes  $P_1, \dots, P_n, Q_1, \dots, Q_n$  and  $R_1, \dots, R_n$  such that

$$\mathcal{D}(P) \subseteq \{P_1, \dots, P_n\}, \mathcal{D}(Q) \subseteq \{Q_1, \dots, Q_n\} \text{ and } \mathcal{D}(R) \subseteq \{R_1, \dots, R_n\}$$

and  $\sigma', \sigma'' \in \mathcal{B}$  such that

$$d_\alpha^c(P, Q) = \sigma'_\alpha(P, Q), d_\alpha^c(Q, R) = \sigma''_\alpha(Q, R)$$

and for each  $i = 1..n$ ,  $Q_i^\sim = \sigma'(P_i^\sim)$ ,  $R_i^\sim = \sigma''(Q_i^\sim)$ .

Suppose also that  $P \rightarrow \mu$ ,  $Q \rightarrow \mu'$  and  $R \rightarrow \mu''$ . Then,  $\sigma''' = \sigma'' \circ \sigma' \in \mathcal{B}$ . Consequently,

$$d_\alpha^c(P, R) \leq \sigma'''_\alpha(P, R) = \sum_{i=1..n} (|\mu(\alpha)(P_i^\sim) - \mu''(\alpha)(R_i^\sim)| + c \cdot d_\alpha^c(P_i, R_i)).$$

From the inductive hypothesis we obtain that for each  $i = 1..n$ ,

$$d_\alpha^c(P_i, R_i) \leq d_\alpha^c(P_i, Q_i) + d_\alpha^c(Q_i, R_i).$$

Moreover,

$$|\mu(\alpha)(P_i) - \mu''(\alpha)(R_i)| \leq |\mu(\alpha)(P_i) - \mu'(\alpha)(Q_i)| + |\mu'(\alpha)(Q_i) - \mu''(\alpha)(R_i)|.$$

Consequently,

$$\begin{aligned} & \sum_{i=1..n} (|\mu(\alpha)(P_i^\sim) - \mu''(\alpha)(R_i^\sim)| + c \cdot d_\alpha^c(P_i, R_i)) \leq \\ & \sum_{i=1..n} (|\mu(\alpha)(P_i^\sim) - \mu'(\alpha)(Q_i^\sim)| + c \cdot d_\alpha^c(P_i, Q_i)) + \\ & \sum_{i=1..n} (|\mu'(\alpha)(Q_i^\sim) - \mu''(\alpha)(R_i^\sim)| + c \cdot d_\alpha^c(Q_i, R_i)) = \\ & \sigma'_\alpha(P, Q) + \sigma''_\alpha(Q, R) = d_\alpha^c(P, Q) + d_\alpha^c(Q, R). \end{aligned}$$

Hence,  $d_\alpha^c(P, R) \leq d_\alpha^c(P, Q) + d_\alpha^c(Q, R)$ . □

It is not difficult to prove that the distance between bisimilar processes is always zero. Also, if for a fixed  $c \neq 0$  the distances  $d_\alpha^c$  between two given processes are zero for all  $\alpha \in \mathbb{A}^+$ , then the processes must be bisimilar.

**Theorem 6.2.3.** *Let  $P, Q \in \mathbb{P}$ .*

- (i) *If  $P \sim Q$ , then for any  $c \in [0, 1]$  and any  $\alpha \in \mathbb{A}^+$ ,  $d_\alpha^c(P, Q) = 0$ .*
- (ii) *If there exists  $c \in (0, 1]$ , such that for any  $\alpha \in \mathbb{A}^+$ ,  $d_\alpha^c(P, Q) = 0$ , then for any  $c' \in [0, 1]$  and any  $\alpha \in \mathbb{A}^+$ ,  $d_\alpha^{c'}(P, Q) = 0$ . Moreover, in this case  $P \sim Q$ .*

Notice that the elements of  $\mathbb{D}_\alpha$  measure only  $\alpha$ -transitions and for this reason their utility is limited. Our main intention is to introduce a metric on processes that can characterize the bisimulation. For achieving this goal, in what follows we will introduce a family of metrics which consider all the transitions. The intuition is that the “general” distance  $d^c$  between two processes is the supremum of the distances  $d_\alpha^c$  for all  $\alpha \in \mathbb{A}^+$ .

**Definition 6.2.4.** *Consider the family  $\mathbb{D}$  of functions*

$$d^c : \mathbb{P} \times \mathbb{P} \rightarrow \mathbb{R}^+, \quad c \in [0, 1],$$

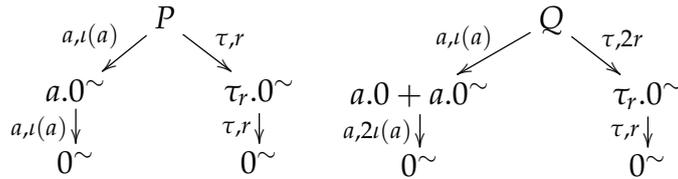
defined for arbitrary  $P', P'' \in \mathbb{P}$  by

$$d^c(P', P'') = \sup_{\alpha \in \mathbb{A}^+} \{d_\alpha^c(P', P'')\}.$$

Consider the processes

$$P = a.a.0 + \tau_r.\tau_r.0 \text{ and } Q = a.(a.0 + a.0) + (\tau_r.0|\tau_r.0)$$

represented below.



For calculating  $d^c(P, Q)$ , we firstly observe that

$$d_a^c(P, Q) = |\iota(a) - \iota(a)| + c \cdot |2\iota(a) - \iota(a)| = c \cdot \iota(a),$$

$$d_\tau^c(P, Q) = |2r - r| + c \cdot |r - r| = r$$

and for any  $\alpha \notin \{a, \tau\}$ ,  $d_\alpha^c(P, Q) = 0$ .

Consequently,  $d^c(P, Q) = \max\{c \cdot \iota(a), r\}$ .

**Lemma 6.2.5.** *For any  $c \in [0, 1]$ ,  $d^c$  is a pseudometric on  $\mathbb{P}$ .*

*Proof.* As for  $d_\alpha^c$ , the only non-trivial axiom is  $d^c(P, R) \leq d^c(P, Q) + d^c(Q, R)$ . From Lemma 6.2.2 we know that  $d_\alpha^c(P, R) \leq d_\alpha^c(P, Q) + d_\alpha^c(Q, R)$ . From here we obtain

$$\begin{aligned} d^c(P, R) &= \sup_{\alpha \in \mathbb{A}^+} d_\alpha^c(P, R) \leq \sup_{\alpha \in \mathbb{A}^+} (d_\alpha^c(P, Q) + d_\alpha^c(Q, R)) \leq \\ &\sup_{\alpha \in \mathbb{A}^+} d_\alpha^c(P, Q) + \sup_{\alpha \in \mathbb{A}^+} d_\alpha^c(Q, R) = d^c(P, Q) + d^c(Q, R). \end{aligned}$$

□

The next theorem states that, indeed, the pseudometrics  $d^c$  generalise the bisimulation of processes. Lifted on the level of bisimulation classes, the pseudometrics became metrics and consequently, they organize the space  $\mathbb{P}^\sim$  as a metric space.

**Theorem 6.2.6.** *Let  $P, Q \in \mathbb{P}$ .*

- (i) *If  $P \sim Q$ , then for any  $c \in [0, 1]$ ,  $d^c(P, Q) = 0$ .*
- (ii) *If for some  $c \in (0, 1]$ ,  $d^c(P, Q) = 0$ , then for any  $c' \in [0, 1]$ ,  $d^{c'}(P, Q) = 0$ . Moreover, in this case  $P \sim Q$ .*

For concluding this section, we notice that the metrics are influenced by the algebraic structure of the processes. The next lemma reveals such a relation for the case of prefixing. However, we believe that more complex relations can be identified and we intend to return to this problem in future works. The possibility of computing the distance between two processes from the relative distances of their sub-processes is an idea that can find interesting applications especially in the case of large systems where it is more convenient to focus on subsystems.

**Lemma 6.2.7.** *For arbitrary  $P, Q \in \mathbb{P}$  and  $\varepsilon \in \mathbb{A}^*$ ,*

$$\text{if } d^c(P, Q) = r > 0, \text{ then } d^c(\varepsilon.P, \varepsilon.Q) = \max\{2 \cdot \iota(\varepsilon), c \cdot r\}.$$

## 6.3 Concluding Remarks

In this final chapter of this part of our monograph we show how one can extend the equivalence-based reasoning on processes to metric reasoning. We defined quantitative extensions of stochastic bisimulation in the form of two classes of metrics that measure the distance between processes in terms of similar behaviours: two processes are at distance zero iff they are bisimilar; two processes are close if their behaviours are similar.

The organisation of the space of algebraic processes as a metric space was a novelty when [17] was published. This work can be extended to other calculi and used in applications, for example, to appreciate the quality of approximations of models or to characterise quantitatively the concept of robustness.

In the Part IV of this monograph we will return to this idea and develop it for general Markov processes up to the level of proving its role into the Stone duality for Markov processes. In Part V we will investigate computability and complexity problems related to such behaviours distances.



# Bibliography

- [1] G. Bacci, G. Bacci, K.G. Larsen, R. Mardare. *Computing Behavioral Distances, Compositionally*, In Proc. MFCS2013:74-85, 2013.
- [2] G. Bacci, G. Bacci, K.G. Larsen, R. Mardare. *The BisimDist Library: Efficient Computation of Bisimilarity Distances for Markovian Models*, In Proc. QEST2013:278-281, 2013.
- [3] G. Bacci, G. Bacci, K.G. Larsen, R. Mardare. *On-the-Fly Exact Computation of Bisimilarity Distances*, In Proc. TACAS2013:1-15, 2013.
- [4] G. Bacci, M. Miculan. *Measurable stochastics for Brane Calculus*. TCS.431: 117-136, 2012.
- [5] G. Bacci, M. Miculan. *Structural Operational Semantics for Continuous State Probabilistic Processes*. In Proc. CMCS2012, LNCS 7399: 71-89, 2012.
- [6] G. Bacci, M. Miculan. *Structural Operational Semantics for Continuous State Probabilistic Processes*. J. Comput. Syst. Sci., 2015 (in press).
- [7] G. Berry, G. Boudol. *The Chemical Abstract Machine*, In Proc. POPL 1990:81-94.
- [8] J.A. Bergstra et al (Eds.) *Handbook of Process Algebra*. Elsevier, 2001.
- [9] M. Bernardo, R. Gorrieri. *A tutorial on EMPA: A theory of concurrent processes with nondeterminism, priorities, probabilities and time*. TCS 202(1-2):1-54, 1998.
- [10] R. Blute et al. *Bisimulation for labeled Markov processes*. In Proc. LICS'97, IEEE Press, 1997.
- [11] M. Bravetti et al. *YMCA: Why Markov Chain Algebra?*, ENTCS 162:107-112, 2006.
- [12] M. Calder et al. *Automatically deriving ODEs from process algebra models of signaling pathways*. In Proc. CMSB'05, 2005.
- [13] L. Cardelli, *A Process Algebra Master Equation*. In Proc. QEST'07, 2007.
- [14] L. Cardelli, *Bitonal membrane systems: Interactions of biological membranes*, Theoretical Computer Science, 404(1-2): 5-18, 2008

- [15] L. Cardelli, A. Gordon, *Mobile ambients*, Theoretical Computer Science, 240-1:177–213, 2000
- [16] L. Cardelli, R. Mardare. Stochastic Pi-Calculus Revisited, In Proc. ICTAC2013, LNCS 8049, 2013.
- [17] L. Cardelli, R. Mardare. *The Measurable Space of Stochastic Processes*. In Proc. QEST'10 2010:171-180, IEEE Press.
- [18] L. Cardelli, R. Mardare. *The Measurable Space of Stochastic Processes*, Fundamenta Informaticae 131(3-4):351-371, 2014.
- [19] P. R. D'Argenio, D. Gebler, M. D. Lee. *Axiomatizing Bisimulation Equivalences and Metrics from Probabilistic SOS Rules*, Proc. FoSSaCS 2014, LNCS 8412 pp. 289-303.
- [20] P. R. D'Argenio, M. D. Lee. *Probabilistic Transition Systems Specifications: Congruence and Full Abstraction of Bisimulation*, Proc. FoSSaCS 2012, LNCS 7213, pp. 452-466, 2012.
- [21] V. Danos et al. *Bisimulation and Cocongruence for Probabilistic Systems*, Inf. and Comp. 204(4):503-523, 2006.
- [22] L. de Alfaro et al. *Discounting the future in systems theory*. LNCS 2719:1022-1037, 2003.
- [23] R. De Nicola et al. *Rate-Based Transition Systems for Stochastic Process Calculi*. LNCS 5556:435-446, 2009
- [24] E.P. de Vink, J. Rutten, *Bisimulation for probabilistic transition systems: A coalgebraic approach*. TCS 221(1-2):271-293, 1999.
- [25] J. Desharnais, *Labelled Markov Processes*. PhD thesis, McGill University, 1999.
- [26] J. Desharnais et al. *A logical characterization of bisimulation for labeled Markov processes*. In Proc LICS'98, IEEE Press, 1998.
- [27] J. Desharnais et al. *Bisimulation for labeled Markov Processes*. Inf. and Comp., 179(2):163-193, 2002.
- [28] J. Desharnais, P. Panangaden. *Continuous Stochastic Logic Characterizes Bisimulation of Continuous-time Markov Processes*. JLAP. 56:99-115, 2003.
- [29] E. Doberkat. *Stochastic Relations. Foundations for Markov Transition Systems*. Chapman and Hall/CRC, 2007.
- [30] M.P.Fiore, D.Turi. *Semantics of name and value passing*. LICS'01, IEEE Press, 2001.

- [31] M.P.Fiore, S.Staton. *A congruence rule format for name-passing process calculi*. Inf. and Comp., 207(2), 2009.
- [32] N. Gotz et al. *TIPP - A language for timed processes and performance evaluation*. Tech.Rep. 4/92 IMMD VII, University of Erlangen-Nurnberg.
- [33] V. Gupta et al. *Approximate reasoning for real-time probabilistic processes*. LMCS 2(1:4) 2006.
- [34] J. Harsanyi. *Games with incomplete information played by bayesian players, part one*. Management Sci., 14:159-182, 1967.
- [35] J. Hillston. *A compositional approach to performance modelling*. Cambridge University Press, 1996.
- [36] J. Hillston. *Process algebras for quantitative analysis*. In Proc. LICS'05, IEEE Press, 2005.
- [37] H. Hermanns. *Interactive Markov Chains*. LNCS 2428, 2008.
- [38] B. Klin, V. Sassone. *Structural Operational Semantics for Stochastic Process Calculi*. LNCS 4968:428-442, 2008.
- [39] M. Kwiatkowska et al. *Automatic Verification of Real-Time Systems With Discrete Probability Distributions*. LNCS 1601:79-95, 1999.
- [40] K. G. Larsen and A. Skou. *Bisimulation through probabilistic testing*. Inf. and Comp., 94:1-28, 1991.
- [41] M.D. Lee, D. Gebler, P.R. D'Argenio. *Tree rules in probabilistic transition system specifications with negative and quantitative premises*, Proc. EXPRESS/SOS'12, EPTCS 89, pp. 115-130, 2012.
- [42] P. D. Lincoln et al. *A probabilistic poly-time framework for protocol analysis*. ACM(CCS-5), 1998.
- [43] R. Milner. *A calculus of communicating systems*. J. of ACM, 1980.
- [44] R. Milner, *Communicating and Mobile Systems: the Pi-Calculus*, Cambridge Univ. Press, 1999.
- [45] L.S. Moss, I.D. Viglizzo. *Harsanyi type spaces and final coalgebras constructed from satisfied theories*. ENTCS 106:279-295, 2004.
- [46] P. Panangaden. *Labelled Markov Processes*. Imperial College Press, 2009.
- [47] C. Priami. *Stochastic  $\pi$ -Calculus*. Computer Journal, 38(7):578-589, 1995.

- [48] J. Rutten. *Universal coalgebra: a theory of systems*, TCS, 249:3-80, 2000.
- [49] R. Segala, N. Lynch. *Probabilistic Simulations for Probabilistic Processes*, Nordic J. of Comp., 2(2):250-273, 1995.
- [50] S. Tini. *Non-expansive epsilon-bisimulations for Probabilistic Processes*, Theor. Comput. Sci. 441(22-24) pp. 2002-2222, 2010.
- [51] D. Turi, G.D. Plotkin. *Towards a mathematical operational semantics*, In Proc. Lics'97, IEEE Press, 1997.
- [52] A.M. Turing. *Computing Machinery and Intelligence*, Mind 59, pp 433-460, 1950
- [53] A.M. Turing. *The Chemical Basis of Morphogenesis*, Phil. Trans. R. Soc. London B 237 pp 37-72, 1952
- [54] F. van Breugel, J. Worrell. *An algorithm for quantitative verification of probabilistic systems*. LNCS 2154:336-350, 2001.
- [55] R. J. van Glabbeek et al. *Reactive, Generative and Stratified Models of Probabilistic Processes*. Inf. and Comp. 121(1):59-80, 1995.

**Part III**  
**Markovian Logics**



In the previous Part of this monograph we have looked at a particular subclass of Markov processes – the processes that can be defined inductively by terms of stochastic process algebras. Starting with this part of the monograph, we extend our research to the most general classes of Markov processes as defined by Pannagaden et al. [14–20, 22, 38]. These classes of processes are defined over arbitrary analytic spaces of states by measurable mappings from states to distributions over the state space. We can thus encode (i) *probabilistic systems* by considering the probabilistic distributions; (ii) *subprobabilistic systems* by taking the subprobabilistic distributions; and also the *stochastic-Markovian systems* by considering the general distributions over the state space (typically used to encode continuous-time Markov processes).

Properties of these classes of MPs can be specified by a few probabilistic or stochastic modal logics that we call in this monograph Markovian Logics (MLs) [4, 10, 15, 22, 27, 34, 35, 49]. They are multi-modal logics that, in addition to the classic Boolean operators, are devised with modalities indexed by rationals, which approximate, from below and from above, the probabilities or the rates of the MP-transitions.

For the probabilistic case, it has been proven that this logic is sufficiently expressive to characterize probabilistic bisimulation of MPs, even in the case when the state space is not finite (hence not image-finite) [14, 15]. This result was surprising, since for the nondeterministic transition system the Hennessy-Milner property can only be demonstrated under the finite-imaginess assumption. Moreover, one can even strip down the probabilistic Markovian logic to a very spartan core—just the modalities and finite conjunction—and still characterize bisimulation for labeled probabilistic Markov processes [14, 15].

Similar logics can be defined also for the classes of subprobabilistic or general stochastic Markov processes. It is therefore tempting to understand if all these logics enjoy similar properties.

The Markovian logics are also challenging from a model theoretic perspective. Goldblatt in [26] presents a proof-theoretic analysis of the class of the logics of  $T$ -coalgebras, where  $T$  is any polynomial functor constructed from a standard monad on the category of measurable spaces. He proves that the semantic consequence relation over  $T$ -coalgebras is equal to the least deducibility relation that satisfies Lindenbaum’s lemma, which states that any consistent set of formulas can be extended to a maximally consistent set. In other words, this consequence relation is equal to the least of all deducibility relations if and only if that least deducibility relation satisfies Lindenbaum’s lemma. These logics are not compact and for proving the aforementioned results in [26] it is used a powerful infinitary axiom scheme named the Countable Additivity Rule (CAR). However, a feature of CAR is that it has an uncountable set of instances and this fact makes it difficult to prove that maximally consistent sets exist for such logics. Moreover, in [51] it has been proved that such a logic is not strongly-complete in the absence of CAR.

The Markovian logics are particular cases of the logics studied by Goldblatt and

consequently all the aforementioned results are reflected in the Markovian case. It is therefore tempting to study the Markovian logic from a model theoretic perspective.

We took these challenges and address them in a series of articles that we will present in this part of the paper. We approach the Markovian logics from a model theoretic perspective.

In the Chapter 7 we introduce the syntax and semantics for the three logics that we study, hereafter called the *probabilistic Markovian logic*, the *subprobabilistic Markovian logic* and the *stochastic Markovian logic* respectively. We propose, for each of them a sound Hilbert-style axiomatic system.

Next, we demonstrate that each of these logics enjoy the finite model property. Extending the finite model construction, we present classic canonic-model constructions. These results allow us to prove that the proposed axiomatic systems are weakly-complete with respect to the corresponding MP-semantics.

The following step in our research focuses on the strongly-complete axiomatizations. In our first papers on this topic we have used Goldblatt's technique to prove the strong completeness, i.e., we extended the axiomatic systems with the CAR rule and assume Lindenbaum's lemma as a meta-axiom. Latter, we discovered that there exists an alternative rule to CAR. We propose a new infinitary axiom schema to replace CAR. Unlike CAR, our axiom has a countable set of instances. This fact allows us to invoke the Rasiowa-Sikorski Lemma and prove the strong completeness theorem via a canonical models construction without needing to assume Lindenbaum's lemma. In fact Lindenbaum's lemma can be directly proven from the Extended Rasiowa-Sikorski lemma (see Lemma 3.7.2 from Preliminaries). Using this new rule, we extend the axiomatizations of the previous chapter to obtain strong-completeness. These results are the ones that lead our steps to Stone dualities that we will present in Part IV of this Monograph.

In Chapter 8 we emphasize how the model theory can be extended towards metric reasoning. For this, we show that whenever the class of models is organized as a (pseudo)metric space, the structure can be lifted to the space of logical properties. This lifting encodes many interesting properties widely used in practice, such as the concepts of *robustness* or *parameter-continuity*. This research has a direct connection to Stone duality as well and we will return to these arguments in Part IV of this Monograph.

In this Part of the Monograph we included results published in the following articles.

- [I] D. Kozen, R. Mardare, P. Panangaden. *Strong Completeness for Markovian Logics*. In Proc. of 38th International Symposium Mathematical Foundations of Computer Science, MFCS 2013, Lecture Notes in Computer Science, LNCS 8087, pages: 655-666, 2013.
- [II] R. Mardare, L. Cardelli, K. G. Larsen. *Continuous Markovian Logics - Axiomatization and Quantified Metatheory*. Logical Methods in Computer Science, LMCS vol.8(4):1-28, 2012.

- [III] K. G. Larsen, R. Mardare, P. Panangaden. *Take it to the Limit: Approximate Reasoning for Markov Processes*. In Proc. of 37th International Symposium Mathematical Foundations of Computer Science, MFCS 2012, Lecture Notes in Computer Science, LNCS 7464, pages: 681-692, 2012.
- [IV] L. Cardelli, K.G. Larsen, R. Mardare. *Continuous Markovian Logics - From Complete Axiomatization to the Metric Space of Formulas*. In Proc. of Computer Science Logic, CSL2011, LIPIcs 12 Schloss Dagstuhl - Leibniz-Zentrum fuer Informatik, vol.12, pages: 144-158, 2011.
- [V] L. Cardelli, K.G. Larsen, R. Mardare. *Modular Markovian Logic*. in Proc. of the 38th International Colloquium on Automata, Languages and Programming, ICALP 2011, Lecture Notes in Computer Science, LNCS 6756: 380-391, 2011.



# Chapter 7

## Markovian Logics

### 7.1 Markov Processes

In this section we introduce the three classes of models of probabilistic systems with a continuous state space: (i) *probabilistic Markov processes* (PMPs), (ii) *subprobabilistic Markov processes* (SMPs) and (iii) *general Markov processes* (GMPs). The first two classes contain the systems for which the transition from a state to a measurable set of states is characterized by its probability. The third class represents the systems with continuous-time transitions, i.e., the probability of a transition from a state to a measurable set of states depends on time. These classes can be extended by adding labelled transitions to emphasize the fact that there might be multiple possible actions. Here we will suppress the labels, as they do not contribute any relevant structure for our results.

In this chapter, to simplify the formalisms, we use the following notations

$$\mathbb{Q}_0 = \mathbb{Q} \cap [0, 1], \quad \mathbb{Q}^+ = \mathbb{Q} \cap [0, \infty), \quad \mathbb{R}_0 = \mathbb{R} \cap [0, 1], \quad \text{and} \quad \mathbb{R}^+ = \mathbb{R} \cap [0, \infty).$$

Given a measurable space  $(M, \Sigma)$ , we denote by  $\Pi(M, \Sigma)$ ,  $\Pi^*(M, \Sigma)$  and  $\Delta(M, \Sigma)$  the set of probabilistic distributions, sub-probabilistic distribution and general distributions on  $(M, \Sigma)$  respectively.

Each of these sets of distributions will be seen as a measurable space of distributions, where the measurable sets, as detailed in the Preliminaries, are elements of the  $\sigma$ -algebra generated by the sets

$$\{\mu \in \Delta(M, \Sigma) \mid \mu(S) \geq r\} \quad \text{for } S \in \Sigma \text{ and } r \in \mathbb{Q}^+.$$

This is the least  $\sigma$ -algebra on the space of distributions such that all the maps  $\mu \mapsto \mu(S) : \Delta(M, \Sigma) \rightarrow \mathbb{R}^+$  for  $S \in \Sigma$  are measurable, where the set of positive reals is endowed with the Borel  $\sigma$ -algebra.

**Definition 7.1.1** (Markov process). *Given an analytic space  $(M, \Sigma)$ ,*

- a probabilistic Markov process is a measurable mapping  $\theta \in \llbracket M \rightarrow \Pi(M, \Sigma) \rrbracket$ ;
- a subprobabilistic Markov process is a measurable mapping  $\theta \in \llbracket M \rightarrow \Pi^*(M, \Sigma) \rrbracket$ ;
- a general Markov process is a measurable mapping  $\theta \in \llbracket M \rightarrow \Delta(M, \Sigma) \rrbracket$ .

In what follows we identify a Markov process with the tuple  $\mathcal{M} = (M, \Sigma, \theta)$ ;  $M$  is called the *support set*, denoted by  $\text{supp}(\mathcal{M})$ , and  $\theta$  is called the *transition function*.

If  $\mathcal{M} = (M, \Sigma, \theta)$  is a (probabilistic/subprobabilistic/general) Markov process, then for  $m \in M$ ,  $\theta(m)$  is a (probabilistic/subprobabilistic/general) measure on the state space  $(M, \Sigma)$ . If  $\mathcal{M}$  is a PMP or a SMP, the value  $\theta(m)(N)$  for  $N \in \Sigma$  represents the probability of a transition from  $m$  to a state in  $N$ ; otherwise, if  $\mathcal{M}$  is a GMP, then  $\theta(m)$  is a measure on the state space and the value  $\theta(m)(N) \in \mathbb{R}^+$  represents the rate of an exponentially distributed random variable that characterizes the time of a transition from  $m$  to a state in  $N$ .

The condition that  $\theta$  is measurable is equivalent to the condition that for fixed  $N \in \Sigma$ , the function  $m \mapsto \theta(m)(N)$  is measurable (see e.g. Proposition 2.9 of [20]).

Given two Markov processes  $\mathcal{M}_i = (M_i, \Sigma_i, \theta_i)$ ,  $i = 1, 2$ , a surjective measurable function  $f : M_1 \rightarrow M_2$  is a *zig-zag* if for any  $m \in M_1$  and  $B \in \Sigma_2$ ,

$$\theta_1(m)(f^{-1}(B)) = \theta_2(f(m))(B).$$

Such a map is essentially a functional version of bisimulation [15].

Now we recall one of the equivalent (for analytic spaces) definitions of the MP bisimulation (for details see the Preliminaries).

A *span* in a category is a pair of morphisms  $f : A \rightarrow B$  and  $g : A \rightarrow C$  with a common domain.

Two Markov processes  $\mathcal{M}_1, \mathcal{M}_2$  are said to be *bisimilar* if there is a third Markov process  $\mathcal{M}$  and a span of zig-zags  $f_i : \mathcal{M} \rightarrow \mathcal{M}_i$ ,  $i = 1, 2$ .

Two states  $m_i \in \text{supp}(\mathcal{M}_i)$ ,  $i = 1, 2$ , are said to be *bisimilar* if there exist a span of zig-zags  $f_i : \mathcal{M} \rightarrow \mathcal{M}_i$ ,  $i = 1, 2$  and  $m \in \text{supp}(\mathcal{M})$  such that  $m_i = f_i(m)$ ,  $i = 1, 2$ .

We write  $(\mathcal{M}_1, m_1) \approx (\mathcal{M}_2, m_2)$  to indicate that  $m_1$  and  $m_2$  are bisimilar in this sense.

## 7.2 Syntax of Markovian Logics

*Markovian logics* are multi-modal logics for semantics based on the three classes of Markov processes introduced in the previous section. They have been introduced and studied in various contexts [3, 4, 10, 22, 27, 34, 35, 49].

In addition to the boolean operators, these logics are equipped with modal operators of type  $L_r$  for rational numbers  $r$  that are used to approximate the numerical labels of the transitions. Intuitively, the formula  $L_r\varphi$  is satisfied by  $m \in \mathcal{M}$  whenever the

probability/rate of a transition from  $m$  to a state satisfying the logical property  $\varphi$  is at least  $r$ .

In this chapter we study three Markovian logics: the *probabilistic Markovian logic* (PML), the *sub-probabilistic Markovian logic* (SML) and the *general Markovian logic* (GML); they are interpreted on PMPs, SMPs and GMPs respectively. Despite their apparent similarities, we have found it necessary to treat these logics separately because of subtle technical differences that make a uniform model theoretic treatment difficult.

**Definition 7.2.1.** *Given a countable set  $\mathcal{P}$  of atomic propositions, the grammars below define the sets of formulas  $\mathcal{L}(\Pi)$  of probabilistic and subprobabilistic Markovian logic,  $\mathcal{L}(\Pi^*)$  of subprobabilistic Markovian logic and  $\mathcal{L}(\Delta)$  of general Markovian logic*

$$\begin{aligned}\mathcal{L}(\Pi) : \quad \varphi &::= p \mid \neg\varphi \mid \varphi \wedge \varphi \mid L_r\varphi, \quad \text{for arbitrary } p \in \mathcal{P} \text{ and } r \in \mathbb{Q}_0 \\ \mathcal{L}(\Pi^*) : \quad \varphi &::= p \mid \neg\varphi \mid \varphi \wedge \varphi \mid L_r\varphi, \quad \text{for arbitrary } p \in \mathcal{P} \text{ and } r \in \mathbb{Q}_0 \\ \mathcal{L}(\Delta) : \quad \varphi &::= p \mid \neg\varphi \mid \varphi \wedge \varphi \mid L_r\varphi, \quad \text{for arbitrary } p \in \mathcal{P} \text{ and } r \in \mathbb{Q}^+\end{aligned}$$

Note that the languages  $\mathcal{L}(\Pi)$  and  $\mathcal{L}(\Pi^*)$  are the same (syntactically identical) and very similar to  $\mathcal{L}(\Delta)$ . Nevertheless, the three logics are interpreted over different structures and, thus, have different semantics and proof systems.

For each of these logics we assume that the usual boolean operators  $\top, \perp, \vee, \rightarrow$  are available as derived constructs as well as the additional derived operator

$$L_{r_1 \dots r_n} \varphi = L_{r_1} \dots L_{r_n} \varphi$$

defined for  $r_1, \dots, r_n \in \mathbb{Q}_0$  for  $\mathcal{L}(\Pi)$  and for  $r_1, \dots, r_n \in \mathbb{R}^+$  for  $\mathcal{L}(\Delta)$ .

### 7.3 Semantics of Markovian logics

In what follows we define *en masse* the semantics for three logics using a generic  $\mathcal{L}$  that ranges over the set  $\{\mathcal{L}(\Pi), \mathcal{L}(\Pi^*), \mathcal{L}(\Delta)\}$ . However, each of the following concepts has to be properly interpreted in each case.

Let  $\mathcal{M} = (M, \Sigma, \theta)$  be a PMP when we consider  $\mathcal{L} = \mathcal{L}(\Pi)$ , an SMP when we consider  $\mathcal{L} = \mathcal{L}(\Pi^*)$  and an GMP when we consider  $\mathcal{L} = \mathcal{L}(\Delta)$ .

Let  $m \in M$  be an arbitrary state and  $i : M \rightarrow 2^{\mathcal{P}}$  an arbitrary interpretation function for the atomic propositions. The semantics of the three logics is defined as follows.

- $\mathcal{M}, m, i \models p$  iff  $p \in i(m)$ ,
- $\mathcal{M}, m, i \models \varphi \wedge \psi$  iff  $\mathcal{M}, m, i \models \varphi$  and  $\mathcal{M}, m, i \models \psi$ ,
- $\mathcal{M}, m, i \models \neg\varphi$  iff not  $\mathcal{M}, m, i \models \varphi$ .
- $\mathcal{M}, m, i \models L_r\varphi$  iff  $\theta(m)(\llbracket \varphi \rrbracket_{\mathcal{M}}^i) \geq r$ ,  
where  $\llbracket \varphi \rrbracket_{\mathcal{M}}^i = \{m \in M \mid \mathcal{M}, m, i \models \varphi\}$ .

For the last clause to make sense,  $\llbracket \varphi \rrbracket_{\mathcal{M}}^i$  must be measurable. This is guaranteed, for each of the three types of Markov process, by the fact that  $\theta$  is a measurable mapping between the measurable space of states and the measurable space of probabilistic/subprobabilistic/general distributions.

- Lemma 7.3.1.** 1. For any  $\varphi \in \mathcal{L}(\Pi)$ , any PMP  $\mathcal{M} = (M, \Sigma, \theta)$  and any interpretation function  $i$ ,  $\llbracket \varphi \rrbracket_{\mathcal{M}}^i \in \Sigma$ .
2. For any  $\varphi \in \mathcal{L}(\Pi^*)$ , any SMP  $\mathcal{M} = (M, \Sigma, \theta)$  and any interpretation function  $i$ ,  $\llbracket \varphi \rrbracket_{\mathcal{M}}^i \in \Sigma$ .
3. For any  $\varphi \in \mathcal{L}(\Delta)$ , any GMP  $\mathcal{M} = (M, \Sigma, \theta)$  and any interpretation function  $i$ ,  $\llbracket \varphi \rrbracket_{\mathcal{M}}^i \in \Sigma$ .

*Proof.* Each case can be proven inductively on the structure of  $\varphi$ . We only present here the case  $\varphi = L_r\psi$  for  $\mathcal{L}(\Delta)$ .

**The case  $\varphi = L_r\psi$ :**  $\llbracket L_r\psi \rrbracket_{\mathcal{M}}^i = \theta^{-1}(\{\mu \in \Delta(M, \Sigma) \mid \mu(\llbracket \psi \rrbracket_{\mathcal{M}}^i) \geq r\})$ . From the inductive hypothesis,  $\llbracket \psi \rrbracket_{\mathcal{M}}^i \in \Sigma$ , hence,  $\{\mu \in \Delta(M, \Sigma) \mid \mu(\llbracket \psi \rrbracket_{\mathcal{M}}^i) \geq r\}$  is measurable in  $\Delta(M, \Sigma)$  and because  $\theta$  is a measurable mapping, we obtain that  $\llbracket L_r\psi \rrbracket_{\mathcal{M}}^i$  is measurable.  $\square$

Given  $\mathcal{M} = (M, \Sigma, \theta)$  and  $i$ , we say that  $m \in M$  satisfies  $\varphi$  if  $\mathcal{M}, m, i \models \varphi$ .

We write  $\mathcal{M}, m, i \not\models \varphi$  if it is not the case that  $\mathcal{M}, m, i \models \varphi$ .

For a set  $\Phi \subseteq \mathcal{L}$ , we write  $\mathcal{M}, m, i \models \Phi$  if  $\mathcal{M}, m, i \models \varphi$  for all  $\varphi \in \Phi$ . We write  $\Phi \models \varphi$  if  $\mathcal{M}, m, i \models \varphi$  whenever  $\mathcal{M}, m, i \models \Phi$ .

A formula or set of formulas is *satisfiable* if there exist an MP  $\mathcal{M}$ , an interpretation function  $i$  for  $\mathcal{M}$  and  $m \in \text{supp}(\mathcal{M})$  that satisfies it. We say that  $\varphi$  is *valid* and write  $\models \varphi$ , if  $\neg\varphi$  is not satisfiable.

In what follows, when we have to differentiate between the three semantics, we will use indexes:  $\models_{\Pi}$  will be used for PML,  $\models_{\Pi^*}$  for SML and  $\models_{\Delta}$  for GML.

In the end of this section we demonstrate that the three Markovian logics are *adequate* with respect to their semantics, in the sense that the semantic equivalence induced by each of these logics over their class of models coincides with the corresponding concept of bisimilarity. This result is often referred to as Hennessy-Milner theorem. For the probabilistic case the result has already been proven in [14, 15, 46].

We state this result for each of the three logics. However, in the following theorem we consider the three Markovian logics defined for the empty set of atomic propositions,

$$\mathcal{P} = \emptyset,$$

in which case we implicitly assume that the grammar that defines the formulas uses  $\top$  as the terminal symbol.

The logics for which  $\mathcal{P} \neq \emptyset$  characterize the labeled-bisimulation of MPs that have the states labeled by the elements in  $\mathcal{P}$  and the definition of bisimulation includes the requirement that bisimilar states have the same labels. This more general case raises

no additional issues and for this reason, in what follows we will only present the case  $\mathcal{P} = \emptyset$ .

**Theorem 7.3.2.** *The following statements are verified for the case  $\mathcal{L}(\Pi)$ ,  $\mathcal{L}(\Pi^*)$  and  $\mathcal{L}(\Delta)$  are defined for  $\mathcal{P} = \emptyset$ .*

1. *For arbitrary PMPs  $\mathcal{M}_j$  and arbitrary  $m_j \in \text{supp}(\mathcal{M}_j)$ ,  $j = 1, 2$ ,  $(\mathcal{M}_1, m_1) \approx (\mathcal{M}_2, m_2)$  iff,*  

$$\text{for any } \varphi \in \mathcal{L}(\Pi) \text{ and } i, \mathcal{M}_1, m_1, i \models_{\Pi} \varphi \text{ iff } \mathcal{M}_2, m_2, i \models_{\Pi} \varphi.$$
2. *For arbitrary SMPs  $\mathcal{M}_j$  and arbitrary  $m_j \in \text{supp}(\mathcal{M}_j)$ ,  $j = 1, 2$ ,  $(\mathcal{M}_1, m_1) \approx (\mathcal{M}_2, m_2)$  iff,*  

$$\text{for any } \varphi \in \mathcal{L}(\Pi^*) \text{ and } i, \mathcal{M}_1, m_1, i \models_{\Pi^*} \varphi \text{ iff } \mathcal{M}_2, m_2, i \models_{\Pi^*} \varphi.$$
3. *For arbitrary GMPs  $\mathcal{M}_j$  and arbitrary  $m_j \in \text{supp}(\mathcal{M}_j)$ ,  $j = 1, 2$ ,  $(\mathcal{M}_1, m_1) \approx (\mathcal{M}_2, m_2)$  iff,*  

$$\text{for any } \varphi \in \mathcal{L}(\Delta) \text{ and } i, \mathcal{M}_1, m_1, i \models_{\Delta} \varphi \text{ iff } \mathcal{M}_2, m_2, i \models_{\Delta} \varphi.$$

*Proof.* Since 1. and 2. has been proven in related papers, hereafter we prove only 3. This proof reproduces at stochastic level the proof of Lemma 7.16 presented in [46] for probabilistic systems.

The only difficult implication is from right to left. To prove this we take firstly the disjoint union  $\mathcal{M} = (M, \Sigma, \theta)$  of the two GMPs and we will treat the bisimulation between  $\mathcal{M}_1$  and  $\mathcal{M}_2$  as an "internal" bisimulation in  $\mathcal{M}$ . Hence,  $m_1, m_2 \in \text{supp}(\mathcal{M})$ .

Before starting the proof, we introduce some additional concepts and present some results that are needed for our proof.

Let  $\cong$  be the relation on  $\text{supp}(\mathcal{M})$  defined by

$$m \cong n \text{ iff for any } \varphi \in \mathcal{L}(\Delta) \text{ and } i, \mathcal{M}, m, i \models_{\Delta} \varphi \text{ iff } \mathcal{M}, n, i \models_{\Delta} \varphi.$$

Because  $\cong$  is an equivalence relation, we can take for a given  $(M, \Sigma)$  the quotient  $(M^{\cong}, \Sigma^{\cong})$  constructed as follows.  $M^{\cong}$  is the set of all equivalence classes of  $M$ ; there exists a projection  $\pi : M \rightarrow M^{\cong}$  which maps each element to its equivalence class.  $\pi$  determines a  $\sigma$ -algebra  $\Sigma^{\cong}$  on  $M^{\cong}$  by  $S \in \Sigma^{\cong}$  iff  $\pi^{-1}(S) \in \Sigma$ . We call  $\pi$  the *canonical projection* from  $(M, \Sigma)$  into  $(M^{\cong}, \Sigma^{\cong})$ .

We state now a few results that allow us to prove the theorem.

Given a set  $X$ , a family of subsets  $\Pi \subset 2^X$  closed under finite intersection is called  $\pi$ -system. A family of subsets  $\Lambda \subset 2^X$  is a  $\lambda$ -system if contains  $X$  and is closed under complementation and countable union of pairwise disjoint sets.

**[Dynkin's  $\lambda - \pi$  theorem]:** If  $\Pi$  is a  $\pi$ -system and  $\Lambda$  is a  $\lambda$ -system, then  $\Pi \subset \Lambda$  implies  $\overline{\Pi} \subseteq \Lambda$ , where  $\overline{\Pi}$  is the  $\sigma$ -algebra generated by  $\Pi$ .

Dynkin's theorem allows us to prove the next lemma.

**[Lemma A.]** Suppose that  $\Pi \subseteq 2^X$  is a  $\pi$ -system with  $X \in \Pi$  and  $\mu, \nu$  are two measures on  $(X, \overline{\Pi})$ . If  $\mu$  and  $\nu$  agree on all the sets in  $\Pi$ , then they agree on  $\overline{\Pi}$ .

We also present two more lemmas (see, e.g., [46] Section 7.7).

**[Lemma B.]** Let  $(M, \Sigma)$  be an analytic set and let  $\Sigma_0$  be a countably generated sub- $\sigma$ -algebra of  $\Sigma$  which separates points in  $M$ , i.e., for any  $m, n \in M$ ,  $m \neq n$ , there exists  $S \in \Sigma_0$  such that  $m \in S \not\equiv n$ . Then  $\Sigma_0 = \Sigma$ .

**[Lemma C.]** Let  $(M, \Sigma)$  be an analytic set and let  $\equiv$  be an equivalence relation on  $M$ . If there exists a sequence  $f_1, f_2, \dots$  of real-valued Borel functions on  $M$  such that  $m \equiv n$  iff for all  $i$ ,  $f_i(m) = f_i(n)$ , then  $(M^{\equiv}, \Sigma^{\equiv})$  is an analytic set.

**[Proposition D].** For any GMP  $\mathcal{M} = (M, \Sigma, \theta)M$ , there exists a GMP  $\mathcal{M}^{\approx} = (M^{\approx}, \Sigma^{\approx}, \theta^{\approx})$  such that the canonical projection  $\pi : (M, \Sigma, \theta) \rightarrow (M^{\approx}, \Sigma^{\approx}, \theta^{\approx})$  is a zigzag morphism.

Now we prove Proposition D:

For the beginning we show that  $(M^{\approx}, \Sigma^{\approx})$  is an analytic set.

Since  $\mathcal{L}(\Delta)$  is countable, we can assume that  $\mathcal{L}(\Delta) = \{\varphi_k \mid k \in \mathbb{N}\}$ . Because  $\llbracket \varphi_k \rrbracket_{\mathcal{M}}^i$  is measurable, the characteristic functions  $1_{\varphi_k} : M \rightarrow \{0, 1\}$  are measurable and  $m \approx n$  iff  $[\forall k \in \mathbb{N}, 1_{\varphi_k}(m) = 1_{\varphi_k}(n)]$ . Lemma C proves further that  $(M^{\approx}, \Sigma^{\approx})$  is an analytic set.

Let  $\mathcal{B} = \{\pi(\llbracket \varphi_k \rrbracket_{\mathcal{M}}^i \mid k \in \mathbb{N})\}$ . We show that the  $\sigma$ -algebra  $\sigma(\mathcal{B})$  generated by  $\mathcal{B}$  coincides with  $\Sigma^{\approx}$ . Obviously,  $\mathcal{B} \subseteq \Sigma^{\approx}$ , because for any  $\pi(\llbracket \varphi_k \rrbracket_{\mathcal{M}}^i) \in \mathcal{B}$ ,  $\pi^{-1}(\pi(\llbracket \varphi_k \rrbracket_{\mathcal{M}}^i)) \in \Sigma$ . Notice that  $\sigma(\mathcal{B})$  separates points in  $M^{\approx}$ : let  $C, D \in M^{\approx}$ ,  $C \neq D$  and let  $m \in \pi^{-1}(C)$ ,  $n \in \pi^{-1}(D)$ ; because  $m \not\approx n$ , there exists  $\varphi \in \mathcal{L}(\Delta)$  and some  $i$  such that  $m \in \llbracket \varphi \rrbracket_{\mathcal{M}}^i \not\equiv n$ . Hence, we can apply Lemma B and we obtain  $\sigma(\mathcal{B}) = \Sigma^{\approx}$ .

Now we define  $\theta^{\approx}$  such that  $\pi$  is a zigzag. Notice first that  $\pi$  is measurable and surjective by definition. For each  $C \in \Sigma^{\approx}$ , let  $\theta^{\approx}(m^{\approx})(C) = \theta(m)(\pi^{-1}(C))$ .

This definition is correct: let  $m, n \in m^{\approx}$ , we prove that  $\theta(m)$  and  $\theta(n)$  agree on  $\Sigma^{\approx}$ . We show first that they agree on  $\llbracket \varphi \rrbracket_{\mathcal{M}}^i \in \mathcal{B}$ .

Suppose that we have  $\theta(m)(\llbracket \varphi \rrbracket_{\mathcal{M}}^i) < r < \theta(n)(\llbracket \varphi \rrbracket_{\mathcal{M}}^i)$ . Then,  $\mathcal{M}, m \models \neg L_r \varphi$  while  $\mathcal{M}, n \models L_r \varphi$  - impossible. Because  $\mathcal{B}$  is closed under finite intersection ( $\llbracket \varphi \rrbracket_{\mathcal{M}}^i \cap \llbracket \psi \rrbracket_{\mathcal{M}}^j = \llbracket \varphi \wedge \psi \rrbracket_{\mathcal{M}}^{i,j}$ ) and  $M = \llbracket \top \rrbracket_{\mathcal{M}} \in \mathcal{B}$ , we apply Lemma A and obtain that  $\theta(m)$  and  $\theta(n)$  agree on  $\Sigma^{\approx}$ .

Now we only need to prove that  $\theta^{\approx}$  is measurable.

Let  $C \in \Sigma^{\approx}$  and  $A$  a Borel set of  $\mathbb{R}^+$ . We have

$$(\theta^{\approx})^{-1}(\{\mu \in \Delta(M^{\approx}, \Sigma^{\approx}) \mid \mu(C) \in A\}) = \pi(\theta^{-1}(\{\nu \in \Delta(M, \Sigma) \mid \nu(\pi^{-1}(C)) \in A\})).$$

But  $\{\nu \in \Delta(M, \Sigma) \mid \nu(\pi^{-1}(B)) \in A\}$  is measurable in  $\Delta(M, \Sigma)$  and because  $\theta$  is measurable, we obtain that  $\theta^{-1}(\{\nu \in \Delta(M, \Sigma) \mid \nu(\pi^{-1}(B)) \in A\}) \in \Sigma$  implying  $\pi(\theta^{-1}(\{\nu \in \Delta(M, \Sigma) \mid \nu(\pi^{-1}(B)) \in A\})) \in \Sigma^{\cong}$ . And this concludes the proof of Proposition D.

Now we have the ingredients to prove Theorem 7.3.2.

We prove that  $\cong$  is a rate bisimulation.

Let  $C \in \Sigma(\cong)$ . Then,  $C = \pi^{-1}(\pi(C))$ , where  $\pi$  is the canonical projection. Because  $\pi$  is measurable, we get that  $\pi(C) \in \Sigma^{\cong}$ .

If  $m \cong m'$ , then  $\pi(m) = \pi(m')$ . Hence,

$$\theta(m)(C) = \theta(m)(\pi^{-1}(\pi(C))) = \theta^{\cong}(\pi(m))(\pi(C)),$$

because  $\pi$  is a zigzag morphism. But

$$\theta^{\cong}(\pi(m))(\pi(C)) = \theta^{\cong}(\pi(m'))(\pi(C))$$

and

$$\theta^{\cong}(\pi(m'))(\pi(C)) = \theta(m')(C).$$

This proves that  $\theta(m)(C) = \theta(m')(C)$  and it concludes the proof.  $\square$

## 7.4 Hilbert-style Axiomatizations

We now present sound Hilbert-style axiomatic systems for each of the three logics. These axiomatic systems include the axioms of propositional logic, even if we will not write them explicitly.

In the next section we prove that these system are *weakly complete* for their semantics, meaning that an arbitrary formula  $\varphi$  can be proven from the axioms if and only if it is satisfied by all the states of all the models.

As we did for the semantics, we introduce the concepts related to the provability *en masse*. However, they have a specific meaning for each logic and depend directly of each particular provability relation.

As usual, for an arbitrary formula  $\varphi$ ,  $\vdash \varphi$  denotes the fact that  $\varphi$  is an axiom or a theorem in the system.

If  $\Phi$  is a set of formulas, we write  $\Phi \vdash \varphi$  and say that  $\Phi$  *derives*  $\varphi$  if  $\varphi$  is provable from the axioms and the extra assumptions  $\Phi$ . In this interpretation, this is, obviously, a *deducibility relation on  $\mathcal{L}$* , i.e., it satisfies the following conditions

(R1) If  $\Phi \vdash \varphi$  and  $\Phi \subseteq \Phi'$ , then  $\Phi' \vdash \varphi$ ;

(R2) If  $\varphi \in \Phi$ , then  $\Phi \vdash \varphi$ ;

(R3) If  $\Phi \vdash \varphi$  and  $\Phi \cup \{\varphi\} \vdash \perp$ , then  $\Phi \vdash \perp$ ;

(R4)  $\Phi \cup \{\neg\varphi\} \vdash \perp$  iff  $\Phi \vdash \varphi$ .

Note that  $\vdash \varphi$  iff  $\emptyset \vdash \varphi$ .

A formula or set of formulas is *consistent* if it cannot derive  $\perp$ . We say that  $\Phi$  is  *$\mathcal{L}$ -maximally consistent* if it is consistent and it has no proper consistent extensions in  $\mathcal{L}$ .

When we have to differentiate between the three provability relations, we will use indexes:  $\vdash_{\Pi}$  will be used for PML,  $\vdash_{\Pi^*}$  for SML and  $\vdash_{\Delta}$  for GML.

### 7.4.1 Axioms of Probabilistic Markovian Logic

The axiomatic system of PML is listed in Table 9.2. The axioms and the rules are stated for arbitrary  $\varphi, \psi \in \mathcal{L}$  and arbitrary  $r, s, r_1, \dots, r_k \in \mathbb{Q}_0$  for  $k \geq 0$ . As mentioned before, these axioms are considered in addition to the axioms and rules of the classic propositional logic.

(A1):	$\vdash_{\Pi} L_0\varphi$
(A2):	$\vdash_{\Pi} L_r\top$
(A3):	$\vdash_{\Pi} L_r\varphi \rightarrow \neg L_s\neg\varphi, \quad r + s > 1$
(A4):	$\vdash_{\Pi} L_r(\varphi \wedge \psi) \wedge L_s(\varphi \wedge \neg\psi) \rightarrow L_{r+s}\varphi, \quad r + s \leq 1$
(A5):	$\vdash_{\Pi} \neg L_r(\varphi \wedge \psi) \wedge \neg L_s(\varphi \wedge \neg\psi) \rightarrow \neg L_{r+s}\varphi, \quad r + s \leq 1$
(R1):	$\frac{\vdash_{\Pi} \varphi \rightarrow \psi}{\vdash_{\Pi} L_r\varphi \rightarrow L_r\psi}$
(R2):	$\{L_{r_1 \dots r_k} \psi \mid r < s\} \vdash_{\Pi} L_{r_1 \dots r_k} \psi$

Table 7.1: The axioms of  $\mathcal{L}(\Pi)$

Axiom (A1) guarantees that the rate of any transition is at least 0 and encodes the fact that the measure of any set cannot be negative.

Axiom (A2) states that the rate of exiting a state is at least  $r$  for any  $r \leq 1$ , i.e., it is exactly 1.

Axiom (A3) relies on the fact that the sum of the probabilities of transition from a given state to a state satisfying  $\varphi$  and from the same state to a state satisfying  $\neg\varphi$  is 1.

The axioms (A4) and (A5) encode the additive properties of measures for disjoint sets:  $\llbracket \varphi \wedge \psi \rrbracket$  and  $\llbracket \varphi \wedge \neg\psi \rrbracket$  are disjoint sets of processes such that  $\llbracket \varphi \wedge \psi \rrbracket \cup \llbracket \varphi \wedge \neg\psi \rrbracket = \llbracket \varphi \rrbracket$ .

The rule (R1) establishes the monotonicity of  $L_r$ .

The rule (R2) reflects the Archimedian property of rationals: if the probability of a transition from a state to a given set of states is at least  $r$  for any  $r < s$ , then it is at least  $s$ ; and this phenomenon can be identified under any finite number of occurrences of the operator  $L_r$ .

A similar axiomatic system was independently proposed in [49, 50] for Harsanyi type spaces. The novelty of our axiomatization is the rule (R2) which is much stronger than the one used in these papers – in fact our rule is sufficient to prove the strong-completeness. In [49], for proving the strong completeness of the axiomatic system, Lindenbaum's lemma is assumed as a meta-axiom and instead of (R2), the rules in Table 7.2 are used, stated for arbitrary  $\varphi \in \mathcal{L}(\Pi)$  and arbitrary set  $\Phi \subseteq \mathcal{L}(\Pi)$  closed under conjunction, where  $L_r\Phi = \{L_r\psi \mid \psi \in \Phi\}$ .

$$(R2'): \quad \{L_r\psi \mid r < s\} \vdash_{\Pi} L_s\psi$$

$$(R2''): \quad \frac{\Phi \vdash_{\Pi} \varphi}{L_r\Phi \vdash_{\Pi} L_r\varphi}$$

Table 7.2: Zhou's rules of  $\mathcal{L}(\Pi)$

While (R2') is an instance of (R2), (R2''), which is Goldblatt's (CAR) rule, is much stronger and it has an uncountable set of instances. This makes the proof of the existence of the maximally consistent sets difficult since one cannot simply apply Zorn's lemma in a constructive proof.

In our case that fact that (R2) has countably many instances allows us to apply the Rasiowa-Sikorski lemma in order to guarantee that any consistent formula belongs to a maximally-consistent set of formulas. On this we will later base the canonical model construction.

The next proposition presents a couple of  $\vdash_{\Pi}$  theorems that are useful in explaining the relation between the three Markovian logics.

**Proposition 7.4.1.** 1.  $\vdash_{\Pi} L_r\perp \rightarrow \perp$ .

$$2. \vdash L_{r+s}\varphi \rightarrow L_r\varphi.$$

*Proof.* 1. From (A2),  $\vdash_{\Pi} \neg L_s\top \rightarrow \perp$  for any  $s$  such that  $r + s > 1$ , and further we use (A3).

2. From (A5),  $\vdash L_{r+s}\varphi \rightarrow L_r(\varphi \wedge \top) \vee L_s(\varphi \wedge \perp)$  and using 1. we get the result.  $\square$

### 7.4.2 Axioms of Sub-probabilistic Markovian Logic

Before introducing the axiomatization of SML, notice that the axiom (A3) of PML guarantees that the semantics must use distributions bounded by 1, while the axiom (A2) guarantees that the distributions are, in fact, probability distributions.

The axiomatic system of SML is listed in Table 9.3. The axioms and the rules are stated for arbitrary  $\varphi, \psi \in \mathcal{L}$  and arbitrary  $r, s, r_1, \dots, r_k \in \mathbb{Q}_0$  for  $k \geq 0$ .

- (A1):  $\vdash_{\Pi^*} L_0\varphi$
- (A2'):  $\vdash_{\Pi^*} L_r\perp \rightarrow \perp$
- (A3):  $\vdash_{\Pi^*} L_r\varphi \rightarrow \neg L_s\neg\varphi, r + s > 1$
- (A4):  $\vdash_{\Pi^*} L_r(\varphi \wedge \psi) \wedge L_s(\varphi \wedge \neg\psi) \rightarrow L_{r+s}\varphi, r + s \leq 1$
- (A5):  $\vdash_{\Pi^*} \neg L_r(\varphi \wedge \psi) \wedge \neg L_s(\varphi \wedge \neg\psi) \rightarrow \neg L_{r+s}\varphi, r + s \leq 1$
- (R1): 
$$\frac{\vdash_{\Pi^*} \varphi \rightarrow \psi}{\vdash_{\Pi^*} L_r\varphi \rightarrow L_r\psi}$$
- (R2):  $\{L_{r_1 \dots r_k} \psi \mid r < s\} \vdash_{\Pi^*} L_{r_1 \dots r_k} \psi$

Table 7.3: The axioms of  $\mathcal{L}(\Pi^*)$

Notice the difference between this axiomatic system and the previous one developed for PML: for SML the axiom (A2) is not sound any more, since for a sub-probability distribution the measure of the entire space can be smaller than 1.

However, (A2'), which replaces (A2), is also sound for PML, as proven in Proposition 7.4.1. Hence, all the  $\vdash_{\Pi^*}$ -theorems are also  $\vdash_{\Pi}$ -theorems, but the reverse implication is not true.

Regarding the axioms and rules that are identical with the ones for PML, we did not change their names.

### 7.4.3 Axioms of the General Markovian Logic

The axiomatic system of GML is listed in Table 7.4. The axioms and the rules are stated for arbitrary  $\varphi, \psi \in \mathcal{L}$  and arbitrary  $r, s, r_1, \dots, r_k \in \mathbb{Q}^+$  for  $k \geq 0$ .

The difference with respect to the axiomatic system of SML is that the axiom (A3) is not sound any more. Moreover, the indexes of the modal operator can be any positive rational, meaning that these axioms have more instances than the corresponding ones in the other two systems.

(A1):	$\vdash_{\Delta} L_0\varphi$
(A2'):	$\vdash_{\Delta} L_r\perp \rightarrow \perp$
(A4):	$\vdash_{\Delta} L_r(\varphi \wedge \psi) \wedge L_s(\varphi \wedge \neg\psi) \rightarrow L_{r+s}\varphi, r + s \leq 1$
(A5):	$\vdash_{\Delta} \neg L_r(\varphi \wedge \psi) \wedge \neg L_s(\varphi \wedge \neg\psi) \rightarrow \neg L_{r+s}\varphi, r + s \leq 1$
(R1):	$\frac{\vdash_{\Delta} \varphi \rightarrow \psi}{\vdash_{\Delta} L_r\varphi \rightarrow L_r\psi}$
(R2):	$\{L_{r_1 \dots r_k} \psi \mid r < s\} \vdash_{\Delta} L_{r_1 \dots r_k} s \psi$
(R3):	$\{L_{r_1 \dots r_k} \psi \mid r \in \mathbb{Q}^+\} \vdash_{\Delta} L_{r_1 \dots r_k} \perp$

Table 7.4: The axioms of  $\mathcal{L}(\Delta)$ 

Since the semantics does not allow infinite measures, rule (R3) guarantees that divergent sequences of modalities prefixing some formula generates an inconsistent set of formulas.

The next theorem states the soundness of the axioms of the three logics for their corresponding semantics

**Theorem 7.4.2.** *[Soundness]*

1. *The axiomatization of PML is sound for the PMPs semantics, i.e.,*

*for any  $\varphi \in \mathcal{L}(\Pi), \vdash_{\Pi} \varphi$  implies  $\vDash_{\Pi} \varphi$ .*

2. *The axiomatization of SML is sound for the SMPs semantics, i.e.,*

*for any  $\varphi \in \mathcal{L}(\Pi^*), \vdash_{\Pi^*} \varphi$  implies  $\vDash_{\Pi^*} \varphi$ .*

3. *The axiomatization of GML is sound for the GMPs semantics, i.e.,*

*for any  $\varphi \in \mathcal{L}(\Delta), \vdash_{\Delta} \varphi$  implies  $\vDash_{\Delta} \varphi$ .*

*Proof.* As usual, the soundness proof consists in proving that each axiom is sound with respect to the corresponding semantics and that the rules preserve soundness. This is sufficient to guarantee that we can only derive sound consequences from sound hypothesis.

In what follows we sketch the soundness proof for of a few of our axioms, the other being proved in a similar way.

Consider an arbitrary  $\mathcal{M} = (M, \Sigma, \theta)$  and an arbitrary  $m \in \text{supp}(\mathcal{M})$ .

**Soundness of (A1):** We have  $\models L_0\varphi$  since  $\theta(m)(\llbracket\varphi\rrbracket^i) \geq 0$  for any  $\varphi$ .

**Soundness of (A3):** Suppose that  $\mathcal{M}, m, i \models L_r(\varphi \wedge \psi) \wedge L_s(\varphi \wedge \neg\psi)$ . Then,

$$\theta(m)(\llbracket\varphi \wedge \psi\rrbracket^i) \geq r \quad \text{and} \quad \theta(m)(\llbracket\varphi \wedge \neg\psi\rrbracket^i) \geq s.$$

But since  $\llbracket\varphi \wedge \psi\rrbracket^i$  and  $\llbracket\varphi \wedge \neg\psi\rrbracket^i$  are disjoint sets of processes such that

$$\llbracket\varphi \wedge \psi\rrbracket^i \cup \llbracket\varphi \wedge \neg\psi\rrbracket^i = \llbracket\varphi\rrbracket^i,$$

$\theta(m)(\llbracket\varphi\rrbracket^i) = \theta(m)(\llbracket\varphi \wedge \psi\rrbracket^i) + \theta(m)(\llbracket\varphi \wedge \neg\psi\rrbracket^i)$ . Hence,  $\theta(m)(\llbracket\varphi\rrbracket^i) \geq r + s$ , i.e.,  $\mathcal{M}, m, i \models L_{r+s}\varphi$ .

**Soundness of (A4):** Suppose that  $\mathcal{M}, m, i \models \neg L_r(\varphi \wedge \psi) \wedge \neg L_s(\varphi \wedge \neg\psi)$ . Then,

$$\theta(m)(\llbracket\varphi \wedge \psi\rrbracket^i) < r \quad \text{and} \quad \theta(m)(\llbracket\varphi \wedge \neg\psi\rrbracket^i) < s.$$

But since  $\llbracket\varphi \wedge \psi\rrbracket^i$  and  $\llbracket\varphi \wedge \neg\psi\rrbracket^i$  are disjoint sets of processes such that

$$\llbracket\varphi \wedge \psi\rrbracket^i \cup \llbracket\varphi \wedge \neg\psi\rrbracket^i = \llbracket\varphi\rrbracket^i,$$

$\theta(m)(\llbracket\varphi\rrbracket^i) = \theta(m)(\llbracket\varphi \wedge \psi\rrbracket^i) + \theta(m)(\llbracket\varphi \wedge \neg\psi\rrbracket^i)$ . Hence,  $\theta(m)(\llbracket\varphi\rrbracket^i) < r + s$ , i.e.,  $\mathcal{M}, m, i \models \neg L_{r+s}\varphi$ .

**Soundness of (R1):** If  $\models \varphi \rightarrow \psi$ , then  $\llbracket\varphi\rrbracket^i \subseteq \llbracket\psi\rrbracket^i$ . Suppose that  $\mathcal{M}, m, i \models L_r\varphi$ . Then,  $\theta(m)(\llbracket\varphi\rrbracket^i) \geq r$ . Since  $\llbracket\varphi\rrbracket^i \subseteq \llbracket\psi\rrbracket^i$ , we derive that  $\theta(m)(\llbracket\varphi\rrbracket^i) \leq \theta(m)(\llbracket\psi\rrbracket^i)$ , hence,  $\theta(m)(\llbracket\psi\rrbracket^i) \geq r$  implying  $\mathcal{M}, m, i \models L_s\psi$ .

**Soundness of (R2):** We have to prove that  $\{L_{r_1 \dots r_k} \psi \mid r < s\} \models L_{r_1 \dots r_k} \psi$ . We do this by induction on  $k$ .

The case  $k = 0$ : Suppose that for all  $r < s$ ,  $\mathcal{M}, m, i \models L_r\varphi$ , i.e., for all  $r < s$ ,  $\theta(m)(\llbracket\varphi\rrbracket^i) \geq r$ . Using the Archimedean property of rationals, we derive that  $\theta(m)(\llbracket\varphi\rrbracket^i) \geq s$ . Hence,  $\mathcal{M}, m, i \models L_s\varphi$ .

The inductive step is similar to the soundness of the rule (R1).

**Soundness of (R3):** We have to prove that  $\{L_{r_1 \dots r_k} \psi \mid r \in \mathbb{Q}^+\} \models_\Delta L_{r_1 \dots r_k} \perp$ . Again, we use induction on  $k$ .

The case  $k = 0$ : Suppose that for all  $r > s$ ,  $\mathcal{M}, m, i \models_\Delta L_r\varphi$ , i.e., for all  $r > s$ ,  $\theta(m)(\llbracket\varphi\rrbracket^i) \geq r$ . But then,  $\theta(m)(\llbracket\varphi\rrbracket^i) = \infty$  - impossible since  $\llbracket\varphi\rrbracket^i$  is measurable and the measure is always finite. Hence, there exists no model with this property, i.e.,  $\{L_r\varphi \mid r > s\}$  is inconsistent.

The inductive step is similar to the soundness of (R1). □

## 7.5 Finite Model Property and Weak-Completeness

In what follows we prove that the Markovian logics enjoy the finite model property which states that any formula that is satisfiable admits a finite model, i.e., an MP defined over a finite state-space. This result will eventually establish the weak-completeness of the three axiomatic systems, meaning that we will succeed to prove that for an arbitrary formula  $\varphi$ , we have that

$$\models \varphi \text{ implies } \vdash \varphi.$$

The finite model property for  $\mathcal{L}(\Pi)$  against a semantics based on Harsanyi type spaces has been proven in [49]. This result implies the finite model property also for the Markovian semantics as well as for the logic  $\mathcal{L}(\Pi^*)$ .

In what follows we prove the similar result for  $\mathcal{L}(\Delta)$ . This result is neither a consequence, nor an adaptation of the proof presented in [49]. To prove the finite model property for the case of GMPs one has to handle the fact that, unlike in the probabilistic cases, the general distributions used in the definition of GMPs are not bounded by some maximal value. This complicates the entire development and makes our proof particularly interesting.

To prove the finite model property, we will construct, for an arbitrary consistent formula  $\psi \in \mathcal{L}(\Delta)$ , a model  $(\mathcal{M}_\psi, \Sigma_\psi, \theta_\psi)$  where  $\text{sup}(\mathcal{M}_\psi)$  is a finite set of  $\mathcal{L}(\Delta)$ -consistent sets of formulas, and an interpretation  $i_\psi$ . As usual with the filtration method that we use here, the key argument is the truth lemma: for  $\Gamma \in \mathcal{M}_\psi$ ,  $\psi \in \Gamma$  iff  $\mathcal{M}_\psi, \Gamma, i_\psi \models_\Delta \psi$ .

Before proceeding with the construction, we fix some notations.

For  $n \in \mathbb{N}$ ,  $n \neq 0$ , let

$$\mathbb{Q}_n = \left\{ \frac{p}{n} : p \in \mathbb{N} \right\}.$$

If  $S \subseteq \mathbb{Q}$  is finite, the *granularity* of  $S$ ,  $gr(S)$ , is the least common multiple of the denominators of the elements of  $S$ .

The *modal depth* of  $\varphi \in \mathcal{L}(\Delta)$  is defined by

- for  $p \in \mathcal{P}$ ,  $md(p) = 0$ ,
- $md(\neg\varphi) = md(\varphi)$ ,
- $md(\varphi \wedge \psi) = \max(md(\varphi), md(\psi))$
- $md(L_r\varphi) = md(\varphi) + 1$ .

The *granularity* of  $\varphi \in \mathcal{L}(\Delta)$  is  $gr(\varphi) = gr(R)$ , where  $R \subseteq \mathbb{Q}_+$  is the set of indexes  $r$  of the operators  $L_r$  present in  $\varphi$ ; the *upper bound* of  $\varphi$  is  $\max(\varphi) = \max(R)$ .

For arbitrary  $n \in \mathbb{N}$ , let  $\mathcal{L}_n$  be the sublanguage of  $\mathcal{L}(\Delta)$  that uses only modal operators  $L_r$  with  $r \in \mathbb{Q}_n$ .

For  $\Lambda \subseteq \mathcal{L}(\Delta)$ , let

$$[\Lambda]_n = \Lambda \cup \{ \varphi \in \mathcal{L}_n : \Lambda \vdash_\Delta \varphi \}.$$

Consider now a consistent formula  $\psi \in \mathcal{L}(\mathbb{A})$  that we keep fixed for the rest of the finite model construction.

Suppose that  $gr(\psi) = n$ . Let

$$\mathcal{L}[\psi] = \{\varphi \in \mathcal{L}_n \mid \max(\varphi) \leq \max(\psi), md(\varphi) \leq md(\psi)\}.$$

In what follows we construct a GMP  $\mathcal{M}_\psi$  such that each  $\Gamma \in \text{supp}(\mathcal{M}_\psi)$  is a consistent set of formulas that contains an  $\mathcal{L}[\psi]$ -maximally consistent set of formulas; and each  $\mathcal{L}[\psi]$ -maximally consistent set is contained in some  $\Gamma \in \text{supp}(\mathcal{M}_\psi)$ . And we will prove that for any  $\varphi \in \mathcal{L}[\psi]$ ,  $\varphi \in \Gamma$  iff  $\mathcal{M}_\psi, \Gamma, i_\psi \models \varphi$  for some interpretation function  $i_\psi$  that we will define as well.

Let  $\Omega[\psi]$  be the set of  $\mathcal{L}[\psi]$ -maximally consistent sets of formulas. By construction,  $\Omega[\psi]$  is finite and any  $\Lambda \in \Omega[\psi]$  contains finitely many *nontrivial formulas*<sup>1</sup>; in the rest of this construction we only count non-trivial formulas while ignoring the rest and we use  $\bigwedge \Lambda$  to denote the conjunction of the nontrivial formulas of  $\Lambda$ .

For each  $\Lambda \in \Omega[\psi]$ , such that  $\{\varphi_1, \dots, \varphi_i\} \subseteq \Lambda$  is its set of its non-trivial formulas, we construct  $\Lambda^+ \supseteq [\Lambda]_n$  with the property that  $\forall \varphi \in \Lambda$  there exists  $\neg L_r \varphi \in \Lambda^+$ .

**The construction step [ $\varphi_1$  versus  $\Lambda$ :]**

The rule (R3) guarantees that there exists  $r \in \mathbb{Q}_n$  s.t.  $[\Lambda]_n \cup \{\neg L_r^a \varphi_1\}$  is consistent.

Indeed, suppose that this is not the case, then  $\vdash \bigwedge \Lambda \rightarrow L_r \varphi_1$  for all  $r \in \mathbb{Q}_n$  implying that  $\bigwedge \Lambda$  is inconsistent - impossible.

Let

$$y_1 = \min\{s \in \mathbb{Q}_n \mid [\Lambda]_n \cup \{\neg L_s \varphi_1\} \text{ is consistent}\}$$

and

$$x_1 = \max\{s \in \mathbb{Q}_n \mid L_s \varphi_1 \in [\Lambda]_n\}.$$

As before, (R3) guarantees the existence of  $\max$ , because otherwise  $\vdash \bigwedge \Lambda \rightarrow L_r \varphi_1$  for all  $r$  implying  $\bigwedge \Lambda$  inconsistent - impossible.

The rule (R2) implies that there exists  $r \in \mathbb{Q} \setminus \mathbb{Q}_n$  such that  $x_1 < r < y_1$  and  $\{\neg L_r \varphi_1\} \cup [\Lambda]_n$  is consistent.

Indeed, suppose otherwise, then  $\vdash \bigwedge \Lambda \rightarrow L_r \varphi_1$  for all  $r < y_1$  and due to (R2),  $\vdash \bigwedge \Lambda \rightarrow L_{y_1} \varphi_1$  - contradiction with the consistency of  $\Lambda$ .

Obviously,  $r \notin \mathbb{Q}_n$ .

Let  $n_1 = \text{gran}\{1/n, r\}$ . And let

$$s_1 = \min\{s \in \mathbb{Q}_{n_1} \mid [\Lambda]_{n_1} \cup \{\neg L_s \varphi_1\} \text{ is consistent}\},$$

and

$$\Lambda_1 = \Lambda \cup \{\neg L_{s_1} \varphi_1\}$$

---

<sup>1</sup>By nontrivial formulas we mean the formulas that are not obtained from more basic consistent ones by boolean derivations, for instance,  $p \vee q \rightarrow p$ ,  $p \wedge p$ ,  $p \vee p$  are trivial formulas. The nontrivial formulas are (isomorphic to) elements of the Lindenbaum algebra  $\mathcal{L}[\psi]^{\leftrightarrow}$ , which is the quotient of  $\mathcal{L}[\psi]$  with respect to logical equivalence.

**The construction step [ $\varphi_2$  versus  $\Lambda_1$ ]:**

As before, let

$$y_2 = \min\{s \in \mathbb{Q}_{n_1} \mid [\Lambda_1]_{n_1} \cup \{\neg L_s \varphi_2\} \text{ is consistent}\}$$

and

$$x_2 = \max\{s \in \mathbb{Q}_{n_1} \mid L_s \varphi_2 \in [\Lambda_1]_{n_1}\}.$$

There exists  $r \in \mathbb{Q} \setminus \mathbb{Q}_{n_1}$  such that  $x_2 < r < y_2$  and  $\{\neg L_r \varphi_2\} \cup [\Lambda_1]_{n_1}$  is consistent.

Let  $n_2 = \text{gran}\{1/n_1, r\}$ . and let

$$s_2 = \min\{s \in \mathbb{Q}_{n_2} : [\Lambda]_{n_2} \cup \{\neg L_s \varphi_2\} \text{ is consistent}\}$$

and

$$\Lambda_2 = \Lambda_1 \cup \{\neg L_{s_2} \varphi_2\}.$$

**The complete construction:**

We repeat this construction step for  $[\varphi_3$  versus  $\Lambda_2], \dots, [\varphi_i$  versus  $\Lambda_{i-1}]$  and in a finite number of steps we eventually obtain  $\Lambda \subseteq \Lambda_1 \subseteq \dots \subseteq \Lambda_i$ , where  $\Lambda_i$  is a consistent set containing a finite set of nontrivial formulas.

Let  $n_\Lambda = \text{gran}\{1/n_1, \dots, 1/n_i\}$ . We make this construction for all  $\Lambda \in \Omega[\psi]$ .

Let  $p = \text{gran}\{1/n_\Lambda \mid \Lambda \in \Omega[\psi]\}$ . Notice that  $p > n$ .

Let

$$\Lambda^+ = [\Lambda_i]_p \quad \text{and} \quad \Omega^+[\psi] = \{\Lambda^+ \mid \Lambda \in \Omega[\psi]\}.$$

**Remark 7.5.1.** Since the axiomatization of  $\mathcal{L}(\Delta)$  is countable<sup>2</sup>, any consistent formula  $\varphi \in \mathcal{L}[\psi]$  is an element of a set  $\Lambda^+ \in \Omega^+[\psi]$ , due to Rasiowa-Sikorski lemma used in the form of Lemma 3.7.1. For each  $\Lambda \in \Omega[\psi]$  and each  $\varphi \in \Lambda$ , there exist  $s, t \in \mathbb{Q}_p$ ,  $s < t$ , such that  $L_s \varphi, \neg L_t \varphi \in \Lambda^+$ . Moreover, since  $\Lambda^+$  is finite modulo logical equivalence, for any  $\Lambda^+$  there exists a formula  $\rho$  such that  $\varphi \in \Lambda^+$  iff  $\vdash \rho \rightarrow \varphi$ .

Let  $\Omega_p$  be the set of  $\mathcal{L}_p$ -maximally consistent sets of formulas. Since  $\mathcal{L}(\Delta)$  (hence,  $\mathcal{L}_p$ ) is countably axiomatized, we can apply the Extended Rasiowa-Sikorski Lemma 3.7.2, which proves the Lindenbaum's Lemma 3.7.3 (see the Preliminaries for details). Consequently, for each  $\Lambda^+ \in \Omega^+[\psi]$  there exists a  $\mathcal{L}_p$ -maximal consistent set that contains  $\Lambda^+$ .

We fix an injective (choice) function

$$f : \Omega^+[\psi] \rightarrow \Omega_p$$

such that for any  $\Lambda^+ \in \Omega^+[\psi]$ ,  $\Lambda^+ \subseteq f(\Lambda^+)$ . We denote by  $\Omega_p[\psi] = f(\Omega^+[\psi])$ .

For  $\varphi \in \mathcal{L}[\psi]$ , let

$$[\varphi] = \{\Gamma \in \Omega_p[\psi] \mid \varphi \in \Gamma\}.$$

Anticipating the further construction, we will use  $\Omega_p[\psi]$  as the support-set for  $\mathcal{M}_\psi$ . For this reason we establish some properties for this set.

<sup>2</sup>Since the modal operators of type  $L_r$  only use rational indices all the axioms and rules, including (R2) and (R3), have countably many instances.

**Lemma 7.5.2.** 1.  $\Omega_p[\psi]$  is finite.

$$2. 2^{\Omega_p[\psi]} = \{[\varphi] \mid \varphi \in \mathcal{L}[\psi]\}.$$

3. For any  $\varphi_1, \varphi_2 \in \mathcal{L}[\psi]$ ,  $\vdash \varphi_1 \rightarrow \varphi_2$  iff  $[\varphi_1] \subseteq [\varphi_2]$ .

4. For any  $\Gamma \in \Omega_p[\psi]$  and  $\varphi \in \mathcal{L}[\psi]$ , there exist

$$x = \max\{r \in \mathbf{Q}_p : L_r \varphi \in \Gamma\}, \quad y = \min\{r \in \mathbf{Q}_p : \neg L_r \varphi \in \Gamma\}$$

and  $y = x + 1/p$ .

*Proof.* 1. It has already been discussed in Remark 7.5.1.

2. Since any element of  $\Omega_p[\psi]$  can be characterized by a formula in  $\mathcal{L}[\psi]$  (see Remark 7.5.1), i.e., it is a principal filter, and because  $\Omega_p[\psi]$  is finite, any subset of it can also be characterized by a formula in  $\mathcal{L}[\psi]$  – for instance by the disjunction of the formulas that characterize each element of the subset.

3. ( $\Rightarrow$ ) Since  $\vdash \varphi_1 \rightarrow \varphi_2$  and  $\varphi_1, \varphi_2 \in \mathcal{L}[\psi]$ , any  $\mathcal{L}[\psi]$ -maximal consistent set that contains  $\varphi_1$  must also contain  $\varphi_2$  since it is closed under deduction of  $\mathcal{L}[\psi]$ -formulas.

( $\Leftarrow$ ) Suppose that all the  $\mathcal{L}[\psi]$ -maximal consistent sets that contain  $\varphi_1$  also contain  $\varphi_2$ . Because  $\varphi_1 \rightarrow \varphi_2 \in \mathcal{L}[\psi]$ , we obtain that all the  $\mathcal{L}[\psi]$ -maximal consistent sets must contain  $\varphi_1 \rightarrow \varphi_2$ . Indeed, suppose that there exists a  $\mathcal{L}[\psi]$ -maximal consistent set that does not contain  $\varphi_1 \rightarrow \varphi_2$ . Since it is maximal, it must contain its negations, which is  $\varphi_1 \wedge \neg \varphi_2$ , i.e. this set contains  $\varphi_1$  but not  $\varphi_2$  – contradiction with our hypothesis.

Since any  $\mathcal{L}[\psi]$ -maximal consistent set is  $\mathcal{L}(\Delta)$  consistent, applying Lindenbaum's lemma 3.7.3, we obtain that any  $\mathcal{L}[\psi]$ -maximal consistent set has a  $\mathcal{L}[\Delta]$ -maximal consistent extension. This property guarantees that if  $\varphi_1 \rightarrow \varphi_2$  is contained in all  $\mathcal{L}[\psi]$ -maximal consistent sets, then  $\varphi_1 \rightarrow \varphi_2$  is also contained in all  $\mathcal{L}[\Delta]$ -maximal consistent sets. Indeed, suppose that there exists a  $\mathcal{L}[\Delta]$ -maximal consistent set that does not contain  $\varphi_1 \rightarrow \varphi_2$ . Then, if we consider the intersection of this set with  $\mathcal{L}[\psi]$  we obtain a  $\mathcal{L}[\psi]$ -maximal consistent set that does not contain  $\varphi_1 \rightarrow \varphi_2$  – contradiction.

Hence,  $\varphi_1 \rightarrow \varphi_2$  is contained in all the  $\mathcal{L}[\Delta]$ -maximal consistent sets. Then it must be a  $\vdash_{\Delta}$ -theorem. Hence  $\vdash \varphi_1 \rightarrow \varphi_2$ .

4.  $L_x \varphi, \neg L_y \varphi \in \Gamma$  implies  $x \neq y$ . If  $x > y$ ,  $L_x \varphi \in \Gamma$  entails (Axiom (A2))  $L_y \varphi \in \Gamma$ , contradicting the consistency of  $\Gamma$ . If  $x + 1/p < y$ , then  $L_{x+1/p} \varphi \notin \Gamma$ , i.e.  $\neg L_{x+1/p} \varphi \in \Gamma$  implying that  $x + 1/p \geq y$  – contradiction.  $\square$

Let  $\Omega$  be the set of  $\mathcal{L}(\Delta)$ -maximally consistent sets of formulas. Since the deducibility relation on  $\mathcal{L}(\Delta)$  is countably axiomatized, we can apply again the Lindenbaum's Lemma 3.7.3 to prove that each  $\Gamma \in \Omega_p$  can be extended to a  $\mathcal{L}(\Delta)$ -maximal consistent set that contains.

We fix an injective (choice) function

$$g : \Omega_p \rightarrow \Omega$$

such that for any  $\Gamma \in \Omega_p$ ,  $\Gamma \subseteq \pi(g(\Gamma))$ ; we denote  $g(\Gamma)$  by  $\Gamma^\infty$ .

**Lemma 7.5.3.** *For any  $\Gamma \in \Omega_p[\psi]$  and any  $\varphi \in \mathcal{L}[\psi]$ , there exists*

$$z = \sup\{r \in \mathbb{Q} : L_r\varphi \in \Gamma^\infty\} = \inf\{r \in \mathbb{Q} : \neg L_r\varphi \in \Gamma^\infty\}$$

and  $x \leq z < y$ , where  $x$  and  $y$  are as defined in Lemma 7.5.2.

*Proof.* Let

$$x^\infty = \sup\{r \in \mathbb{Q} : L_r\varphi \in \Gamma^\infty\} \quad \text{and} \quad y^\infty = \inf\{r \in \mathbb{Q} : \neg L_r\varphi \in \Gamma^\infty\}.$$

Suppose that  $x^\infty < y^\infty$ . Then, there exists  $r \in \mathbb{Q}$  such that  $x^\infty < r < y^\infty$ . From  $x^\infty < r$  we obtain  $\neg L_r\varphi \in \Gamma^\infty$ . From  $r < y^\infty$  we obtain  $L_r\varphi \in \Gamma^\infty$ . Hence,  $\neg L_r\varphi, L_r\varphi \in \Gamma^\infty$  - impossible because  $\Gamma^\infty$  is consistent.

Suppose that  $x^\infty > y^\infty$ . Then there exists  $r \in \mathbb{Q}$  such that  $x^\infty > r > y^\infty$ . As  $\Gamma^\infty$  is maximally consistent we have either  $L_r\varphi \in \Gamma^\infty$  or  $\neg L_r\varphi \in \Gamma^\infty$ . The first case contradicts the definition of  $x^\infty$  while the second the definition of  $y^\infty$ .

Hence,  $x \leq z \leq y$ .

If  $z = y$ , then  $z \in \mathbb{Q}_p$  implying  $L_z\varphi, \neg L_z\varphi \in \Gamma$  - contradiction with the consistency of  $\Gamma$ .  $\square$

We denote  $z$  as defined in Lemma 7.5.3 by  $a_\varphi^\Gamma$  and now we can proceed with the definition of  $\mathcal{M}_\psi$ .

**Lemma 7.5.4.** *If  $\theta_\psi : \Omega_p[\psi] \rightarrow \Delta(\Omega_p[\psi], 2^{\Omega_p[\psi]})$  is defined, for arbitrary  $\Gamma \in \Omega_q[\psi]$  and  $\varphi \in \mathcal{L}[\psi]$  by*

$$\theta_\psi(\Gamma)(\llbracket \varphi \rrbracket) = a_\varphi^\Gamma,$$

then  $\mathcal{M}_\psi = (\Omega_p[\psi], 2^{\Omega_p[\psi]}, \theta_\psi)$  is a GMP.

*Proof.* The central problem is to prove that for arbitrary  $\Gamma \in \Omega_p[\psi]$ , the function  $\theta_\psi(\Gamma) : 2^{\Omega_p[\psi]} \rightarrow \mathbb{R}^+$  is well defined and a measure on  $(\Omega_p[\psi], 2^{\Omega_p[\psi]})$ . Further, because the space is discrete with finite support, we obtain that  $\theta_\psi$  is indeed measurable.

Suppose that for  $\varphi_1, \varphi_2 \in \mathcal{L}[\psi]$  we have  $\llbracket \varphi_1 \rrbracket = \llbracket \varphi_2 \rrbracket$ . Then, from Lemma 7.5.2,  $\vdash \varphi_1 \leftrightarrow \varphi_2$  and applying (R1),  $\vdash L_r\varphi_1 \leftrightarrow L_r\varphi_2$ . Hence,  $a_{\varphi_1}^\Gamma = a_{\varphi_2}^\Gamma$  proving that  $\theta_\psi(\Gamma)$  is well defined.

Now we prove that  $\theta_\psi(a)(\Gamma)$  is a measure.

For showing  $\theta_\psi(\Gamma)(\emptyset) = 0$ , we show that for any  $r > 0$ ,  $\vdash \neg L_r\perp$ . This is sufficient, as (A1) guarantees that  $\vdash L_0\perp$  and  $\llbracket \perp \rrbracket = \emptyset$ . Suppose that there exists  $r > 0$  such that  $L_r\perp$  is consistent. Let  $\varepsilon \in (0, r) \cap \mathbb{Q}$ . Then (A2) gives  $\vdash L_r\perp \rightarrow L_\varepsilon\perp$ . Hence,

$$\vdash L_r\perp \rightarrow (L_r(\perp \wedge \perp) \wedge L_\varepsilon(\perp \wedge \neg\perp));$$

and applying (A3),  $\vdash L_r\perp \rightarrow L_{r+\varepsilon}\perp$ . Repeating this argument, we can prove that  $\vdash L_r\perp \rightarrow L_s\perp$  for any  $s$  and (R3) confirms the inconsistency of  $L_r\perp$ .

We show now that if  $A, B \in 2^{\Omega_p[\psi]}$  with  $A \cap B = \emptyset$ , then

$$\theta_\psi(\Gamma)(A) + \theta_\psi(\Gamma)(B) = \theta_\psi(\Gamma)(A \cup B).$$

Let  $A = \llbracket \varphi_1 \rrbracket$ ,  $B = \llbracket \varphi_2 \rrbracket$  with  $\varphi_1, \varphi_2 \in \mathcal{L}[\psi]$  and  $\vdash \varphi_1 \rightarrow \neg \varphi_2$ .

Let  $x_1 = \theta_\psi(\Gamma)(A)$ ,  $x_2 = \theta_\psi(\Gamma)(B)$  and  $x = \theta_\psi(\Gamma)(A \cup B)$ . We prove that  $x_1 + x_2 = x$ .

Suppose that  $x_1 + x_2 < x$ . Then, there exist  $\varepsilon_1, \varepsilon_2 \in \mathbb{Q}^+$  such that  $x'_1 + x'_2 < x$ , where  $x'_i = x_i + \varepsilon_i$  for  $i = 1, 2$ .

From the definition of  $x_i$ ,  $\neg L_{x'_i} \varphi_i \in \Gamma^\infty$ . Further, using (A4), we obtain  $\neg L_{x'_1 + x'_2}(\varphi_1 \vee \varphi_2) \in \Gamma^\infty$ , implying that  $x'_1 + x'_2 \geq x$  - contradiction.

Suppose now that  $x_1 + x_2 > x$ . Then, there exist  $\varepsilon_1, \varepsilon_2 \in \mathbb{Q}^+$  such that  $x''_1 + x''_2 > x$ , where  $x''_i = x_i - \varepsilon_i$  for  $i = 1, 2$ .

The definition of  $x_i$  implies that  $L_{x''_i} \varphi_i \in \Gamma^\infty$ . Further, (A3) gives  $L_{x''_1 + x''_2}(\varphi_1 \vee \varphi_2) \in \Gamma^\infty$ , i.e.  $x''_1 + x''_2 \leq x$  - contradiction. □

Consequently,  $\mathcal{M}_\psi$  is indeed a general Markov process. Moreover, it has a finite state space. In what follows we prove the Truth Lemma for  $\mathcal{M}_\psi$  that states that any consistent formula in  $\mathcal{L}[\psi]$  is satisfied by a state of  $\mathcal{M}_\psi$ .

**Lemma 7.5.5 (Truth Lemma).** *Let  $\psi \in \mathcal{L}(\Delta)$  be a  $\vdash_\Delta$ -consistent formula. For arbitrary  $\varphi \in \mathcal{L}[\psi]$  and  $\Gamma \in \text{supp}(\mathcal{M}_\psi)$ ,*

$$\mathcal{M}_\psi, \Gamma, i_\psi \models \varphi \text{ iff } \varphi \in \Gamma,$$

where  $i_\psi : \text{supp}(\mathcal{M}_\psi) \rightarrow 2^{\mathcal{P}}$  is defined by  $i_\psi(\Gamma) = \Gamma \cap \mathcal{P}$ .

*Proof.* Induction on the structure of  $\varphi$ .

The only nontrivial case is  $\varphi = L_r \varphi'$ .

( $\implies$ ) Suppose that  $\mathcal{M}_\psi, \Gamma, i_\psi \models \varphi$  and  $\varphi \notin \Gamma$ . Hence  $\neg \varphi \in \Gamma$ . Let

$$y = \min\{r \in \mathbb{Q}_p \mid \neg L_r \varphi \in \Gamma\}.$$

Then, from  $\neg L_r \varphi' \in \Gamma$ , we obtain  $r \geq y$ .

But  $\mathcal{M}_\psi, \Gamma, i_\psi \models L_r \varphi'$  is equivalent to  $\theta_\psi(\Gamma)(\llbracket \varphi' \rrbracket) \geq r$ , i.e.  $a_{\varphi'}^\Gamma \geq r$ .

On the other hand, from Lemma 7.5.2,  $a_{\varphi'}^\Gamma < y$  - contradiction.

( $\impliedby$ ) If  $L_r \varphi' \in \Gamma$ , then  $r \leq a_{\varphi'}^\Gamma$  and  $r \leq \theta_\psi(\Gamma)(\llbracket \varphi \rrbracket)$ . Hence,  $\mathcal{M}_\psi, \Gamma, i_\psi \models L_r \varphi'$ . □

A consequence of the previous lemma is the finite model property for our logic.

**Theorem 7.5.6 (Finite model property).** *For any  $\mathcal{L}(\Delta)$ -consistent formula  $\varphi$ , there exists an GMP  $\mathcal{M}$  with finite support of cardinality bound by the structure of  $\varphi$ , there exists  $m \in \text{supp}\mathcal{M}$  and an interpretation function  $i$  such that*

$$\mathcal{M}, m, i \models \varphi.$$

Obviously, the model of which the previous theorem states is  $\mathcal{M}_\varphi$ , the interpretation is  $i_\varphi$  and  $m$  is just any  $\mathcal{L}[\varphi]$ -maximal consistent set that contains  $\varphi$  – Rasiowa-Sikorski lemma guarantees that such a set exists.

The finite model property proves the weak-completeness of the axiomatic system, i.e., we can now prove that for an arbitrary  $\varphi \in \mathcal{L}(\Delta)$ , we have that if  $\varphi$  is satisfied by all the models then it is provable.

**Theorem 7.5.7** (Weak Completeness). *The axiomatic system of  $\mathcal{L}(\Delta)$  is weakly-complete with respect to the Markovian semantics, i.e.*

$$\models_{\Delta} \varphi \text{ implies } \vdash_{\Delta} \varphi.$$

*Proof.* We have that

$$[\models_{\Delta} \psi \text{ implies } \vdash_{\Delta} \psi]$$

is equivalent with

$$[\not\models_{\Delta} \psi \text{ implies } \not\vdash_{\Delta} \psi],$$

that is equivalent to

[the  $\vdash_{\Delta}$ -consistency of  $\neg\psi$  implies the existence of a model  $(\mathcal{M}, m)$  for  $\neg\psi$ ] and this is exactly what the finite model property guarantees.  $\square$

These entire arguments can be reproduced both for the case of probabilistic and sub-probabilistic Markovian logics where, in fact, some steps of the finite model construction are much simpler. As in the case of the general Markovian logic, we will obtain that the axiomatizations proposed before for sub-probabilistic and probabilistic logics are weakly-complete.

## 7.6 Canonical Models and Strong Completeness

In this section we construct canonical models for the three logics. The canonical model for a logic  $\mathcal{L} \in \{\mathcal{L}(\Pi), \mathcal{L}(\Pi^*), \mathcal{L}(\Delta)\}$  is a Markov processes  $\mathcal{M}_{\mathcal{L}} = (\mathcal{U}_{\mathcal{L}}, \Sigma_{\mathcal{L}}, \theta_{\mathcal{L}})$  having the set  $\mathcal{U}_{\mathcal{L}}$  of  $\mathcal{L}$ -maximally consistent sets of formulas as the state space and satisfying the property that for any  $\varphi \in \mathcal{L}$  and  $u \in \mathcal{U}_{\mathcal{L}}$ ,  $\mathcal{M}_{\mathcal{L}}, u, i_{\mathcal{L}} \models \varphi$  iff  $\varphi \in u$ , where  $i_{\mathcal{L}}$  is an appropriate interpretation function.

With respect to the construction proposed in the previous section, the construction of the canonical model is universal, in the sense that it does not depend on one formula only, but it applies uniformly to any formula. Obviously this also means that the Markov process that we will construct will not be finite any more.

In order to complete such a construction, we have to:

- prove that  $\mathcal{U}_{\mathcal{L}} \neq \emptyset$  for each  $\mathcal{L} \in \{\mathcal{L}(\Pi), \mathcal{L}(\Pi^*), \mathcal{L}(\Delta)\}$ ;
- define  $\Sigma_{\mathcal{L}}$  such that  $(\mathcal{U}_{\mathcal{L}}, \Sigma_{\mathcal{L}})$  is an analytic space;

- define a measure  $\theta_{\mathcal{L}}$  on  $(\mathcal{U}_{\mathcal{L}}, \Sigma_{\mathcal{L}})$ ;
- define an interpretation function  $i_{\mathcal{L}}$  such that for any  $\varphi \in \mathcal{L}$ ,  $\llbracket \varphi \rrbracket_{\mathcal{M}_{\mathcal{L}}}^{i_{\mathcal{L}}} \in \Sigma_{\mathcal{L}}$ ;
- and prove the Truth Lemma stating that  $\mathcal{M}_{\mathcal{L}}, u, i_{\mathcal{L}} \models \varphi$  iff  $\varphi \in u$ .

If realized, this construction will allow us to prove that the axiomatic systems proposed before for the three Markovian logics are not only weak-complete, but also strongly-complete, i.e., for an arbitrary set  $\Phi \subseteq \mathcal{L}$  of formulas and an arbitrary formula  $\varphi \in \mathcal{L}$ , if all the models of (all the formulas of)  $\Phi$  are also models of  $\varphi$ , written  $\Phi \models \varphi$ , then from the formulas in  $\Phi$  as hypothesis we can prove  $\varphi$ , written  $\Phi \vdash \varphi$ .

In general (for the compact logics) the weak and the strong completeness are equivalent; this is the case with propositional logic or with normal modal logic. However, if the logic is not compact, meaning that there exists a set  $S$  of formulas that admits no model even if each finite subset of  $S$  has a model, then the weak and strong completeness are not necessarily equivalent.

This is the case of Markovian logics: take, for instance, the set

$$S = \{L_r \varphi \mid r < s\} \cup \{\neg L_s \varphi\}$$

defined for a fixed rational  $s$  and for some consistent formula  $\varphi$ . Obviously, any finite subset of  $S$  has a model: it is sufficient to consider a rational between the maximum value of  $r$  in this set and  $s$ , and to look at a model that can do a transition with that value to  $\varphi$ . However, due to the rule (R2), the set  $S$  admits no model because from the soundness of (R2) we know that

$$\{L_r \varphi \mid r < s\} \models L_s \varphi,$$

hence, any model of  $\{L_r \varphi \mid r < s\}$  is also a model for  $L_s \varphi$ . And this proves that all the Markovian logics are not compact.

The fact they are not compact also explains why for the complete axiomatization we had to add infinitary axioms. Because a logic with a finitary weak-complete axiomatization is always compact. Indeed, in a finitary logic all the theorems have a finite proof. So, if  $S$  has no model, this means, due to completeness, that  $S \vdash \perp$ . But since all the proofs must be finite, there exists a finite proof for  $\perp$  using the formulas in  $S$  as hypothesis. Since the axioms are finite, this implies that there exists a finite subset  $S' \subset S$  such that  $S' \vdash \perp$ . Now the soundness gives us that  $S'$  cannot have a model. Hence,  $S$  has a finite subset without models. And this proves that the logic is compact.

In what follows we proceed with the construction of the canonical models.

**Lemma 7.6.1.** *For  $\mathcal{L} \in \{\mathcal{L}(\Pi), \mathcal{L}(\Pi^*), \mathcal{L}(\Delta)\}$  with the proof systems previously defined, the set  $\mathcal{U}_{\mathcal{L}}$  of  $\mathcal{L}$ -maximally consistent sets is nonempty.*

*Proof.* Note that in each case  $\mathcal{L}$  forms a Boolean algebra and the instances of all the axioms and rules define a countable family of subsets of  $\mathcal{L}$ , each member of which has the meet in  $\mathcal{L}$ . In particular, the instances of (R2) define the subsets  $\{L_{r_1 \dots r_k s} \psi \mid r < s\}$  of  $\mathcal{L}$  each having the meet  $L_{r_1 \dots r_k s} \psi \in \mathcal{L}$ .

Observe also that a set  $u \subseteq \mathcal{L}$  is a  $\mathcal{L}$ -maximally consistent set iff it is a boolean ultrafilter that respects all the instances of the axioms of  $\mathcal{L}$ . Consequently, the Rasiowa-Sikorski Lemma guarantees that  $\mathcal{U}_{\mathcal{L}} \neq \emptyset$ .  $\square$

Consider the set

$$(\mathcal{L}) = \{(\varphi) \mid \varphi \in \mathcal{L}\}, \quad \text{where } (\varphi) = \{u \in \mathcal{U}_{\mathcal{L}} \mid \varphi \in u\}.$$

Using this, we define  $\Sigma_{\mathcal{L}} = \sigma((\mathcal{L}))$ . The space  $(\mathcal{U}_{\mathcal{L}}, \Sigma_{\mathcal{L}})$  is an analytic space for each  $\mathcal{L} \in \{\mathcal{L}(\Pi), \mathcal{L}(\Pi^*), \mathcal{L}(\Delta)\}$ . This proof will be detailed later, in the Part IV of this monograph as it requires complex mathematical machinery that we will develop in the context of Stone duality. For now we will only state this result.

**Lemma 7.6.2.** *For  $\mathcal{L} \in \{\mathcal{L}(\Pi), \mathcal{L}(\Pi^*), \mathcal{L}(\Delta)\}$ ,  $(\mathcal{U}_{\mathcal{L}}, \Sigma_{\mathcal{L}})$  is an analytic space.*

The next step in our construction is to define an appropriate measure  $\theta_{\mathcal{L}}$  on  $(\mathcal{U}_{\mathcal{L}}, \Sigma_{\mathcal{L}})$ . To do this we prove the following lemma.

**Lemma 7.6.3.** 1. *For arbitrary  $u \in \mathcal{U}_{\mathcal{L}(\Pi)}$  and  $\varphi \in \mathcal{L}(\Pi)$ , or  $u \in \mathcal{U}_{\mathcal{L}(\Pi^*)}$  and  $\varphi \in \mathcal{L}(\Pi^*)$ ,*

$$x_u^\varphi = \sup\{r \in \mathbb{Q}_0 \mid L_r \varphi \in u\} = \inf\{r \in \mathbb{Q}_0 \mid \neg L_r \varphi \in u\}.$$

*Moreover, if  $x_u^\varphi \in \mathbb{Q}$ , then  $L_{x_u^\varphi} \varphi \in u$ .*

2. *For arbitrary  $u \in \mathcal{U}_{\mathcal{L}(\Delta)}$  and  $\varphi \in \mathcal{L}(\Delta)$ ,*

$$x_u^\varphi = \sup\{r \in \mathbb{Q}^+ \mid L_r \varphi \in u\} = \inf\{r \in \mathbb{Q}^+ \mid \neg L_r \varphi \in u\} \in \mathbb{R}^+.$$

*Moreover, if  $x_u^\varphi \in \mathbb{Q}$ , then  $L_{x_u^\varphi} \varphi \in u$ .*

*Proof.* 1. This proof is similar to the proof of Lemma 7.5.3. Let

$$x^\infty = \sup\{r \in \mathbb{Q}^+ \mid L_r \varphi \in u\} \quad \text{and} \quad y^\infty = \inf\{r \in \mathbb{Q}^+ \mid \neg L_r \varphi \in u\}.$$

Suppose that  $x^\infty < y^\infty$ . Then, there exists  $r \in \mathbb{Q}^+$  such that  $x^\infty < r < y^\infty$ . From  $x^\infty < r$  we obtain  $\neg L_r \varphi \in u$ . From  $r < y^\infty$  we obtain  $L_r \varphi \in u$ . Hence,  $\neg L_r \varphi, L_r \varphi \in u$  - impossible because  $u$  is consistent.

Suppose that  $x^\infty > y^\infty$ . Then there exists  $r \in \mathbb{Q}^+$  such that  $x^\infty > r > y^\infty$ . As  $u$  is maximally consistent, we have either  $L_r \varphi \in u$  or  $\neg L_r \varphi \in u$ . The first case contradicts the definition of  $x^\infty$  while the second the definition of  $y^\infty$ .

Hence,  $x^\infty = y^\infty = x_u^\varphi$ .

2. This result can be proven similarly to the previous case. However, to show that  $x_u^\varphi \in \mathbb{R}^+$ , we use (R3) that guarantees the existence of a finite upper bound for

$$\{r \in \mathbb{Q}^+ \mid L_r \varphi \in u\}.$$

□

The previous lemma allows us to define, for each  $\mathcal{L} \in \{\mathcal{L}(\Pi), \mathcal{L}(\Pi^*), \mathcal{L}(\Delta)\}$  and arbitrary  $u \in \mathcal{U}_{\mathcal{L}}$ ,  $\varphi \in \mathcal{L}$ ,

$$\theta_{\mathcal{L}}(u)(\langle \varphi \rangle) = \sup\{r \in \mathbb{Q}^+ \mid L_r \varphi \in u\}.$$

Obviously,  $\theta_{\mathcal{L}}(u)$  is a set function defined on the field  $\langle \mathcal{L} \rangle$  and Theorem 3.2.3 ensures us that it can be uniquely extended to a measure on  $\Sigma_{\mathcal{L}}$  if it is finitely additive and countable subadditive on  $\langle \mathcal{L} \rangle$ . This is what we prove next.

**Lemma 7.6.4.** *For all  $u \in \mathcal{U}_{\mathcal{L}}$ , the function  $\theta_{\mathcal{L}}(u)$  previously defined is finitely additive.*

*Proof.* Suppose  $\varphi, \psi \in \mathcal{L}$  and  $\langle \varphi \rangle \cap \langle \psi \rangle = \emptyset$ . Then  $\vdash \varphi \wedge \psi \rightarrow \perp$ .

We wish to show that

$$\theta_{\mathcal{L}}(u)(\langle \varphi \vee \psi \rangle) = \theta_{\mathcal{L}}(u)(\langle \varphi \rangle) + \theta_{\mathcal{L}}(u)(\langle \psi \rangle).$$

It suffices to show the inequality in both directions. For  $\leq$ , by the definition of  $\theta$ , it suffices to show

$$\begin{aligned} & \sup\{t \mid L_t(\varphi \vee \psi) \in u\} \\ & \leq \inf\{r \mid \neg L_r \varphi \in u\} + \inf\{s \mid \neg L_s \psi \in u\} \\ & = \inf\{r + s \mid \neg L_r \varphi \in u \text{ and } \neg L_s \psi \in u\} \\ & = \inf\{r + s \mid \neg L_r \varphi \wedge \neg L_s \psi \in u\}; \end{aligned}$$

that is, if  $L_t(\varphi \vee \psi) \in u$  and  $\neg L_r \varphi \wedge \neg L_s \psi \in u$ , then  $t \leq r + s$ . But

$$\begin{aligned} & \vdash \neg L_r \varphi \wedge \neg L_s \psi \leftrightarrow \neg L_r((\varphi \vee \psi) \wedge \varphi) \wedge \neg L_s((\varphi \vee \psi) \wedge \neg \varphi) \text{ and} \\ & \vdash \neg L_r((\varphi \vee \psi) \wedge \varphi) \wedge \neg L_s((\varphi \vee \psi) \wedge \neg \varphi) \rightarrow \neg L_{r+s}(\varphi \vee \psi) \quad \text{by (A5),} \end{aligned}$$

thus  $\neg L_{r+s}(\varphi \vee \psi) \in u$ , and necessarily  $t \leq r + s$ .

The inequality in the opposite direction is similar, using (A4). We need to show

$$\inf\{t \mid \neg L_t(\varphi \vee \psi) \in u\} \geq \sup\{r + s \mid L_r \varphi \wedge L_s \psi \in u\};$$

that is, if  $\neg L_t(\varphi \vee \psi) \in u$  and  $L_r \varphi \wedge L_s \psi \in u$ , then  $t \geq r + s$ . But

$$\begin{aligned} & \vdash L_r \varphi \wedge L_s \psi \leftrightarrow L_r((\varphi \vee \psi) \wedge \varphi) \wedge L_s((\varphi \vee \psi) \wedge \neg \varphi) \text{ and} \\ & \vdash L_r((\varphi \vee \psi) \wedge \varphi) \wedge L_s((\varphi \vee \psi) \wedge \neg \varphi) \rightarrow L_{r+s}(\varphi \vee \psi) \quad \text{by (A4),} \end{aligned}$$

thus  $L_{r+s}(\varphi \vee \psi) \in u$  implying  $r + s \leq t$ . □

Now we prove that the function  $\theta_{\mathcal{L}}(u)$  is also countable subadditive and this is a central result of this section where we make use of (R2).

In related papers [26, 51], to prove a similar result a so-called countable additivity rule (CAR) was used and, in addition, Lindenbaum's lemma was used as a meta-axiom. This is because the rule (CAR) does not allow us to apply the extended Rasiowa-Sikorski lemma to prove Lindenbaum's lemma.

The main technical lemma that lies at the heart of the construction below is proved in our paper on Stone duality [32] for the probabilistic case. The proof can be similarly done for the other two cases.

**Lemma 7.6.5.** *For  $u \in \mathcal{U}_{\mathcal{L}}$ , the function  $\theta_{\mathcal{L}}(u)$  is countably subadditive.*

*Proof.* To prove that  $\theta_{\mathcal{L}}$  is countably subadditive, since we know that it is finitely additive, it is sufficient to prove that it is continuous from above at  $\emptyset$  and apply Theorem 3.2.2. Hence, we need to show that if  $u \in \mathcal{U}_{\mathcal{L}}$  and  $\langle \psi_0 \rangle \geq \langle \psi_1 \rangle \geq \dots$  with  $\bigcap_i \langle \psi_i \rangle = \emptyset$ , then

$$\inf_i \theta_{\mathcal{L}}(u)(\langle \psi_i \rangle) = 0.$$

1. We prove this firstly for  $\mathcal{L}(\Pi)$  and  $\mathcal{L}(\Pi^*)$ .

Consider the countable set  $\mathcal{F}$  of elements of the form  $\alpha^r = L_{t_1 \dots t_n} \varphi$  for  $\varphi \in \mathcal{L}$  and rational  $t_1, \dots, t_n, r \geq 0$ , parameterized by  $r$ . If  $r < s$ , then  $\vdash \alpha^s \rightarrow \alpha^r$ .

Using (A4),

$$\theta_{\mathcal{L}}(u)(\langle \alpha^r \wedge \neg \alpha^s \rangle) \leq \theta_{\mathcal{L}}(u)(\langle \alpha^r \rangle) - \theta(u)(\langle \alpha^s \rangle). \quad (7.6.1)$$

Since  $L_t \alpha^r \in u$  for all  $r < s$  iff  $L_t \alpha^s \in u$ , therefore

$$\theta_{\mathcal{L}}(u)(\langle \alpha^s \rangle) = \inf_{r < s} \theta_{\mathcal{L}}(u)(\langle \alpha^r \rangle). \quad (7.6.2)$$

Let  $\varepsilon > 0$  be an arbitrarily small positive number. For each  $\alpha \in \mathcal{F}$  and  $s \in \mathbb{Q}_0$ , choose  $\varepsilon_{\alpha}^s > 0$  such that

$$\sum_{\alpha \in \mathcal{F}} \sum_{s \in \mathbb{Q}_0} \varepsilon_{\alpha}^s = \varepsilon.$$

By (7.6.1) and (7.6.2), we can choose  $r_{\alpha}^s < s$  such that

$$\theta_{\mathcal{L}}(u)(\langle \alpha^{r_{\alpha}^s} \wedge \neg \alpha^s \rangle) \leq \theta_{\mathcal{L}}(u)(\langle \alpha^{r_{\alpha}^s} \rangle) - \theta_{\mathcal{L}}(u)(\langle \alpha^s \rangle) \leq \varepsilon_{\alpha}^s.$$

We call a set of formulas of  $\mathcal{L}$  *finitely-consistent* if each finite subset of it is consistent. For an arbitrary  $\psi \in \mathcal{L}$ , let  $\langle \psi \rangle^*$  be the set maximally-finitely-consistent sets of formulas that contain  $\psi$ .

The assumption  $\bigcap_i \langle \psi_i \rangle = \emptyset$  implies that  $\bigcap_i \langle \psi_i \rangle^*$  contains no maximally-consistent set, but only maximally-finitely-consistent sets that are inconsistent (due to (R2)). We have

$$\mathcal{U}_{\mathcal{L}} = \bigcap_{\alpha \in \mathcal{F}} \bigcap_{s \in \mathbb{Q}_0} \left( \bigcup_{r < s} (\langle \neg \alpha^r \rangle^* \cup \langle \alpha^s \rangle^*) \right). \quad (7.6.3)$$

Thus  $\bigcap_i (\psi_i) = \emptyset$  is equivalent to the condition

$$\left( \bigcap_{\alpha \in \mathcal{F}} \bigcap_{s \in \mathbb{Q}_0} \left( \bigcup_{r < s} (\neg \alpha^r)^* \cup (\alpha^s)^* \right) \right) \cap \bigcap_i (\psi_i)^* = \emptyset.$$

From this it follows that

$$\left( \bigcap_{\alpha \in \mathcal{F}} \bigcap_{s \in \mathbb{Q}_0} \left( (\neg \alpha^{r_s})^* \cup (\alpha^s)^* \right) \right) \cap \bigcap_i (\psi_i)^* = \emptyset.$$

The set of maximally-finitely-consistent sets form a Stone space (the maximally-finitely-consistent sets are the ultrafilters of the support boolean algebra), which is a compact space and  $(\psi)^*$  is a clopen for any  $\psi \in \mathcal{L}$  – due to the classic Stone duality. Consequently, there exist finite sets  $C_0 \subseteq \mathcal{F}$  and  $S_0 \subseteq \mathbb{Q} \cap [0, 1]$  and  $j \in \mathbb{N}$  such that

$$\bigcap_{\alpha \in C_0} \bigcap_{s \in S_0} (\neg \alpha^{r_s} \vee \alpha^s)^* \cap (\psi_j)^* = \emptyset,$$

or in other words,

$$(\psi_j)^* \subseteq \bigcup_{\alpha \in C_0} \bigcup_{s \in S_0} (\alpha^{r_s} \wedge \neg \alpha^s)^* = \left( \bigvee_{\alpha \in C_0} \bigvee_{s \in S_0} (\alpha^{r_s} \wedge \neg \alpha^s) \right)^*$$

This proves that,

$$\vdash \psi_j \rightarrow \bigvee_{\alpha \in C_0} \bigvee_{s \in S_0} (\alpha^{r_s} \wedge \neg \alpha^s). \quad (7.6.4)$$

Consequently,

$$\begin{aligned} \theta_{\mathcal{L}}(u)((\psi_j)) &\leq \theta_{\mathcal{L}}(u)\left(\left(\bigvee_{\alpha \in C_0} \bigvee_{s \in S_0} (\alpha^{r_s} \wedge \neg \alpha^s)\right)\right) \leq \sum_{\alpha \in C_0} \sum_{s \in S_0} \theta_{\mathcal{L}}(u)((\alpha^{r_s} \wedge \neg \alpha^s)) \\ &\leq \sum_{\alpha \in C_0} \sum_{s \in S_0} \varepsilon_{\alpha}^s \leq \varepsilon. \end{aligned}$$

As  $\varepsilon > 0$  was arbitrarily chosen,  $\inf_i \theta_{\mathcal{L}}(u)((\psi_i)) = 0$ .

2. To prove it for  $\mathcal{L}(\Delta)$  we follow the same strategy only that we need to also consider the instances of (R3). We use the same notation.

As for the other case, we consider the countable set  $\mathcal{F}$  of elements of the form

$$\alpha^r = L_{t_1 \dots t_n} r \varphi \text{ for } \varphi \in \mathcal{L} \text{ and rational } t_1, \dots, t_n, r \geq 0,$$

parameterized by  $r$  only that, in addition we also add the elements  $\alpha^\infty = L_{t_1 \dots t_n} \perp$ .

As before, if  $r < s$ , then  $\vdash \alpha^s \rightarrow \alpha^r$  and the inequality 7.6.1 holds. Similarly, the inequality 7.6.2 holds and, in addition, we also have

$$\theta_{\mathcal{L}}(u)((\alpha^\infty)) = \inf_{r \in \mathbb{Q}^+} \theta_{\mathcal{L}}(u)((\alpha^r)). \quad (7.6.5)$$

Let  $\varepsilon > 0$  be an arbitrarily small positive number. For each  $\alpha \in \mathcal{F}$  and  $s \in \mathbb{Q}^+ \cup \{\infty\}$ , we choose  $\varepsilon_\alpha^s > 0$  such that

$$\sum_{\alpha \in \mathcal{F}} \sum_{s \in \mathbb{Q}^+ \cup \{\infty\}} \varepsilon_\alpha^s = \varepsilon.$$

As before, we can choose  $r_\alpha^s < s$  such that

$$\theta_{\mathcal{L}}(u)(\langle \alpha^{r_\alpha^s} \wedge \neg \alpha^s \rangle) \leq \theta_{\mathcal{L}}(u)(\langle \alpha^{r_\alpha^s} \rangle) - \theta_{\mathcal{L}}(u)(\langle \alpha^s \rangle) \leq \varepsilon_\alpha^s.$$

and  $r_\alpha^\infty \in \mathbb{Q}^+$  big enough such that

$$\theta_{\mathcal{L}}(u)(\langle \alpha^{r_\alpha^\infty} \wedge \neg \alpha^\infty \rangle) \leq \theta_{\mathcal{L}}(u)(\langle \alpha^{r_\alpha^\infty} \rangle) - \theta_{\mathcal{L}}(u)(\langle \alpha^\infty \rangle) \leq \varepsilon_\alpha^\infty.$$

This time we have

$$\mathcal{U}_{\mathcal{L}} = \bigcap_{\alpha \in \mathcal{F}} \bigcap_{s \in \mathbb{Q}^+ \cup \{\infty\}} \left( \bigcup_{r < s} (\langle \neg \alpha^r \rangle^* \cup \langle \alpha^s \rangle^*) \right). \quad (7.6.6)$$

and following the same argument as in the previous case we obtain that there exist finite sets  $C_0 \subseteq \mathcal{F}$  and  $S_0 \subseteq \mathbb{Q}^+ \cup \{\infty\}$  and  $j \in \mathbb{N}$  such that

$$\bigcap_{\alpha \in C_0} \bigcap_{s \in S_0} (\langle \neg \alpha^{r_\alpha^s} \vee \alpha^s \rangle^* \cap \langle \psi_j \rangle^*) = \emptyset,$$

From here we obtain the same conclusion as for case 1, which is

$$\vdash \psi_j \rightarrow \bigvee_{\alpha \in C_0} \bigvee_{s \in S_0} (\alpha^{r_\alpha^s} \wedge \neg \alpha^s) \quad (7.6.7)$$

implying further

$$\begin{aligned} \theta_{\mathcal{L}}(u)(\langle \psi_j \rangle) &\leq \theta_{\mathcal{L}}(u)(\langle \bigvee_{\alpha \in C_0} \bigvee_{s \in S_0} (\alpha^{r_\alpha^s} \wedge \neg \alpha^s) \rangle) \leq \sum_{\alpha \in C_0} \sum_{s \in S_0} \theta_{\mathcal{L}}(u)(\langle \alpha^{r_\alpha^s} \wedge \neg \alpha^s \rangle) \\ &\leq \sum_{\alpha \in C_0} \sum_{s \in S_0} \varepsilon_\alpha^s \leq \varepsilon. \end{aligned}$$

□

At this point we have demonstrated that  $\theta_{\mathcal{L}}$  is a set function defined on the field  $(\mathcal{L})$  that is finitely additive and countably subadditive. In the light of Theorem 3.2.3, it can be uniquely extended to a measure on  $\Sigma_{\mathcal{L}}$ , which was defined as the  $\sigma$ -algebra generated by  $(\mathcal{L})$ .

From the previous construction we also obtain that  $\theta_{\mathcal{L}(\Pi)}$  is a probabilistic measure and that  $\theta_{\mathcal{L}(\Pi^*)}$  is a subprobabilistic measure.

With these results in hand we can proceed to the proof of one of the main results of this section.

- Theorem 7.6.6** (Canonical models). 1.  $\mathcal{M}_{\mathcal{L}(\Pi)} = (\mathcal{U}_{\mathcal{L}(\Pi)}, \Sigma_{\mathcal{L}(\Pi)}, \theta_{\mathcal{L}(\Pi)})$  is a probabilistic Markov process;
2.  $\mathcal{M}_{\mathcal{L}(\Pi^*)} = (\mathcal{U}_{\mathcal{L}(\Pi^*)}, \Sigma_{\mathcal{L}(\Pi^*)}, \theta_{\mathcal{L}(\Pi^*)})$  is a subprobabilistic Markov process;
3.  $\mathcal{M}_{\mathcal{L}(\Delta)} = (\mathcal{U}_{\mathcal{L}(\Delta)}, \Sigma_{\mathcal{L}(\Delta)}, \theta_{\mathcal{L}(\Delta)})$  is a general Markov process.

*Proof.* In the generic case we only need to verify that  $\theta_{\mathcal{L}}$  is a measurable function.

Let  $\varphi \in \mathcal{L}$ , and  $r \in [0, 1]$  for  $\mathcal{L}(\Pi)$  and  $\mathcal{L}(\Pi^*)$  and  $r \in \mathbb{R}^+$  for  $\mathcal{L}(\Delta)$ .

Consider  $(r_i)_i \subseteq \mathbb{Q}$  an increasing sequence with supremum  $r$ .

Let

$$\begin{aligned} X &= \{\mu \in \Pi(\mathcal{U}_{\mathcal{L}(\Pi)}, \Sigma_{\mathcal{L}(\Pi)}) \mid \mu(\llbracket \varphi \rrbracket) \geq r\} \text{ for } \mathcal{L}(\Pi), \\ X &= \{\mu \in \Pi^*(\mathcal{U}_{\mathcal{L}(\Pi^*)}, \Sigma_{\mathcal{L}(\Pi^*)}) \mid \mu(\llbracket \varphi \rrbracket) \geq r\} \text{ for } \mathcal{L}(\Pi^*) \text{ and} \\ X &= \{\mu \in \Delta(\mathcal{U}_{\mathcal{L}(\Delta)}, \Sigma_{\mathcal{L}(\Delta)}) \mid \mu(\llbracket \varphi \rrbracket) \geq r\} \text{ for } \mathcal{L}(\Delta). \end{aligned}$$

It suffices to prove, in each case, that  $\theta_{\mathcal{L}}^{-1}(X) \in \Sigma_{\mathcal{L}}$ . But

$$\begin{aligned} \theta_{\mathcal{L}}^{-1}(X) &= \{u \in \mathcal{U}_{\mathcal{L}} \mid \theta(u)(\llbracket \varphi \rrbracket) \geq r\} = \\ &= \bigcap_i \{u \in \mathcal{U}_{\mathcal{L}} \mid \theta_{\mathcal{L}}(u)(\llbracket \varphi \rrbracket) \geq r_i\} = \\ &= \bigcap_i \llbracket L_{r_i} \varphi \rrbracket \in \Sigma_{\mathcal{L}}. \end{aligned}$$

□

It remains to prove that indeed the MPs previously constructed are canonical models.

We define the interpretation function  $i_{\mathcal{L}}$  for arbitrary  $u \in \mathcal{U}_{\mathcal{L}}$  by

$$i_{\mathcal{L}}(u) = u \cap \mathcal{P}.$$

Now we are ready to prove the Extended Truth Lemma.

**Lemma 7.6.7** (Extended Truth Lemma). For  $\mathcal{L} \in \{\mathcal{L}(\Pi), \mathcal{L}(\Pi^*), \mathcal{L}(\Delta)\}$ ,  $\Phi \subseteq \mathcal{L}$  and  $u \in \mathcal{U}_{\mathcal{L}}$ ,

$$\mathcal{M}_{\mathcal{L}}, u, i_{\mathcal{L}} \models \Phi \text{ iff } \Phi \subseteq u.$$

*Proof.* It is sufficient to prove inductively that for any  $\varphi \in \mathcal{L}$ ,

$$\mathcal{M}_{\mathcal{L}}, u, i_{\mathcal{L}} \models \varphi \text{ iff } \varphi \in u.$$

The case  $\varphi \in \mathcal{P}$  and the boolean cases are trivial.

**The case**  $\varphi = L_r \psi$ :

( $\implies$ ) Suppose that  $\mathcal{M}_{\mathcal{L}}, u, i_{\mathcal{L}} \models \varphi$  and  $\varphi \notin u$ . Hence  $\neg \varphi \in u$ . Let

$$x_u^\varphi = \inf \{r \in \mathbb{Q} \mid \neg L_r \psi \in u\}.$$

Then, from  $\neg L_r\psi \in u$ , we obtain  $r \geq x_u^\psi$ . But  $\mathcal{M}_{\mathcal{L}, u, i_{\mathcal{L}}} \models L_r\psi$  implies  $\theta_{\mathcal{L}}(u)(\|\psi\|) \geq r$ , i.e.  $x_u^\psi \geq r$ .

Hence,  $x_u^\psi = r \in \mathbb{Q}$  and applying Lemma 7.6.3 we obtain that  $L_{x_u^\psi}\psi \in u$ . This means that  $\psi \in u$  – contradiction.

( $\Leftarrow$ ) If  $L_r\psi \in u$ , then  $r \leq x_u^\psi$ , i.e.,  $r \leq \theta_{\mathcal{L}}(u)(\|\psi\|)$ . Hence,  $\mathcal{M}_{\mathcal{L}, u, i_{\mathcal{L}}} \models L_r\psi$ .  $\square$

The Extended Truth Lemma demonstrates that, indeed, the MPs constructed in this section from the sets of maximal-consistent sets of the three logics are canonical models. These allow us to finally prove the strong completeness for all three logics.

**Theorem 7.6.8 (Completeness).** For  $\mathcal{L} \in \{\mathcal{L}(\Pi), \mathcal{L}(\Pi^*), \mathcal{L}(\Delta)\}$ ,  $\Phi \subseteq \mathcal{L}$  and  $\varphi \in \mathcal{L}$

$$\Phi \models \varphi \text{ iff } \Phi \vdash \varphi.$$

*Proof.* ( $\Leftarrow$ ) This is a consequence of soundness, Theorem 7.4.2.

( $\Rightarrow$ ) If  $\Phi$  is inconsistent, the statement is trivially true.

Suppose that  $\Phi$  is consistent, and let  $u \in \mathcal{U}_{\mathcal{L}}$  be an arbitrary maximally consistent set. We have that  $\Phi \subseteq u$  iff  $\mathcal{U}_{\mathcal{L}, u, i_{\mathcal{L}}} \models \Phi$  (from the Extended Truth Lemma). But if  $\mathcal{U}_{\mathcal{L}, u, i_{\mathcal{L}}} \models \Phi$ , since  $\Phi \models \varphi$ , we obtain that  $\mathcal{U}_{\mathcal{L}, u, i_{\mathcal{L}}} \models \varphi$ . Applying again the Extended Truth Lemma we get  $\varphi \in u$ . Consequently, for an arbitrary maximally-consistent set  $u \in \mathcal{U}_{\mathcal{L}}$ ,  $\Phi \subseteq u$  implies  $\varphi \in u$ . Hence,  $\Phi \vdash \varphi$ .  $\square$

## 7.7 Concluding Remarks

The most closely related work to ours is the work of Goldblatt [25] on the role of the Baire category theorem in completeness proofs, and his work on deduction systems for coalgebras [26]. The main difference between his work and ours is that we have replaced the Countable additivity Rule (CAR) that he uses, with a different infinitary axiom that has only countably many instances. Goldblatt uses CAR in order to show countable additivity of the measures that he defines; this is where we have been able to use of the Rasiowa–Sikorski lemma. As far as we know this is a new idea. Furthermore, Goldblatt’s results are contingent on the assumption that consistent sets can be expanded to maximally consistent sets; we have essentially proved this fact for our logics.

Regarding the completeness proofs for Markovian logics, the results for probabilistic case were proved by Zhou in [49] and for the general case by Mardare-Cardelli-Larsen in [35]. In these papers the strong completeness is solved using CAR for an axiomatization containing a lighter version of (R2) and (R3).

Regarding the strong completeness results presented in this chapter, the proof was presented in [33] and generalized in [32].

A very tempting future research project is to extend these completeness theorems to the entire class of systems described as coalgebras of polynomial functors described by

Goldblatt [26]. It is possible that the results of Pattinson and Schröder [42] will be useful for this.

Though the focus of the present chapter has been on Markov systems and Markovian logics, the techniques may well apply to any non-compact modal logic. We are investigating whether there is a general way of introducing an infinitary axiom that will allow us to mimic the techniques of the present paper.

# Chapter 8

## The Topologies of Markovian Logics

### 8.1 Introduction

Probabilistic bisimulation, introduced by Larsen and Skou [34] has become the key concept for reasoning about the equivalence of probabilistic and stochastic Markov processes. The theory of probabilistic bisimulation has been extended to stochastic bisimulation relating processes with continuous-state spaces and continuous distributions [10, 15]. These papers provided a characterization of bisimulation using a negation-free logic, as we have seen in the previous Chapter.

However, it is also widely realized that probabilistic and stochastic bisimulations are too “exact” for most purposes — they only relate processes with identical behaviours. In applications we need instead to know whether two processes that may differ by a small amount in the real-valued parameters (rates or probabilities) have similar behaviours. These motivated the search for a relaxation of the notion of equivalence of processes.

The metric theory for Markov processes was initiated by Desharnais et al. [18] and greatly developed and explored by van Breugel, Worrell and others [46, 47]. The key idea was to consider a behavioral *pseudometric*, i.e. a variation of the concept of metric for processes where pairs of distinct processes are at distance 0 whenever the processes are bisimilar.

Though behavioural pseudometrics were defined, approximate *reasoning principles* as such did not develop.

The work presented in this Chapter is a step in that direction. We lift the metric between processes to a metric between logical formulas by standard techniques, using the *Hausdorff metric*; but then we break new ground by exploring the relationship between convergence of processes and of formulas. We thus lay the groundwork for a notion of *approximate reasoning* not by getting rid of the logic but by fusing metric and logical principles. The completeness theorems of [10, 11, 32] was a powerful impetus for the research presented in this Chapter.

Consider the sequence of stochastic processes represented in Figure 8.1. The pro-

cess  $m$  has only one state and one self-transition at rate 5; similarly, for each  $k \in \mathbb{N}$ , the process  $m_k$  has one state and one transition at rate  $4.\underbrace{9..9}_k$ . Using a behavioral pseudometric, we expect to prove that the sequence  $(m_k)_{k \in \mathbb{N}}$  of processes converges to  $m$ . We often meet such problems in practice where  $m$  is a natural process that we need to analyze, while  $m_k$  are increasingly accurate models of  $m$ . If, in addition, we have a convergent sequence of logical formulas  $\varphi_k$  with limit  $\varphi$  such that  $m_k \models \varphi_k$  for each  $k$ , we want to understand whether we can infer  $m \models \varphi$ .

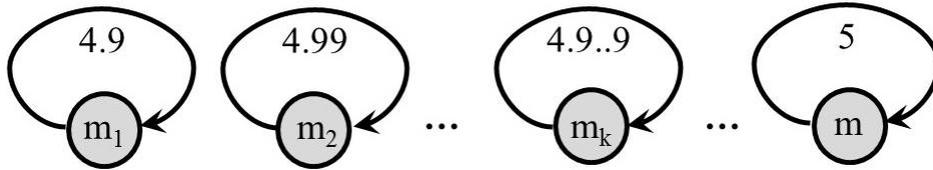


Figure 8.1: A sequence of convergent stochastic processes and their limit

In order to address such problems, we identify a general metrical notion that we call *dynamical continuity*. It characterizes the behavioral pseudometrics for which a sequence of processes as the one in our example is convergent; and it allows us to relate the convergence in formulas with convergence of the processes.

Using this concept we can address the above mentioned problem and prove that in general we do not have, at the limit,  $m \models \varphi$ .

For the probabilistic case  $m \models \varphi$  only if  $\varphi$  is a *positive* formula. Positive formulas will be defined in the paper; they are restricted, but they suffice for the modal characterization of probabilistic bisimulation.

For the stochastic (general) case we have to restrict the set of formulas slightly more, remaining however within a set of formulas that characterize bisimulation. In either case, even if  $m \not\models \varphi$ , there exists a sequence of processes  $(n_k)_{k \in \mathbb{N}}$  such that  $\lim_{k \rightarrow \infty} n_k = m$  and  $n_k \models \varphi$  for each  $k \in \mathbb{N}$ .

So this gives a handle on constructing approximations satisfying prescribed conditions. Along the way we give topological characterizations of various classes of formulas as defining open, closed,  $G_\delta$  or  $F_\sigma$  sets<sup>1</sup>. All these results hold whenever one has a dynamically-continuous metric bisimulation, as it is the case with the behavioral pseudometrics introduced in [18,46,47].

### The relevance of this work.

We prove that the process of extrapolating properties from arbitrary accurate approximations of a system to the system itself – a method widely accepted as valid and used in applications – is not always consistent. Often one constructs better and better approximations of a system, proves properties of these approximations and extrapolates

<sup>1</sup>In topology, a  $G_\delta$  set is a countable intersection of open sets and a  $F_\sigma$  set is a countable union of closed sets – see Preliminaries for detailed definitions.

the results to the original system. But can we indeed be sure that if, for instance, the approximants show oscillatory behaviours [5] then the original system also oscillates?

The mathematical framework developed in this Chapter allows us to address such a question and to prove that the answer is **no**, in general: there may exist sequences of arbitrarily accurate approximations of a system showing properties that are not preserved to the limit; and this already happens for fragments of modal probabilistic and stochastic logics, less expressive than CSL or pCTL.

We prove that the *preservation to the limit* depends on the logical structure of the property; “negative information” and “approximations from above”, for instance, are obstructions to this kind of limiting argument. Moreover, different logics behave differently to the limit. In this Chapter we show that there is a considerable difference between probabilistic and stochastic logical properties.

## 8.2 Preliminaries

In this section we fix the notation used in this Chapter, introduce a few useful concepts related to pseudometric spaces and prove a few lemmas that will be used in what follows.

### 8.2.1 Pseudometric Spaces

Given a pseudometric space  $(M, d)$ , we define the following variations of the Hausdorff pseudometric defined and discussed in Preliminaries.

For arbitrary  $a \in M$  and  $A, B \subseteq M$  with  $A \neq \emptyset \neq B$ , let

1.  $d^h(a, B) = \inf_{b \in B} d(a, b)$ ,
2.  $d^{H/2}(A, B) = \sup_{a \in A} \inf_{b \in B} d(a, b)$ ,
3.  $d^H(A, B) = \max\{\sup_{a \in A} \inf_{b \in B} d(a, b), \sup_{b \in B} \inf_{a \in A} d(a, b)\}$ .

We call  $d^H$  the *Hausdorff pseudometric* (associated to  $d$ ).

**Lemma 8.2.1.** *If  $d$  is a pseudometric on  $M$ , then  $d^H$  is a pseudometric on  $2^M$ . Moreover, for arbitrary  $x \in M$  and  $Y, Z \subseteq M$ ,*

$$d^h(x, Y) \leq d^h(x, Z) + d^H(Y, Z).$$

*Proof.* That  $d^H$  is a pseudometric, we have proven in the Preliminaries, see Lemma 3.4.1.

To prove the additional inequality, we start from arbitrary  $y \in Y$  and  $z \in Z$  and the triangle inequality

$$d(x, y) \leq d(x, z) + d(y, z).$$

This implies the following sequence of inequalities

$$\begin{aligned} \inf_{y \in Y} d(x, y) &\leq d(x, z) + d(y, z), \\ d^h(x, Y) &\leq d(x, z) + \sup_{z \in Z} d(y, z), \\ d^h(x, Y) &\leq d(x, z) + \sup_{z \in Z} \inf_{y \in Y} d(y, z), \\ d^h(x, Y) &\leq d(x, z) + d^{H/2}(Y, Z), \\ d^h(x, Y) &\leq d(x, z) + d^H(Y, Z), \\ d^h(x, y) &\leq d^h(x, Z) + d^H(Y, Z). \end{aligned}$$

□

In what follows, given a pseudometric space  $(M, d)$ , we use the induced open ball topology  $\mathcal{T}_d$ . We denote by  $\overline{X}$  the closure of  $X \subseteq M$  and by  $\widetilde{X}$  the interior of  $X$  – see the Preliminaries for the definitions of these topological concepts.

**Lemma 8.2.2.** *If  $(M, d)$  is a pseudometric space and  $\overline{X}$  is the closure of  $X \subseteq M$  in the open ball topology  $\mathcal{T}_d$ , then for arbitrary  $A, B \subseteq M$ ,*

1.  $d^H(A, B) = 0$  iff  $\overline{A} = \overline{B}$ ,
2.  $d^H(A, B) = d^H(\overline{A}, B) = d^H(A, \overline{B}) = d^H(\overline{A}, \overline{B})$ .

*Proof.* 1. We have the following sequence of equivalences

$$\begin{aligned} d^H(A, B) = 0 &\text{ iff} \\ \max\{d^{H/2}(A, B), d^{H/2}(B, A)\} = 0 &\text{ iff} \\ [d^{H/2}(A, B) = 0 \text{ and } d^{H/2}(B, A) = 0] &\text{ iff} \\ [\sup_{a \in A} \inf_{b \in B} d(a, b) = 0 \text{ and } \sup_{b \in B} \inf_{a \in A} d(a, b) = 0] &\text{ iff} \\ [\text{for arbitrary } a \in A \text{ and } b \in B, d(a, B) = d(b, A) = 0] &\text{ iff} \\ [\text{for arbitrary } a \in A \text{ and } b \in B, a \in \overline{B} \text{ and } b \in \overline{A}] &\text{ iff} \\ \overline{A} = \overline{B}. \end{aligned}$$

2. We have the triangle inequalities

$$d^H(A, \overline{B}) \leq d^H(A, B) + d^H(B, \overline{B})$$

and

$$d^H(A, B) \leq d^H(A, \overline{B}) + d^H(\overline{B}, B).$$

From 1 we have that

$$d^H(B, \overline{B}) = d^H(\overline{B}, B) = 0,$$

implying

$$d^H(A, \overline{B}) = d^H(A, B).$$

The other equalities can be proved in the same way.

□

In what follows, we consider for pseudometric spaces  $(M, d)$  the following notions of convergence in the open ball topologies  $\mathcal{T}_d$  and  $\mathcal{T}_{d^H}$  respectively<sup>2</sup>:

- For an arbitrary sequence  $(m_k)_{k \in \mathbb{N}}$  of elements of  $M$  and an arbitrary  $m \in M$ , we write  $m \in \lim_{k \rightarrow \infty} m_k$  (or  $\lim_{k \rightarrow \infty} m_k \ni m$ ) to denote that  $\lim_{k \rightarrow \infty} d(m_k, m) = 0$ .
- For an arbitrary sequence  $(S_k)_{k \in \mathbb{N}}$  of subsets of  $M$  and an arbitrary set  $S \subseteq M$ , we write  $S \in \lim_{k \rightarrow \infty} S_k$  (or  $\lim_{k \rightarrow \infty} S_k \ni S$ ) to denote that  $\lim_{k \rightarrow \infty} d^H(S_k, S) = 0$ .

**Lemma 8.2.3.** *Let  $(M, d)$  be a pseudometric space,  $(B_i)_{i \in \mathbb{N}}$  a decreasing sequence of non-empty compact subsets of  $M$  in the topology  $\mathcal{T}_d$  and  $A$  an arbitrary subset of  $M$ .*

$$(1): d^{H/2}(A, \bigcap_{i \in \mathbb{N}} B_i) = \sup_{i \in \mathbb{N}} d^{H/2}(A, B_i)$$

$$(2): \text{If } \lim_{k \rightarrow \infty} B_k \ni A, \text{ then } d^H(A, \bigcap_{i \in \mathbb{N}} B_i) = 0.$$

*Proof.* Notice that  $(2^M, d^H)$  is a pseudometric space and, hence, not a Hausdorff space, so a convergent sequence might have more than one limit. In fact if a sequence of sets converges to some  $X$  and  $d^H(X, Y) = 0$  then the sequence will also converge to  $Y$ .

Let  $B = \bigcap_{i \in \mathbb{N}} B_i$ . Since  $B_i \neq \emptyset$  and are all compact,  $B \neq \emptyset$ .

1. Consider arbitrary  $a \in A$ . Since  $B_i \supseteq B_{i+1} \supseteq B$  for each  $i$ , the sequence  $d^h(a, B_i)$  is increasing and  $d^h(a, B) \geq d^h(a, B_i)$ . Hence,  $d^h(a, B_i)$  is convergent and

$$\lim_{i \rightarrow \infty} d^h(a, B_i) = \sup_{i \in \mathbb{N}} d^h(a, B_i) \leq d^h(a, B).$$

Since  $d^h(a, B_i) = \inf_{b \in B_i} d(a, b)$  and  $B_i$  is compact, there exists  $b_i \in B_i$  such that  $d(a, b_i) = d^h(a, B_i)$ . Consequently,  $\lim_{i \rightarrow \infty} d^h(a, B_i) = \lim_{i \rightarrow \infty} d(a, b_i)$ .

Let  $b = \lim_{i \rightarrow \infty} b_i$ , which exists (for a subsequence) since  $B_i$  is compact. Clearly,  $b \in B_1$  since  $B_1$  is compact and  $(b_i)_{i \in \mathbb{N}} \subseteq B_1$ .

Suppose  $b \notin B$ . Then, there exists  $k \in \mathbb{N}$  such that  $b \in B_k$  and for any  $i \geq 1$ ,  $b \notin B_{k+i}$ . Then, for all  $i \geq k$ ,  $b_i \in B_k \setminus B_{k+1}$ , which is impossible since  $b_i \in B_i$  converges to  $b$ .

Consequently,  $b \in B$ ,  $\lim_{i \rightarrow \infty} d(a, b_i) = d(a, b)$  and since  $d(a, b) \geq d^h(a, B)$ , we obtain

$$\lim_{i \rightarrow \infty} d^h(a, B_i) \geq d^h(a, B).$$

<sup>2</sup>A pseudometric space is not a Hausdorff space and consequently the limits are not unique.

Hence,  $\lim_{i \rightarrow \infty} d^h(a, B_i) = d^h(a, B)$  implying  $\sup_{a \in A} \lim_{i \rightarrow \infty} d^h(a, B_i) = \sup_{a \in A} d^h(a, B)$  and further  $d^{H/2}(A, B) = \sup_{i \in \mathbb{N}} d^{H/2}(A, B_i)$ .

2. Consider  $x \in B$ . This means that for any  $k$ ,  $x \in B_k$ .

We have  $d^h(x, A) \leq d^H(B_k, A)$  and the latter sequence of numbers converges to 0 by assumption. Therefore  $d^h(x, A) = 0$ , i.e.  $x \in \overline{A}$ .

We have proved that  $B \subseteq \overline{A}$  and hence  $d^{H/2}(B, \overline{A}) = 0$ .

We prove now that  $d^{H/2}(\overline{A}, B) = 0$ .

Since  $d^H(A, B_k)$  converges to 0 by assumption, using Lemma 8.2.2, we obtain that  $d^H(\overline{A}, B_k)$  converges to 0 and further that  $d^{H/2}(\overline{A}, B_k)$  converges to 0. This implies that for any  $a \in \overline{A}$ ,  $d^h(a, B_k)$  converges to 0, since  $d^h(a, B_k) \leq d^{H/2}(\overline{A}, B_k)$ .

Consequently, the sequence  $d^h(a, B_k)$  of positive numbers converges to 0; moreover, this sequence is increasing, since  $(B_k)_{k \in \mathbb{N}}$  is decreasing. Hence, for each  $k \in \mathbb{N}$ ,  $d^h(a, B_k) = 0$  and this is true for each  $a \in \overline{A}$ . From here we get that  $d^{H/2}(\overline{A}, B_k) = 0$ , for all  $k \in \mathbb{N}$ . Using 1,  $d^{H/2}(\overline{A}, B) = \sup_{k \in \mathbb{N}} d^{H/2}(\overline{A}, B_k) = 0$ .

We have proved that  $d^{H/2}(B, \overline{A}) = d^{H/2}(\overline{A}, B) = 0$  which implies  $d^H(\overline{A}, B) = 0$ . Using Lemma 8.2.2,  $d^H(A, B) = 0$ .  $\square$

**Lemma 8.2.4.** *Let  $(M, d)$  be a pseudometric space and  $(m_k)_{k \in \mathbb{N}} \subseteq M$ ,  $(S_k)_{k \in \mathbb{N}} \subseteq 2^M$  convergent sequences with  $m \in \lim_{k \rightarrow \infty} m_k$  and  $S \in \lim_{k \rightarrow \infty} S_k$ . If  $m_k \in S_k$  for each  $k \in \mathbb{N}$ , then  $d^h(m, S) = 0$ . In particular, if  $S$  is closed, then  $m \in S$ .*

*Proof.* Observe that  $d^H(S', S'') = \max\{\sup_{m' \in S'} d^h(m', S''), \sup_{m'' \in S''} d^h(m'', S')\}$ .

Because  $m_k \in S_k$  implies  $d_h(m_k, S) \leq d^H(S_k, S)$  and because  $\lim_{k \rightarrow \infty} d^H(S_k, S) = 0$ , we obtain that  $\lim_{k \rightarrow \infty} d_h(m_k, S) = 0$ .

On the other hand,  $\lim_{k \rightarrow \infty} m_k = m$  implies  $\lim_{k \rightarrow \infty} d_h(m_k, S) = d_h(m, S)$ , hence  $d_h(m, S) = 0$ . If  $S$  is closed,  $d_h(m, S) = 0$  implies  $m \in S$ .  $\square$

## 8.2.2 Markov Processes and Markovian Logics - Basic Notations

In this chapter we recall some definitions of the concepts that we use in this chapter and discuss the pseudometric space of Markov processes.

As in the previous chapter we look both to the probabilistic and stochastic (general) Markov processes. The distinction between probabilistic and subprobabilistic processes does not play a role in this chapter and the following results applies in both cases; for this reason, we will call all these processes discrete-time Markov processes (DMPs) [15, 20, 38]. To maintain a name similarity, we will call the general Markov processes continuous-time Markov processes (CMPs) [10]. Also in this chapter we decided

not to use action-labels in the definition of processes. However, the labels can easily be added without changing any of these aspects of the theory.

Now we briefly recall the definition of Markov processes as introduced before in this monograph.

Let  $(M, \Sigma)$  be an *analytic space*, where  $\Sigma$  is its Borel algebra.

- A *discrete-time Markov process* (DMP) is a tuple  $\mathcal{M} = (M, \Sigma, \theta)$ , where

$$\theta \in \llbracket M \rightarrow \Pi(M, \Sigma) \rrbracket;$$

- A *continuous-time Markov process* (CMP) is a tuple  $\mathcal{M} = (M, \Sigma, \theta)$ , where

$$\theta \in \llbracket M \rightarrow \Delta(M, \Sigma) \rrbracket;$$

where  $\llbracket (X, \Sigma) \rightarrow (Y, \Theta) \rrbracket$  denotes the class of measurable mappings between the two measurable spaces.

For both types of processes,  $M$  is called the *support set* of  $\mathcal{M}$  denoted by  $\text{supp}(\mathcal{M})$ .

If  $m$  is the current state of a DMP and  $N$  is a measurable set of states, the transition function  $\theta(m)$  is a probability measure on the state space and  $\theta(m)(N) \in [0, 1]$  represents the *probability* of a transition from  $m$  to an arbitrary state  $n \in N$ .

Similarly, if  $m$  is the current state of a CMP and  $N$  is a measurable set of states, the transition function  $\theta(m)$  is a measure on the state space and  $\theta(m)(N) \in \mathbb{R}^+$  represents the *rate* of an exponentially distributed random variable that characterizes the duration of a transition from  $m$  to an arbitrary state  $n \in N$ .

Indeterminacy in such systems is resolved by races between events executing at different probabilities/rates.

Notice that, in both cases,  $\theta$  is a measurable mapping between the space of processes and the space of (probabilistic/stochastic) measures. These requirements are equivalent to the conditions on the corresponding two-variable *probabilistic/rate function* used in [15, 20, 38] to define labelled Markov processes and in [19] to define continuous Markov processes (for the proof see, Proposition 2.9 [20]).

The definitions of bisimulation for DMPs and CMPs follow the line of the Larsen-Skou definition of probabilistic bisimulation [34]. Hereafter, it is useful to use the relational definition of bisimulation.

Formally, given the DMK (CMK)  $\mathcal{M} = (M, \Sigma, \theta)$ , a *bisimulation relation* on  $\mathcal{M}$  is a relation  $\mathfrak{R} \subseteq M \times M$  such that whenever  $(m, n) \in \mathfrak{R}$ , for any  $C \in \Sigma(\mathfrak{R})$ ,

$$\theta(m)(C) = \theta(n)(C).$$

Two processes  $(\mathcal{M}, m)$  and  $(\mathcal{M}, n)$  are *bisimilar*, written  $m \sim_{\mathcal{M}} n$ , if they are related by a bisimulation relation.

As before, the bisimulation relation between processes with different supports is defined by taking the disjoint union of the two [10, 11, 15, 38]. For this reason, in what

follows we use  $\sim$  without extra indexes to denote the largest bisimulation relation. We call the largest bisimulation of DMPs *probabilistic bisimilarity* and the largest bisimulation of CMPs *stochastic bisimilarity*.

Also in this chapter we make use of the Markovian logics.

**The probabilistic case.** The Markovian logic defined for the semantics based on DMPs will be called in this chapter *discrete Markovian Logic* (DML). It corresponds to PML introduced in the previous chapter. For this reason we will continue denoting it by  $\mathcal{L}(\Pi)$ . The formulas are defined inductively by the following grammar.

$$\mathcal{L}(\Pi) : \quad \varphi := \top \mid \neg\varphi \mid \varphi \wedge \varphi \mid L_r\varphi, \quad r \in \mathbb{Q}_+ \cap [0, 1].$$

The semantics is defined as in the previous Chapter. We only remind the reader the semantics of the  $L_r$  operator, for an arbitrary DMP  $\mathcal{M} = (M, \Sigma, \theta)$  and arbitrary  $m \in \text{supp}(\mathcal{M})$ .

$$\mathcal{M}, m \models L_r\varphi \quad \text{iff} \quad \theta(m)(\llbracket \varphi \rrbracket) \leq r,$$

where  $\llbracket \varphi \rrbracket = \{n \in \text{supp}(\mathcal{M}) \mid \mathcal{M}, n \models \varphi\}$ .

Observe that this semantics is different from the one introduced in the previous chapter, as it does not involve interpretation functions. This is because in this Chapter we do not consider atomic propositions in the definition of  $\mathcal{L}(\Pi)$ . All the results we are about to prove can also be adapted, without difficulties, to the case where atomic propositions are involved. However, to keep the arguments simple, we decided not to include them.

For PML we assume to have all the classic Boolean derived operators. In addition, it is useful to consider the De Morgan dual of  $L_r$  defined by

$$M_r\varphi = L_{1-r}\neg\varphi.$$

One can easily verify the semantics of  $M_r$ :

$$\mathcal{M}, m \models M_r\varphi \quad \text{iff} \quad \theta(m)(\llbracket \varphi \rrbracket) \geq r.$$

Hence,  $L_r$  stays for "at least  $r$ " and  $M_r$  for "at most  $r$ ".

The axiomatization of  $\mathcal{L}(\Pi)$  that we will use in what follows is the one introduced in the previous chapter for  $\vdash_{\Pi}$ .

**The stochastic (general) case.** Similarly, the Markovian logic defined for a semantics based on CMPs will be referred to as the *continuous Markovian logic* (CML), and its formulas are inductively defined by the following grammar.

$$\mathcal{L}(\Delta) : \quad \varphi := \top \mid \neg\varphi \mid \varphi \wedge \varphi \mid L_r\varphi \mid M_r\varphi, \quad r \in \mathbb{Q}_+.$$

Notice that, with respect to GML discussed in the previous Chapter, CML has an extra class of operators –  $M_r$ . While in the probabilistic case  $L_r$  and  $M_r$  are De Morgan duals (because the sum of the probabilities of going from a state  $m$  to a state satisfying  $\varphi$  and

to a state satisfying  $\neg\varphi$  is 1), in the general case they are independent (because the rate of going to  $\varphi$  does not depend of the rate of going to  $\neg\varphi$ ). For this reason, in addition to the semantics that we defined in previous Chapter, we have to add the semantics of  $M_r$ :

$$\mathcal{M}, m \models M_r\varphi \text{ iff } \theta(m)(\llbracket\varphi\rrbracket) \geq r.$$

Due to this extra operator, we have to update the axiomatization of  $\mathcal{L}(\Delta)$  proposed in the previous Chapter to include axioms that describe the relation of the  $M_r$  operators to the other operators. In Table 8.1 we present an axiomatic system for CML. Notice the similarity to the axiomatic system of GML presented in Table 7.4.

$$\begin{aligned} \text{(B1): } & \vdash L_0\varphi \\ \text{(B2): } & \vdash L_{r+s}\varphi \rightarrow \neg M_r\varphi, s > 0 \\ \text{(B3): } & \vdash \neg L_r\varphi \rightarrow M_r\varphi \\ \text{(B4): } & \vdash \neg L_r(\varphi \wedge \psi) \wedge \neg L_s(\varphi \wedge \neg\psi) \rightarrow \neg L_{r+s}\varphi \\ \text{(B5): } & \vdash \neg M_r(\varphi \wedge \psi) \wedge \neg M_s(\varphi \wedge \neg\psi) \rightarrow \neg M_{r+s}\varphi \\ \text{(S1): } & \text{If } \vdash \varphi \rightarrow \psi \text{ then } \vdash L_r\varphi \rightarrow L_r\psi \\ \text{(S2): } & \{L_{r_1\dots r_k r}\psi \mid r < s\} \vdash L_{r_1\dots r_k s}\psi \\ \text{(S3): } & \{M_{r_1\dots r_k r}\psi \mid r > s\} \vdash M_{r_1\dots r_k s}\psi \\ \text{(S4): } & \{L_{r_1\dots r_k r}\psi \mid r > s\} \vdash L_{r_1\dots r_k}\perp \end{aligned}$$

Table 8.1: The axiomatic system of CML

A proof strategy similar to the ones we presented in the previous Chapter can be used to show that

- the axiomatic system in Table 8.1 is weak-complte for the CMP-semantics;
- CML enjoys the finite model property and the finite model construction presented in the previous Chapter for GML can be adapted to CML;
- the canonical model construction of GML can be extended to obtain a similar canonical model for CML;
- the axiomatic system in Table 8.1 is strong-complete with respect to the CMP-semantics.

All these results are detailed in [10]. Being their similarities with the proofs in the previous Chapter we have decided to omit them in this monograph.

In the end of this section recall that there exist strong relations between logical equivalence and bisimulation both for the probabilistic and for the stochastic cases that we have demonstrated and discussed in the previous Chapter. This is particularly relevant for our interest in this Chapter, because we expect that a pseudometric extension of the bisimulation will enjoy a similar property. This is exactly what we try to identify in the rest of this Chapter.

### 8.3 Dynamic-Continuous Pseudometric Bisimulations

As we have already underlined in the introduction, the concept of bisimulation for probabilistic or stochastic processes is very strict. We can however relax it by introducing a behavioral pseudometric [18, 38] which, formally, is a distances between processes that measure their similarity in terms of quantitative behaviour: the kernel of a behavioral pseudometric is a bisimulation. Moreover, we expect that a behavioral pseudometric can prove that a sequence of processes as  $(m_k)_{k \in \mathbb{N}}$  represented in Figure 8.1 is convergent to  $m$ .

In what follows, we identify a sufficient condition satisfied by any behavioral pseudometric that can prove such a convergence.

Before proceeding with the definition, recall that the convergences are in the corresponding open ball topologies, as defined in the previous section. In addition, we define the *kernel* of  $d$  as being the set

$$\ker(d) = \{(m, n) \in M \times M \mid d(m, n) = 0\}.$$

**Definition 8.3.1** (Dynamically-continuous pseudometric-bisimulation). *Given the DMK (CMK)  $\mathcal{M} = (M, \Sigma, \theta)$ , a pseudometric  $d : M \times M \rightarrow \mathbb{R}^+$  is*

- a pseudometric bisimulation if  $\ker(d) = \sim$
- dynamically-continuous if for any sequence  $(m_k)_{k \in \mathbb{N}} \in M$  such that  $\lim_{k \rightarrow \infty} m_k \ni m$  in the open-ball topology  $\mathcal{T}_d$ , then for any  $S \in \Sigma(\sim)$  there exists a decreasing sequence  $(S_k)_{k \in \mathbb{N}} \subseteq \Sigma(\sim)$  of compact sets in the topology  $\mathcal{T}_d$  such that
  - $\lim_{k \rightarrow \infty} S_k \ni S$  in the topology  $\mathcal{T}_{d^H}$
  - $\lim_{k \rightarrow \infty} \theta(m_k)(S_k) = \theta(m)(S)$  in  $\mathbb{R}$ .

Notice the coinductive nature of this definition, which is reminiscent of the general definition of bisimulation.

**Example 8.3.2.** *Let us now convince ourselves that any behavioral pseudometric that can prove the convergence in Figure 8.1 is indeed a dynamically-continuous pseudometric bisimulation.*

*Formally, in Figure 8.1 we have represented the CMK  $\mathcal{M} = (M, \Sigma, \theta)$  where*

- $M = \{m, m_1, ..m_k, ..\}$ ,
- $\Sigma = 2^M$ ,
- $\theta(m_k)(\underbrace{\{m_k\}}_k) = 4 \cdot \underbrace{9..9}_k$  for  $k \in \mathbb{N}$  and  $\theta(m)(\{m\}) = 5$ .

If in the open ball topology we can prove that  $\lim_{k \rightarrow \infty} m_k \ni m$ , then for each  $k \in \mathbb{N}$ ,  $S_k = \{m, m_k, m_{k+1}, \dots\}$  is a compact set – since it is closed and bounded in a complete pseudometric space;  $S_k \supseteq S_{k+1}$  and  $\lim_{k \rightarrow \infty} S_k \ni \{m\}$ .

Because  $\theta(m_k)$  is a measure,

$$\theta(m_k)(S_k) = \theta(m_k)(\{m_k\}) + \theta(m_k)(S_{k+1}).$$

But  $\theta(m_k)(S_{k+1}) = 0$ , hence  $\theta(m_k)(S_k) = \theta(m_k)(\{m_k\})$ . This implies that

$$\lim_{k \rightarrow \infty} \theta(m_k)(S_k) = \lim_{k \rightarrow \infty} \theta(m_k)(\{m_k\}) = \theta(m)(\{m\})$$

and verifies the second condition of the previous definition.

All these arguments motivate our choice for the definition of dynamically-continuous metric bisimulation.  $\square$

## 8.4 The Topological Space of Logical Formulas

Since a dynamically-continuous pseudometric bisimulation is a relaxation of the bisimulation relation, in what follows we try to identify similar logical characterization results for dynamically-continuous pseudometric bisimulation. In order to do this, we organize the space of the logical formulas as a pseudometric space, by identifying a logical formula with the set of its models and using the Hausdorff distance.

Formally, assume that the space  $\mathcal{M}$  of the continuous (or discrete) Markov processes is a pseudometric space defined by a pseudometric

$$d : \mathcal{M} \times \mathcal{M} \rightarrow \mathbb{R}^+.$$

The Hausdorff pseudometric  $d^H$  associated to  $d$  is also a distance between the sets  $\llbracket \varphi \rrbracket$  of models, for arbitrary  $\varphi \in \mathcal{L}(\Delta)$  (or  $\varphi \in \mathcal{L}(\Pi)$  respectively). Consequently, we can define, for arbitrary  $\varphi, \psi \in \mathcal{L}(\Delta)$  (or  $\varphi, \psi \in \mathcal{L}(\Pi)$  respectively), a distance  $\delta$  by

$$\delta(\varphi, \psi) = d^H(\llbracket \varphi \rrbracket, \llbracket \psi \rrbracket).$$

The properties of Hausdorff distances guarantees the following proposition.

**Proposition 8.4.1.**  $(\mathcal{L}(\Delta), \delta)$  and  $(\mathcal{L}(\Pi), \delta)$  are pseudometric spaces.

Since the space of models and the space of logical formulas are both pseudometric spaces, we can use the inequality of Lemma 8.2.1 to prove a robustness theorem.

**Theorem 8.4.2 (Robustness).** *Given a discrete (continuous) Markov process  $\mathcal{M}$ , an arbitrary state  $m \in \text{supp}(\mathcal{M})$  and arbitrary formulas  $\varphi$  and  $\psi$  of discrete (continuous) Markovian logic, then*

$$d^h(m, \llbracket \varphi \rrbracket) \leq d^h(m, \llbracket \psi \rrbracket) + \delta(\varphi, \psi).$$

If up to this point the two logics have similar properties, when we deeper focus on their topologies we see major differences that are studied in the next two subsections.

### 8.4.1 The topology of Discrete Markovian Logic

In this subsection we concentrate on the discrete Markovian logic.

Let  $\mathcal{M} = (M, \Sigma, \theta)$  be the universal DMP organized as a pseudometric space by the behavioural pseudometric  $d$ .

Let  $(\mathcal{L}(\Pi), \delta)$  be the pseudometric space of logical formulas defined as before.

To understand deeper the relation between the induced topologies, in what follows we isolate the following fragments of  $\mathcal{L}(\Pi)$ .

$$\mathcal{L}(\Pi)^+ : f := \top \mid f \vee f \mid f \wedge f \mid L_r \varphi \mid M_r \varphi, \quad \varphi \in \mathcal{L}(\Pi),$$

$$\mathcal{L}(\Pi)^- = \{\neg f \mid f \in \mathcal{L}(\Pi)^+\}.$$

In what follows we use  $\mathcal{T}_d$  to denote the open ball topology,  $\bar{X}$  and  $X^{int}$  to denote the closure and the interior of  $X \subseteq \mathcal{M}$ , respectively in the topology  $\mathcal{T}_d$ .

The question that we rise in what follows is if the formulas represent closed, open or different type of sets.

**Proposition 8.4.3.** *Let  $d$  be a dynamically-continuous metric bisimulation on  $\mathcal{M}$ .*

1. If  $\varphi \in \mathcal{L}(\Pi)^+$ , then  $\llbracket \varphi \rrbracket$  is a closed set in the topology  $\mathcal{T}_d$ .
2. If  $\varphi \in \mathcal{L}(\Pi)^-$ , then  $\llbracket \varphi \rrbracket$  is an open set in the topology  $\mathcal{T}_d$ .
3.  $\overline{\llbracket \neg M_r \varphi \rrbracket} = \llbracket L_r \varphi \rrbracket$  and  $\overline{\llbracket \neg L_r \varphi \rrbracket} = \llbracket M_r \varphi \rrbracket$ .
4.  $\llbracket M_r \varphi \rrbracket^{int} = \llbracket \neg L_r \varphi \rrbracket$  and  $\llbracket L_r \varphi \rrbracket^{int} = \llbracket \neg M_r \varphi \rrbracket$ .

*Proof.* 1. Induction on  $\varphi \in \mathcal{L}(\Pi)^+$ .

The Boolean cases are trivial, since the entire universe  $\mathcal{M} = \llbracket \top \rrbracket$  and the intersection and the union of two closed sets are closed.

**[The case  $\varphi = L_r \psi$  for some  $\psi \in \mathcal{L}$ ]:**

Suppose that  $\lim_{k \rightarrow \infty} m_k = m$  and for each  $k \in \mathbb{N}$ ,  $m_k \models L_r \psi$ .

Because  $d$  is a dynamically-continuous pseudometric bisimulation and  $\llbracket \psi \rrbracket \in \Sigma(\sim)$ , there exists a decreasing sequence  $(S_k)_{k \in \mathbb{N}} \subseteq \Sigma(\sim)$  of compact sets in  $\mathcal{T}_d$  such that  $\lim_{k \rightarrow \infty} S_k \ni \llbracket \psi \rrbracket$  and  $\lim_{k \rightarrow \infty} \theta(m_k)(S_k) = \theta(m)(\llbracket \psi \rrbracket)$ .

From Lemma 8.2.3,  $d^H(\llbracket \psi \rrbracket, \bigcap_{k \in \mathbb{N}} S_k) = 0$  and using Lemma 8.2.2,  $\overline{\llbracket \psi \rrbracket} = \bigcap_{k \in \mathbb{N}} S_k$ .

Hence,  $\llbracket \psi \rrbracket \subseteq \overline{\llbracket \psi \rrbracket} \subseteq S_k$  for any  $k$ .

Since  $m_k \models L_r \psi$ ,  $\theta(m_k)(\llbracket \psi \rrbracket) \geq r$  implying  $\theta(m_k)(S_k) \geq \theta(m_k)(\llbracket \psi \rrbracket)$  and further,  $\theta(m_k)(S_k) \geq r$ .

Hence,  $\lim_{k \rightarrow \infty} \theta(m_k)(S_k) \geq r$ , implying  $\theta(m)(\llbracket \psi \rrbracket) \geq r$ , i.e.,  $m \models L_r \psi$ .

**[The case  $\varphi = M_r \psi$  for some  $\psi \in \mathcal{L}$ ]:**

From the soundness of  $\vdash M_r\varphi \leftrightarrow L_{1-r}\neg\varphi$  we obtain  $\llbracket M_r\varphi \rrbracket = \llbracket L_{1-r}\neg\varphi \rrbracket$ . Now we can use the fact that  $\llbracket L_{1-r}\neg\varphi \rrbracket$  is open.

2. It is a direct consequence of 1, since  $\llbracket \neg\psi \rrbracket$  is the complement of  $\llbracket \psi \rrbracket$ .

3. Because  $\models \neg M_r\varphi \rightarrow L_r\varphi$ ,  $\llbracket \neg M_r\varphi \rrbracket \subseteq \llbracket L_r\varphi \rrbracket$ .

Consider a sequence  $(m_k)_{k \in \mathbb{N}}$  of elements of  $\mathcal{M}$  such that  $\lim_{k \rightarrow \infty} m_k = m \in \mathcal{M}$  and for each  $k$ ,  $m_k \models \neg M_r\varphi$ . To prove that  $\overline{\llbracket \neg M_r\varphi \rrbracket} = \llbracket L_r\varphi \rrbracket$ , we need to verify that  $m \models L_r\psi$ .

Since  $d$  is a dynamically-continuous pseudometric bisimulation, there exists a decreasing sequence of compact sets  $(S_k)_{k \in \mathbb{N}}$  such that  $\lim_{k \rightarrow \infty} S_k \ni \llbracket \psi \rrbracket$  and

$$\lim_{k \rightarrow \infty} \theta(m_k)(S_k) = \theta(m)(\llbracket \psi \rrbracket).$$

Because  $S_k \supseteq \llbracket \psi \rrbracket$  (see the proof of 1.),  $\theta(m_k)(S_k) \geq \theta(m_k)(\llbracket \psi \rrbracket)$  and from  $m_k \models \neg M_r\psi$  we get  $\theta(m_k)(\llbracket \psi \rrbracket) > r$ . Hence,  $\theta(m_k)(S_k) > r$ , implying  $\lim_{k \rightarrow \infty} \theta(m_k)(S_k) \geq r$ . This is equivalent to  $\theta(m)(\llbracket \psi \rrbracket) \geq r$ , i.e.,  $m \models L_r\psi$ .

To prove the other equality, notice that we have

$$\overline{\llbracket \neg L_r\varphi \rrbracket} = \overline{\llbracket \neg M_{1-r}\neg\varphi \rrbracket}.$$

We apply the first equality and obtain  $\overline{\llbracket \neg M_{1-r}\neg\varphi \rrbracket} = \llbracket L_{1-r}\neg\varphi \rrbracket$  and further we get  $\llbracket L_{1-r}\neg\varphi \rrbracket = \llbracket M_r\varphi \rrbracket$ .

4. It is a direct consequence of 3. □

At this point we want to understand more about the kernel of  $\delta$  and its relation to provability. Since the axiomatic system of DML is complete, the next theorem follows from the definition of Hausdorff distance.

**Theorem 8.4.4.** *If  $\varphi, \psi \in \mathcal{L}(\Pi)$ , then*

$$\vdash \varphi \leftrightarrow \psi \text{ implies } \delta(\varphi, \psi) = 0.$$

The next results and the example will show that actually the reverse of Theorem 8.4.4 is not true: not all the formulas at distance zero are logically equivalent.

**Theorem 8.4.5.** *Let  $d$  be a dynamically-continuous pseudometric bisimulation and  $\varphi, \psi \in \mathcal{L}(\Pi)$ . such that  $\delta(\varphi, \psi) = 0$ .*

1. *If  $\varphi \in \mathcal{L}(\Pi)^+$ , then  $[\delta(\varphi, \psi) = 0 \text{ implies } \vdash \psi \rightarrow \varphi]$ .*

2. *If  $\varphi, \psi \in \mathcal{L}(\Pi)^+$ , then  $[\delta(\varphi, \psi) = 0 \text{ iff } \vdash \varphi \leftrightarrow \psi]$ .*

*Proof.* 1.  $\delta(\varphi, \psi) = 0$  is equivalent to  $d^H(\llbracket \varphi \rrbracket, \llbracket \psi \rrbracket) = 0$ , which is equivalent, as stated in Lemma 8.2.2, to  $\overline{\llbracket \varphi \rrbracket} = \overline{\llbracket \psi \rrbracket}$ .

If  $\varphi \in \mathcal{L}(\Pi)^+$ , by Proposition 8.4.3,  $\llbracket \varphi \rrbracket$  is closed. Hence,  $\overline{\llbracket \psi \rrbracket} = \llbracket \varphi \rrbracket$ . This implies that  $\llbracket \psi \rrbracket \subseteq \llbracket \varphi \rrbracket$ , i.e.,  $\models \psi \rightarrow \varphi$ , which, due to completeness, is equivalent to  $\vdash \psi \rightarrow \varphi$ .

2. is proved by applying 1. in both directions.  $\square$

**Example 8.4.6.** *There exist logical formulas that are at distance zero without being logically equivalent:*

$$\delta(L_r\varphi, \neg M_r\varphi) = 0,$$

$$\vdash \neg M_r\varphi \rightarrow L_r\varphi$$

$$\text{but } \not\vdash L_r\varphi \rightarrow \neg M_r\varphi.$$

To prove these, observe that for any model  $m \in \mathcal{M}$ , if  $m \models \neg M_r\varphi$ , then  $m \models L_r\varphi$ . This guarantees that the closure of  $\llbracket \neg M_r\varphi \rrbracket$  is included in  $\llbracket L_r\varphi \rrbracket$ .

Observe that if  $\theta(m)(\llbracket \varphi \rrbracket) = r$ , then  $m \models L_r\varphi$ , but  $m \not\models \neg M_r\varphi$ .

Suppose that there exists a model  $m \in \mathcal{M}$  such that  $m \in \llbracket L_r\varphi \rrbracket$  and  $d^h(m, \llbracket \neg M_r\varphi \rrbracket) > 0$ . Then,  $\theta(m)(\llbracket \varphi \rrbracket) \geq r$  and  $\theta(m)(\llbracket \varphi \rrbracket) < r$  - impossible.

Hence the closure of  $\llbracket \neg M_r\varphi \rrbracket$  coincides with  $\llbracket L_r\varphi \rrbracket$  and this proves that

$$\delta(L_r\varphi, \neg M_r\varphi) = 0.$$

$\square$

The next theorem states that whenever  $(m_k)_{k \in \mathbb{N}}$  is a sequence of increasingly accurate approximations of  $m$ , if  $m_k \models \varphi_k$  for each  $k$ , we cannot guarantee that  $m$  satisfies the limit  $\varphi$  of  $(\varphi_k)_{k \in \mathbb{N}}$ . However, we can prove that there exists a sequence of approximations of  $m$  satisfying  $\varphi$ .

**Theorem 8.4.7.** *If  $d$  is a dynamically-continuous pseudometric bisimulation and*

$$(\varphi_k)_{k \in \mathbb{N}} \subseteq \mathcal{L}(\Pi), \quad (m_k)_{k \in \mathbb{N}} \subseteq \mathcal{M}$$

*are two convergent sequences such that  $\lim_{k \rightarrow \infty} \varphi_k \ni \varphi$ ,  $\lim_{k \rightarrow \infty} m_k \ni m$ ; and*

$$m_k \models \varphi_k \text{ for each } k \in \mathbb{N},$$

*then there exists a convergent sequence  $(n_k)_{k \in \mathbb{N}} \subseteq \mathcal{M}$  such that  $\lim_{k \rightarrow \infty} n_k \ni m$  and*

$$n_k \models \varphi \text{ for each } k \in \mathbb{N}.$$

*Proof.* If we apply Lemma 8.2.4 for  $S_k = \llbracket \varphi_k \rrbracket$  and  $S = \llbracket \varphi \rrbracket$ , we obtain that  $d^h(m, \llbracket \varphi \rrbracket) = 0$  which implies that there exists a sequence  $(n_k)_{k \in \mathbb{N}} \subseteq \llbracket \varphi \rrbracket$  such that  $\lim_{k \rightarrow \infty} n_k = m$ .  $\square$

There exist, however, properties that can be "taken to the limit", as the following theorem proves.

**Theorem 8.4.8.** *Let  $d$  be a dynamically-continuous pseudometric bisimulation and*

$$(\varphi_k)_{k \in \mathbb{N}} \subseteq \mathcal{L}(\Pi), \quad (m_k)_{k \in \mathbb{N}} \subseteq \mathcal{M}$$

*two convergent sequences such that  $\lim_{k \rightarrow \infty} \varphi_k \ni \varphi$ ,  $\lim_{k \rightarrow \infty} m_k \ni m$  and*

$$m_k \models \varphi_k \text{ for each } k \in \mathbb{N}.$$

*If  $\varphi \in \mathcal{L}(\Pi)^+$ , then  $m \models \varphi$ .*

*Proof.* As in Theorem 8.4.7,  $d^h(m, \llbracket \varphi \rrbracket) = 0$ .

Since  $\varphi \in \mathcal{L}(\Pi)^+$ , Proposition 8.4.3 guarantees that  $\llbracket \varphi \rrbracket$  is closed and using the second part of Lemma 8.2.4 we obtain  $m \in \llbracket \varphi \rrbracket$ .  $\square$

## 8.4.2 The topology of Continuous Markovian Logic

In this subsection we investigate the similar problems for the case of CMPs and continuous Markovian logic. Hereafter, let  $\mathcal{M}$  be the universal CMP organized as a pseudometric space by the behavioural pseudometric  $d$ .

Let  $(\mathcal{L}(\Delta), \delta)$  be the pseudometric space of logical formulas.

As for the probabilistic case, we use  $\mathcal{T}_d$  to denote the open ball topology,  $\overline{X}$  and  $X^{int}$  to denote the closure and the interior of  $X \subseteq \mathcal{M}$ , respectively in the topology  $\mathcal{T}_d$ .

**Lemma 8.4.9.** *For arbitrary  $\varphi \in \mathcal{L}(\Delta)$ ,*

1.  $\overline{\llbracket \neg M_r \varphi \rrbracket} = \llbracket L_r \varphi \rrbracket$ ,
2.  $\llbracket M_r \varphi \rrbracket^{int} = \llbracket \neg L_r \varphi \rrbracket$ ,
3.  $\llbracket M_r \varphi \rrbracket = \bigcap_{k \in \mathbb{N}} \llbracket \neg L_{r+\frac{1}{k}} \varphi \rrbracket$
4.  $\llbracket L_r \varphi \rrbracket = \bigcap_{k \in \mathbb{N}} \llbracket \neg M_{r-\frac{1}{k}} \varphi \rrbracket$ .

*Proof.* 1. Because  $\models \neg M_r \varphi \rightarrow L_r \varphi$ , we get  $\llbracket \neg M_r \varphi \rrbracket \subseteq \llbracket L_r \varphi \rrbracket$ .

Consider a sequence  $(m_k)_{k \in \mathbb{N}}$  of elements of  $\mathcal{M}$  such that  $\lim_{k \rightarrow \infty} m_k \ni m \in \mathcal{M}$  and for each  $k$ ,  $m_k \models \neg M_r \varphi$ . To prove that  $\overline{\llbracket \neg M_r \varphi \rrbracket} = \llbracket L_r \varphi \rrbracket$ , we need to verify that  $m \models L_r \varphi$ .

Since  $d$  is a dynamically-continuous pseudometric bisimulation, there exists a decreasing sequence  $(S_k)_{k \in \mathbb{N}}$  of compact elements of  $\Sigma(\sim)$  such that  $\lim_{k \rightarrow \infty} S_k \ni \llbracket \psi \rrbracket$  and

$$\lim_{k \rightarrow \infty} \theta(m_k)(S_k) = \theta(m)(\llbracket \psi \rrbracket).$$

Because  $S_k \supseteq \llbracket \psi \rrbracket$ , we obtain that  $\theta(m_k)(S_k) \geq \theta(m_k)(\llbracket \psi \rrbracket)$  and from  $m_k \models \neg M_r \psi$  we get  $\theta(m_k)(\llbracket \psi \rrbracket) > r$ .

Hence  $\theta(m_k)(S_k) > r$ , implying  $\lim_{k \rightarrow \infty} \theta(m_k)(S_k) \geq r$ . This is equivalent to

$$\theta(m)(\llbracket \psi \rrbracket) \geq r, \text{ i.e., } m \models L_r \psi.$$

2. It is a direct consequence of 1.

3. It is a consequence of the fact that  $[0, r] = \bigcap_{k \in \mathbb{N}} [0, r + \frac{1}{k}]$ .

4. This follows from  $[r, \infty) = \bigcap_{k \in \mathbb{N}} (r - \frac{1}{k}, \infty)$ . □

In the following examples we show that, unlike in the probabilistic case,  $\llbracket M_r \psi \rrbracket$  is sometimes neither open nor closed in  $\mathcal{T}_d$ .

**Example 8.4.10.** We return to the stochastic system described in Example 8.3.2 and represented in Figure 8.1.

Notice that,

$$\text{for each } k \in \mathbb{N}, \quad m_k \models M_0 L_5 \top$$

meaning that each  $m_k$  cannot do a transition to a state (which is equivalent with "it does it at rate 0") where from it is possible to do a transition at rate at least 5. But at the limit,  $m \models \neg M_0 L_5 \top$  since  $\theta(m)(\llbracket L_5 \top \rrbracket) = 5 > 0$ .

Consequently,  $\llbracket M_0 L_5 \top \rrbracket$  is not closed in  $\mathcal{T}_d$ , since we found a sequence of processes from  $\llbracket M_0 L_5 \top \rrbracket$  with a limit outside  $\llbracket M_0 L_5 \top \rrbracket$ .

To prove that sometimes  $\llbracket M_r \psi \rrbracket$  is not open either, consider the same processes as before only that

$$\text{for each } k \in \mathbb{N}, \quad \theta(m_k)(\{m_k\}) = r_k,$$

where  $(r_k)_{k \in \mathbb{N}} \in \mathbb{Q}^+$  is a strictly decreasing sequence with limit 5.

In this case, for each  $k \in \mathbb{N}$ ,

$$m_k \models \neg M_5 \top, \text{ since } \theta(m_k)(\{m_k\}) > 5.$$

However, to the limit we have  $m \models M_5 \top$  proving that  $\llbracket \neg M_5 \top \rrbracket$  is not closed in  $\mathcal{T}_d$ , hence,  $\llbracket M_5 \top \rrbracket$  is not open. □

To understand this topology more deeply, we isolate the following fragments of  $\mathcal{L}(\Delta)$ .

$$\mathcal{L}(\Delta)^+ : \quad f := \top \mid f \wedge f \mid f \vee f \mid L_r \varphi \mid M_r \varphi,$$

$$\mathcal{L}(\Delta)^- = \{\neg f \mid f \in \mathcal{L}(\Delta)^+\},$$

$$\mathcal{L}(\Delta)_0 : \quad f := \top \mid f \wedge f \mid \neg f \mid L_r \varphi,$$

$$\mathcal{L}(\Delta)_0^+ : \quad f := \top \mid f \wedge f \mid f \vee f \mid L_r \varphi,$$

$$\mathcal{L}(\Delta)_0^- = \{\neg f \mid f \in \mathcal{L}(\Delta)_0^+\},$$

where in the previous definitions  $\varphi \in \mathcal{L}(\Delta)$ .

The next lemma marks essential differences between the topology of DML formulas and the topology of CML formulas. Recall that, in topology, a  $G_\delta$  set is a countable intersection of open sets and a  $F_\sigma$  set is a countable union of closed sets (see the Preliminaries for the definitions).

**Theorem 8.4.11.** *If  $d$  is a dynamically-continuous pseudometric bisimulation, then*

1. *If  $\varphi \in \mathcal{L}(\Delta)_0^+$ , then  $\llbracket \varphi \rrbracket$  is a closed set in the topology  $\mathcal{T}_d$ .*
2. *If  $\varphi \in \mathcal{L}(\Delta)_0^-$ , then  $\llbracket \varphi \rrbracket$  is an open set in the topology  $\mathcal{T}_d$ .*
3. *If  $\varphi \in \mathcal{L}(\Delta)^+$ , then  $\llbracket \varphi \rrbracket$  is a  $G_\delta$  set in the topology  $\mathcal{T}_d$ .*
4. *If  $\varphi \in \mathcal{L}(\Delta)^-$ , then  $\llbracket \varphi \rrbracket$  is a  $F_\sigma$  set in the topology  $\mathcal{T}_d$ .*

*Proof.* 1. Induction on  $\varphi \in \mathcal{L}(\Delta)_0^+$ .

The Boolean cases are trivial and use the fact that the entire universe  $\mathcal{M} = \llbracket \top \rrbracket$ , hence it is closed and that the intersection and the union of two closed sets are closed.

**The case  $\varphi = L_r\psi$  for some  $\psi \in \mathcal{L}(\Delta)$ :**

Suppose that  $\lim_{k \rightarrow \infty} m_k = m$  and for each  $k \in \mathbb{N}$ ,  $m_k \models L_r\psi$ ; we have to prove that  $m \models L_r\psi$ .

Because  $d$  is a dynamically-continuous pseudometric bisimulation,  $\lim_{k \rightarrow \infty} m_k = m$  and  $\llbracket \psi \rrbracket \in \Sigma(\sim)$ , there exists a decreasing sequence  $(S_k)_{k \in \mathbb{N}} \subseteq \Sigma(\sim)$  of compact sets in  $\mathcal{T}_d$  such that  $\lim_{k \rightarrow \infty} S_k \ni \llbracket \psi \rrbracket$  and  $\lim_{k \rightarrow \infty} \theta(m_k)(S_k) = \theta(m)(\llbracket \psi \rrbracket)$ .

From Lemma 8.2.3,  $d^H(\llbracket \psi \rrbracket, \bigcap_{k \in \mathbb{N}} S_k) = 0$  and using Lemma 8.2.2,  $\overline{\llbracket \psi \rrbracket} = \bigcap_{k \in \mathbb{N}} S_k$ , since  $\bigcap_{k \in \mathbb{N}} S_k$  is closed. Hence,  $\llbracket \psi \rrbracket \subseteq \overline{\llbracket \psi \rrbracket} \subseteq S_k$  for any  $k$ .

Because  $m_k \models L_r\psi$ , we obtain that  $\theta(m_k)(\llbracket \psi \rrbracket) \geq r$ . Further,  $\llbracket \psi \rrbracket \subseteq S_k$  implies  $\theta(m_k)(S_k) \geq \theta(m_k)(\llbracket \psi \rrbracket)$  and  $\theta(m_k)(S_k) \geq r$ . Hence,  $\lim_{k \rightarrow \infty} \theta(m_k)(S_k) \geq r$ , implying  $\theta(m)(\llbracket \psi \rrbracket) \geq r$ , i.e.,  $m \models L_r\psi$ .

2. It is a direct consequence of 1 since  $\psi \in \mathcal{L}(\Delta)_0^-$  iff  $\neg\psi \in \mathcal{L}(\Delta)_0^+$  and  $\llbracket \neg\psi \rrbracket$  is the complement of  $\llbracket \psi \rrbracket$ .

3. Induction on  $\varphi \in \mathcal{L}^+$ .

The Boolean cases are trivial, since the union and the intersection of two countable interactions of open sets is a countable intersection of open sets.

**The case  $\varphi \in \mathcal{L}(\Delta)_0^+$ :** follows from 1, since any closed set in a pseudometrizable space is  $G_\delta$ .

**The case  $\varphi = M_r\psi$  for some  $\psi \in \mathcal{L}(\Delta)$ :**

From Lemma 8.4.9, we have that  $\llbracket M_r\psi \rrbracket = \bigcap_{k \in \mathbb{N}} \llbracket \neg L_{r+\frac{1}{k}}\psi \rrbracket$ . Because  $\llbracket \neg L_{r+\frac{1}{k}}\psi \rrbracket$  are open, we obtain that  $\llbracket M_r\psi \rrbracket$  is a  $G_\delta$  set.

4. This is a consequence of 3, since  $\llbracket \neg\psi \rrbracket$  is the complement of  $\llbracket \psi \rrbracket$ .  $\square$

As for DML, logical equivalence is a subset of the kernel of  $\delta$ .

**Theorem 8.4.12.** *If  $\vdash \varphi \leftrightarrow \psi$ , then  $\delta(\varphi, \psi) = 0$ .*

However, Theorem 8.4.5 does not hold for CML. Instead we have the following weaker result relying on the fact that  $\llbracket \varphi \rrbracket$  is closed whenever  $\varphi \in \mathcal{L}(\Delta)_0^+$ .

**Theorem 8.4.13.** *Let  $d$  be a dynamically-continuous pseudometric bisimulation and  $\varphi, \psi \in \mathcal{L}(\Delta)$ .*

1. *If  $\varphi \in \mathcal{L}(\Delta)_0^+$ , then  $[\delta(\varphi, \psi) = 0$  implies  $\vdash \psi \rightarrow \varphi$ ].*
2. *If  $\varphi, \psi \in \mathcal{L}(\Delta)_0^+$ , then  $[\delta(\varphi, \psi) = 0$  iff  $\vdash \varphi \leftrightarrow \psi$ ].*

A similar result to Theorem 8.4.7 holds for CML.

**Theorem 8.4.14.** *If  $d$  is a dynamically-continuous pseudometric bisimulation and*

$$(\varphi_k)_{k \in \mathbb{N}} \subseteq \mathcal{L}(\Delta), \quad (m_k)_{k \in \mathbb{N}} \subseteq \mathcal{M}$$

*are two convergent sequences such that  $\lim_{k \rightarrow \infty} \varphi_k \ni \varphi$ ,  $\lim_{k \rightarrow \infty} m_k \ni m$  and*

$$m_k \vDash \varphi_k \text{ for each } k \in \mathbb{N},$$

*then there exists a convergent sequence  $(n_k)_{k \in \mathbb{N}} \subseteq \mathcal{M}$  such that  $\lim_{k \rightarrow \infty} n_k \ni m$  and*

$$n_k \vDash \varphi \text{ for each } k \in \mathbb{N}.$$

Theorem 8.4.8 does not hold for CML. But since  $\llbracket \varphi \rrbracket$  is closed whenever  $\varphi \in \mathcal{L}(\Delta)_0^+$ , we have a weaker version of it that does not involve the operators of type  $M_r$ .

**Theorem 8.4.15.** *Let  $d$  be a dynamically-continuous pseudometric bisimulation and*

$$(\varphi_k)_{k \in \mathbb{N}} \subseteq \mathcal{L}(\Delta), \quad (m_k)_{k \in \mathbb{N}} \subseteq \mathcal{M}$$

*two convergent sequences such that  $\lim_{k \rightarrow \infty} \varphi_k \ni \varphi$ ,  $\lim_{k \rightarrow \infty} m_k \ni m$  and*

$$m_k \vDash \varphi_k \text{ for each } k \in \mathbb{N}.$$

*If  $\varphi \in \mathcal{L}(\Delta)_0^+$ , then  $m \vDash \varphi$ .*

## 8.5 Concluding Remarks

The main contributions of the present chapter are the following results:

- The definition of dynamically-continuous metric bisimulation which is the correct distance-based counterpart of the concept of probabilistic/stochastic bisimulation.
- The definition of the topology of logical formulas canonically induced by the behavioural pseudometrics.
- Theorems that establish when parallel sequences of (probabilistic or stochastic) processes and formulas converge to give satisfaction *in the limit*; these results reveal important differences between the probabilistic and stochastic Markovian logics.
- Theorems regarding the relationships between logical formulas being at zero distance and logical equivalence/provability.
- A theorem that relates how “close” a process  $m$  is to a formula  $\varphi$  to how close  $m$  is to another formula  $\psi$  and the distance between  $\varphi$  and  $\psi$  (robustness).
- Topological characterization of various classes of formulas.

There are many new things to explore further and some of them will be addressed in the following chapters.

A coalgebraic presentation of this work that helped us understanding a *metric analogue of Stone duality* for Markov processes will be presented in the next Part of this monograph.



# Bibliography

- [1] S. Abramsky. Domain theory in logical form. *Annals of Pure and Applied Logic*, 51:1–77, 1991.
- [2] W. Arveson. *An Invitation to  $C^*$ -Algebra*. Springer-Verlag, 1976.
- [3] Robert Aumann. Interactive epistemology I: knowledge. *International Journal of Game Theory*, 28:263–300, 1999.
- [4] Robert Aumann. Interactive epistemology II: probability. *International Journal of Game Theory*, 28:301–314, 1999.
- [5] P. Ballarini, R. Mardare, I. Mura. Analysing Biochemical Oscillations through Probabilistic Model Checking. In *FBTC 2008, ENTCS 229(1):3-19*, 2009.
- [6] Nick Bezhanishvili, Clemens Kupke, and Prakash Panangaden. Minimization via duality. In *Logic, Language, Information and Computation - 19th International Workshop, WoLLIC 2012, Buenos Aires, Argentina, September 3-6, 2012. Proceedings*, volume 7456 of *Lecture Notes in Computer Science*, pages 191–205. Springer, 2012.
- [7] P. Billingsley. *Probability and Measure*. Wiley-Interscience, 1995.
- [8] Filippo Bonchi, Marcello Bonsangue, Jan Rutten, and Alexandra Silva. Brzozowski’s algorithm (co)algebraically. In Robert Constable and Alexandra Silva, editors, *Logics and Program Semantics: Essays Dedicated to Dexter Kozen*, volume 7230 of *Lecture Notes in Computer Science*, pages 12–23. Springer-Verlag, 2012.
- [9] M. M. Bonsangue and A. Kurz. Duality for logics of transition systems. In *FoSSaCS*, pages 455–469, 2005.
- [10] Luca Cardelli, Kim G. Larsen, and Radu Mardare. Continuous markovian logic - from complete axiomatization to the metric space of formulas. In *CSL*, pages 144–158, 2011.
- [11] Luca Cardelli, Kim G. Larsen, and Radu Mardare. Modular markovian logic. In *ICALP (2)*, pages 380–391, 2011.

- [12] Vincent Danos, Josée Desharnais, François Laviolette, and Prakash Panangaden. Bisimulation and cocongruence for probabilistic systems. *Information and Computation*, 204(4):503–523, 2006.
- [13] E. de Vink and J. J. M. M. Rutten. Bisimulation for probabilistic transition systems: A coalgebraic approach. *Theoretical Computer Science*, 221(1/2):271–293, June 1999.
- [14] J. Desharnais, A. Edalat, and P. Panangaden. A logical characterization of bisimulation for labelled Markov processes. In *proceedings of the 13th IEEE Symposium On Logic In Computer Science, Indianapolis*, pages 478–489. IEEE Press, June 1998.
- [15] J. Desharnais, A. Edalat, and P. Panangaden. Bisimulation for labeled Markov processes. *Information and Computation*, 179(2):163–193, Dec 2002.
- [16] J. Desharnais, V. Gupta, R. Jagadeesan, and P. Panangaden. Metrics for labeled Markov systems. In *Proceedings of CONCUR99*, number 1664 in Lecture Notes in Computer Science. Springer-Verlag, 1999.
- [17] J. Desharnais, V. Gupta, R. Jagadeesan, and P. Panangaden. Approximating labeled Markov processes. *Information and Computation*, 184(1):160–200, July 2003.
- [18] J. Desharnais, V. Gupta, R. Jagadeesan, and P. Panangaden. A metric for labelled Markov processes. *Theoretical Computer Science*, 318(3):323–354, June 2004.
- [19] Josée Desharnais and Prakash Panangaden. Continuous stochastic logic characterizes bisimulation for continuous-time Markov processes. *Journal of Logic and Algebraic Programming*, 56:99–115, 2003. Special issue on Probabilistic Techniques for the Design and Analysis of Systems.
- [20] E.-E. Doberkat. *Stochastic Relations. Foundations for Markov Transition Systems*. Chapman and Hall, New York, 2007.
- [21] R. M. Dudley. *Real Analysis and Probability*. Wadsworth and Brookes/Cole, 1989.
- [22] R. Fagin and J. Y. Halpern. Reasoning about knowledge and probability. *Journal of the ACM*, 41(2):340–367, 1994.
- [23] Mai Gehrke, Serge Grigorieff, and Jean-Eric Pin. Duality and equational theory of regular languages. In *ICALP (2)*, pages 246–257, 2008.
- [24] A. Giacalone, C. Jou, and S. Smolka. Algebraic reasoning for probabilistic concurrent systems. In *Proceedings of the Working Conference on Programming Concepts and Methods*, IFIP TC2, 1990.
- [25] R. Goldblatt. On the role of the Baire category theorem in the foundations of logic. *Journal of Symbolic logic*, pages 412–422, 1985.

- [26] R. Goldblatt. Deduction systems for coalgebras over measurable spaces. *Journal of Logic and Computation*, 20(5):1069–1100, 2010.
- [27] Aviad Heifetz and Philippe Mongin. Probability logic for type spaces. *Games and Economic Behavior*, 35(1-2):31–53, April 2001.
- [28] Bart Jacobs. Probabilities, distribution monads, and convex categories. *Theor. Comput. Sci.*, 412(28):3323–3336, 2011.
- [29] Peter Johnstone. *Stone Spaces*, volume 3 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, 1982.
- [30] B. Jonsson and A. Tarski. Boolean algebras with operators I. *American Journal of Mathematics*, 73:891–939, 1951.
- [31] D. Kozen. A probabilistic PDL. *Journal of Computer and Systems Sciences*, 30(2):162–178, 1985.
- [32] D. Kozen, K. G. Larsen, R. Mardare, and P. Panangaden. Stone duality for markov processes. In *Proceedings of the 28th Annual IEEE Symposium on Logic in Computer Science: LICS 2013*. IEEE Computer Society, 2013.
- [33] D. Kozen, R. Mardare, and P. Panangaden. Strong completeness for markovian logics. In *Proceedings of the 38th International Symposium, MFCS 2013*, pages 655–666, 2013.
- [34] K. G. Larsen and A. Skou. Bisimulation through probabilistic testing. *Information and Computation*, 94:1–28, 1991.
- [35] Radu Mardare, Luca Cardelli, and Kim G. Larsen. Continuous markovian logics - axiomatization and quantified metatheory. *Logical Methods in Computer Science*, 8(4), 2012.
- [36] M. Mislove, J. Ouaknine, D. Pavlovic, and J. Worrell. Duality for labelled Markov processes. In I. Walukiewicz, editor, *Foundations of Software Science and Computation Structures, FOSSACS*, volume 2987 of *Lecture Notes In Computer Science*, pages 393–407, 2004.
- [37] L. S. Moss and I. D. Viglizzo. Harsanyi type spaces and final coalgebras constructed from satisfied theories. *Electr. Notes Theor. Comput. Sci.*, 106:279–295, 2004.
- [38] Prakash Panangaden. *Labelled Markov Processes*. Imperial College Press, 2009.
- [39] G. D. Plotkin. Lecture notes on domain theory. Available from Plotkin’s web page as The Pisa Notes, 1983.

- [40] Gordon D. Plotkin. Dijkstra's predicate transformers and Smyth's power domains. In *Abstract Software Specifications*, volume 86 of *Lecture Notes in Computer Science*, pages 527–553. Springer, 1979.
- [41] H. Rasiowa and R. Sikorski. A proof of the completeness theorem of gödel. *Fund. Math*, 37:193–200, 1950.
- [42] Lutz Schröder and Dirk Pattinson. Modular algorithms for heterogeneous modal logics. In L. Arge, C. Cachin, T. Jurdzinski, and A. Tarlecki, editors, *34th Int. Colloq. Automata, Languages and Programming (ICALP 2007)*, volume 4596 of *Lecture Notes in Computer Science*, pages 459–471. Springer, 2007.
- [43] Alexandra Silva. *Kleene Coalgebra*. PhD thesis, University of Nijmegen, 2010.
- [44] M. Smyth. Powerdomains and predicate transformers. In J. Diaz, editor, *Proceedings of the International Colloquium On Automata Languages And Programming*, pages 662–676. Springer-Verlag, 1983. *Lecture Notes In Computer Science* 154.
- [45] Marshall H. Stone. The theory of representations for boolean algebras. *Trans. Amer. Math. Soc.*, 40:37–111, 1936.
- [46] F. van Breugel, M. Mislove, J. Ouaknine, and J. Worrell. An intrinsic characterization of approximate probabilistic bisimilarity. In *Proceedings of FOSSACS 03*, volume 2620 of *Lecture Notes In Computer Science*. Springer-Verlag, 2003.
- [47] F. van Breugel and J. Worrell. An algorithm for quantitative verification of probabilistic systems. In *Proceedings of CONCUR'01*, volume 2154 of *Lecture Notes In Computer Science*. Springer-Verlag, 2001.
- [48] Franck van Breugel and James Worrell. Towards quantitative verification of probabilistic systems. In *Proceedings of the Twenty-eighth International Colloquium on Automata, Languages and Programming*. Springer-Verlag, July 2001.
- [49] C. Zhou. *A complete deductive system for probability logic with application to Harsanyi type spaces*. PhD thesis, Indiana University, 2007.
- [50] Chunlai Zhou. Probability logic of finitely additive beliefs. *J. Logic, Language and Information*, 19(3):247–282, 2010.
- [51] Chunlai Zhou and Mingsheng Ying. Approximating Markov processes through filtration. *Theoretical Computer Science*, 446(0):75–97, 2012.

## **Part IV**

# **Stone Dualities for Markov Processes**



Stone duality is a recognized landmark of mathematics. Formally, it states that the category of Boolean algebras and Boolean algebra homomorphisms and the category of Stone spaces and continuous maps are contravariantly equivalent. Jonsson and Tarski [24] extended this to modal logic and showed a duality between Boolean algebras with additional operators and certain topological spaces with additional operators.

Stone duality embodies completeness theorems, but goes far beyond them. Proofs of completeness typically work by constructing an instance of a model from maximal consistent sets of formulas. Stone duality works in the same way, but gives a correspondence not just for syntactically generated algebras, but for *any* suitable algebra. This includes both smaller algebras, which could be finite and generate finite structures, or larger algebras, which could be uncountable, thus not syntactically generated. Furthermore, homomorphisms of algebras give rise to maps between the corresponding structures in the opposite direction. Thus mathematical arguments can be transferred in both directions.

We have seen in the previous chapters that for Markov processes, the natural logic is a simple modal logic in which bounds on probabilities or rates enter into the modalities. We call this logic Markovian logic and we have proposed such a logic for each type of Markov process: probabilistic, subprobabilistic and stochastic.

The Markovian logics can be stripped down to very spartan cores — just the modalities and finite conjunction — and still characterize the corresponding bisimulation for Markov processes [8, 11, 12]. Moreover, we have demonstrated in the previous Chapters that these logics can be completely axiomatized with respect to the MP-semantics.

It is therefore tempting to understand these logics algebraically in the same way that Boolean algebras capture propositional reasoning and the Jonsson–Tarski results [24] give duality for algebras arising from modal logics.

This is exactly the challenge that we take in this Part of the monograph. We develop a Stone-type duality for continuous-space probabilistic transitions systems and a certain kind of algebras that we have named *Aumann algebras*. These are Boolean algebras with operators that behave like the probabilistic modalities of type  $L_r$  used in the previous Part of the monograph.

The representation theorem that establishes the Stone duality, captures and extends the completeness that we have previously proven, in a stronger algebraic format. Our duality only considers the probabilistic case. However, similar generalizations of the completeness theorems can be done both for the subprobabilistic and stochastic case. These results are comprised in Chapter 9.

In Chapter 10 we will extend this duality to include results regarding the behavioural pseudometrics. Concretely we show that a version of Aumann algebra devised with a concept of “norm” or “diameter” of elements corresponds, via the duality, to the concept of metrized Markov process. These results provide a conclusive answer to the question that we have risen in the previous Part of the monograph when we have tried to identify the conditions that guarantee that certain sequences of MPs converge to a limit.

The results summarized in this Part of the Monograph have been published in the following articles.

- [I] D. Kozen, R. Mardare, P. Panangaden. *A Metric Analog of Stone Duality for Markov Processes*. In Proc. of 30th Conference Mathematical Foundation of Programming Semantics, MFPS2014, Electronic Notes in Theoretical Computer Science, ENTCS vol. 308, pages: 211-227, 2014.
- [II] D. Kozen, K. G. Larsen, R. Mardare, P. Panangaden. *Stone Duality for Markov Processes*. In Proc. of 28th Annual ACM/IEEE Symposium on Logic in Computer Science, LICS 2013, IEEE Computer Society, pages: 321-330, 2013.
- [III] D. Kozen, R. Mardare, P. Panangaden. *Strong Completeness for Markovian Logics*. In Proc. of 38th International Symposium Mathematical Foundations of Computer Science, MFCS 2013, Lecture Notes in Computer Science, LNCS 8087, pages: 655-666, 2013.

# Chapter 9

## Classic Stone Duality for Markov Processes

### 9.1 Introduction

In this chapter we demonstrate that behind the strong completeness proofs of Markovian logics presented in the previous Part of the monography, one can identify a Stone duality. We demonstrate this for the probabilistic case. However similar developments can be done both for subprobabilistic and for general- stochastic Markov processes. We call this Stone duality "classic" to differentiate it from the Stone duality that we will prove in the next Chapter, which is a metrized version of the Stone duality.

Stone-type dualities are now recognized as being ubiquitous in computer science. Plotkin [33] and Smyth [37] emphasized that the duality between state-transformer semantics of Dijkstra and predicate-transformer semantics is an instance of a Stone-type duality. Kozen [26] discovered such a duality for probabilistic transition systems. Abramsky [1] identified dualities in domain and concurrency theory. Recently several authors (e.g. [6,36]) have emphasized the duality between logics and transition systems from a coalgebraic perspective. And there have been many recent investigations of Stone-type dualities from the viewpoint of coalgebra and automata theory [5,7,18]. Recent very interesting work by Jacobs [28] has explored convex dualities for probability and quantum mechanics.

A detailed and eloquent introduction to Stone duality and its ramifications, as well as an account of several other related dualities that appear in mathematics, like Gelfand duality, can be found in Johnstone [25].

Regarding the dualities for Markov processes, Mislove et al. [31] found a duality between labeled Markov processes and  $C^*$ -algebras based on the closely related classical Gelfand duality. This work generalizes the concept of testing for MPs.

Our work that we summarize in this Chapter goes in a different direction: we aim at understanding, from an algebraical and topological point of view the strong-completeness

proofs presented in the previous chapters as well as to make sense of the so-called Hennessy-Milner property that has been demonstrated for Markovian logics.

The most closely related work to ours is the work of Goldblatt [20] on the role of the Baire category theorem in completeness proofs, and even more closely his work on deduction systems for coalgebras [21]. This connection has mainly to do with our completeness proof that we have discussed in the previous Part of the monograph. However, in this Chapter we will extend this arguments beyond completeness and identify a Stone duality between a class of Markov processes that we call Stone Markov processes and a special class of Boolean algebra with Operator that we named Aumann algebra.

Our key results presented in this Chapter are:

- a description of a new class of algebras that capture in algebraic form the probabilistic modal (Markovian) logics used for continuous-state Markov processes;
- a version of duality for countable algebras and a certain class of countably-generated Markov processes; and
- a complete axiomatization that does not involve infinitary axiom schemes with uncountably many instances and avoids postulating Lindenbaum's lemma as a meta-axioms.

The duality is represented in the diagram below. Here **SMP** stands for Stone Markov processes and **AA** for countable Aumann algebras. The formal definitions are given in §§9.3–9.4.

$$\begin{array}{ccc} & \mathbb{A} & \\ \text{SMP} & \begin{array}{c} \xrightarrow{\quad} \\ \xleftarrow{\quad} \end{array} & \mathbb{AA}^{\text{op}} \\ & \mathbb{M} & \end{array}$$

A general Markov process gives rise to a countable Aumann algebra but a countable Aumann algebra gives rise to a Stone Markov process. However, starting from a Markov process  $\mathcal{M}$ , constructing the corresponding Aumann algebra  $\mathbb{A}$ , then constructing the corresponding Stone Markov process corresponding  $\mathbb{A}$ , one obtains a Markov process bisimilar to  $\mathcal{M}$ .

### 9.1.1 A Summary for Experts

The duality theorem proved in this Chapter has some novel features that distinguish it from many others that have appeared in the literature.

We have avoided the assumption that every consistent set of formulas can be expanded to a maximal consistent set axioms [21] by using the Rasiowa–Sikorski lemma (whose proof uses the Baire category theorem) in the following way. In going from the

algebra to the dual Markov process, we look at ultrafilters that do *not* respect the infinitary axioms of Aumann algebras. We call these *bad* ultrafilters. We show that these form a meager set (in the standard topological sense) and can be removed without affecting the transition probabilities that we are trying to define. Countability is essential here. In order to show that we do not affect the algebra of clopen sets by doing this, we introduce a distinguished base of clopen sets in the definition of Markov process, which has to satisfy some conditions. We show that this forms an Aumann algebra. We are able to go from a Markov process to an Aumann algebra by using this distinguished base. Morphisms of Markov processes are required to preserve distinguished base elements backwards; that is,

$$\text{if } f : \mathcal{M} \rightarrow \mathcal{N} \text{ and } A \in \mathbb{A}_{\mathcal{N}}, \text{ then } f^{-1}(A) \in \mathbb{A}_{\mathcal{M}}.$$

Thus we get Boolean algebra homomorphisms in the dual for free.

Removing bad points has the effect of destroying compactness of the resulting topological space. We introduce a new concept called *saturation* that takes the place of compactness. The idea is that a saturated model has *all* the good ultrafilters. The Stone dual of an Aumann algebra is saturated, because it is constructed that way. However, it is possible to have a Markov process that is unsaturated but still represents the same algebra. For example, we removed bad points and could, in principle, remove a few more; as long as the remaining points are still dense, we have not changed the algebra. One can saturate a model by a process akin to compactification. We explicitly describe how to do this below.

## 9.2 Stone Duality - An Overview

We begin by recalling the definition of a Boolean algebra.

A *Boolean algebra* (BA) over the set  $B \neq \emptyset$  can be given abstractly as a structure  $\mathcal{B} = (B, \top, \perp, \neg, \vee, \wedge, \leq)$  that satisfy the equational axioms in table 9.1, where  $\top, \perp \in B$ ,  $\neg : B \rightarrow B$  is a monadic operation on  $B$ ,  $\vee, \wedge : B \times B \rightarrow B$  are dyadic operations on  $B$  and  $\leq \subseteq B \times B$  is a relation on  $B$ .

Starting from a Boolean algebra  $\mathcal{B}$ , one can construct a topological space called a *Stone space* as follows.

- The points of the space are the ultrafilters (maximal filters) of  $\mathcal{B}$ , which are in one-to-one correspondence with the Boolean algebra homomorphisms  $\mathcal{B} \rightarrow \mathbb{2}$ , where  $\mathbb{2}$  is the two-element Boolean algebra.
- For each  $x \in \mathcal{B}$ , let  $x'$  be the set of ultrafilters containing  $x$ .
- The sets  $x'$  form a base for the topology. The resulting space is compact, Hausdorff, and totally disconnected.

(BA1): $a \vee (b \vee c) = (a \vee b) \vee c$
(BA2): $a \wedge (b \wedge c) = (a \wedge b) \wedge c$
(BA3): $a \vee b = b \vee a$
(BA4): $a \wedge b = b \wedge a$
(BA5): $a \vee (b \wedge a) = a \wedge (b \vee a) = a$
(BA6): $a \vee (b \wedge c) = (a \vee b) \wedge (a \vee c)$
(BA7): $a \wedge (b \vee c) = (a \wedge b) \vee (a \wedge c)$
(BA8): $a \vee (\neg a) = \top$
(BA8): $a \wedge (\neg a) = \perp$
(BA9): $a \leq b$ iff $a = a \wedge b$ iff $b = a \vee b$

Table 9.1: Boolean Algebra

- The basic open sets  $x'$  are both closed and open (clopens); recall that spaces with a base of clopen sets are *zero-dimensional*.

For compact Hausdorff spaces, the notions of zero-dimensionality and total disconnectedness coincide.

A *Stone space* is defined to be a zero-dimensional compact Hausdorff space. The family of clopen sets of a Stone space forms a Boolean algebra.

One can go back and forth from Boolean algebras to Stone spaces; in both cases one obtains an object isomorphic to the starting object. Jonsson and Tarski [24] extended this to Boolean algebras with additional operators, essentially algebraic versions of modal operators, and corresponding topological spaces equipped with suitable closure operators.

The correspondence lifts to a contravariant equivalence between the category of Boolean algebras and Boolean algebra homomorphisms and the category of Stone spaces and continuous maps. Many other dualities in mathematics are recognized as being of this type [25].

$$\begin{array}{ccc}
 & \mathbf{C} & \\
 \mathbf{SS} & \begin{array}{c} \xrightarrow{\quad} \\ \xleftarrow{\quad} \end{array} & \mathbf{BA}^{\text{op}} \\
 & \mathbf{U} & 
 \end{array}$$

### 9.3 Markov Processes and Markovian Logic

In this section we briefly recall the concept of Markov process and the Markovian logic with which we work in this Chapter. they correspond to PMPs and PML studied in the previous Part of the monograph.

**Markov processes** (MPs) are models of probabilistic systems with a continuous state space and probabilistic transitions [12, 15, 32]. In earlier papers, they were called *labeled* Markov processes to emphasize the fact that there were multiple possible actions, but here we will suppress the labels, as they do not contribute any relevant structure for our results.

Formally, given an analytic space  $(M, \Sigma)$ , where  $\Sigma$  is the Borel algebra induced by the topology, a *Markov process* is a measurable mapping  $\theta \in \llbracket M \rightarrow \Delta(M, \Sigma) \rrbracket$ .

In what follows we identify a Markov process with the tuple  $\mathcal{M} = (M, \Sigma, \theta)$ ;  $M$  is the *support set*, denoted by  $\text{supp}(\mathcal{M})$ , and  $\theta$  is the *transition function*. For  $m \in M$ ,  $\theta(m) : \Sigma \rightarrow [0, 1]$  is a probability measure on the state space  $(M, \Sigma)$ . For  $N \in \Sigma$ , the value  $\theta(m)(N) \in [0, 1]$  represents the probability of a transition from  $m$  to a state in  $N$ .

The condition that  $\theta$  is a measurable function  $\llbracket M \rightarrow \Delta(M, \Sigma) \rrbracket$  is equivalent to the condition that for fixed  $N \in \Sigma$ , the function  $m \mapsto \theta(m)(N)$  is a measurable function  $\llbracket M \rightarrow [0, 1] \rrbracket$  (see e.g. Proposition 2.9 of [15]).

The **Markovian logic** (ML) that we discussed in this Chapter is a multi-modal logic for semantics based on MPs [2, 3, 8, 17, 22, 28, 30, 41]. In addition to the Boolean operators, this logic is equipped with *probabilistic modal operators*  $L_r$  for  $r \in \mathbb{Q}_0$  that bound the probabilities of transitions. Intuitively, the formula  $L_r \varphi$  is satisfied by  $m \in \mathcal{M}$  whenever the probability of a transition from  $m$  to a state satisfying  $\varphi$  is at least  $r$ .

**[Syntax]:** The formulas of ML are collected in the set  $\mathcal{L}$  defined inductively, for a set  $\mathcal{P}$  of atomic propositions, by the grammar

$$\varphi ::= p \mid \perp \mid \varphi \rightarrow \varphi \mid L_r \varphi$$

where  $p$  can be any element of  $\mathcal{P}$  and  $r$  any element of  $\mathbb{Q}_0$ .

The Boolean operators  $\vee$ ,  $\wedge$ ,  $\neg$ , and  $\top$  are defined from  $\rightarrow$  and  $\perp$  as usual. And for  $r_1, \dots, r_n \in \mathbb{Q}_0$  and  $\varphi \in \mathcal{L}$ , let

$$L_{r_1 \dots r_n} \varphi = L_{r_1} \dots L_{r_n} \varphi.$$

**[Markovian semantics]:** for  $\mathcal{L}$  is defined for a given MP  $\mathcal{M} = (M, \Sigma, \theta)$  and  $m \in M$ , on top of an interpretation function  $i : M \rightarrow 2^{\mathcal{P}}$ , as follows.

- $\mathcal{M}, m, i \models p$  if  $p \in i(m)$ ,
- $\mathcal{M}, m, i \models \perp$  never,

- $\mathcal{M}, m, i \models \varphi \rightarrow \psi$  if  $\mathcal{M}, m, i \models \psi$  whenever  $\mathcal{M}, m, i \models \varphi$ ,
- $\mathcal{M}, m, i \models L_r \varphi$  if  $\theta(m)(\llbracket \varphi \rrbracket) \geq r$ ,  
where  $\llbracket \varphi \rrbracket = \{m \in M \mid \mathcal{M}, m, i \models \varphi\}$ .

It follows that

- $\mathcal{M}, m, i \models \top$  always,
- $\mathcal{M}, m, i \models \varphi \wedge \psi$  iff  $\mathcal{M}, m, i \models \varphi$  and  $\mathcal{M}, m, i \models \psi$ ,
- $\mathcal{M}, m, i \models \varphi \vee \psi$  iff  $\mathcal{M}, m, i \models \varphi$  or  $\mathcal{M}, m, i \models \psi$ ,
- $\mathcal{M}, m, i \models \neg \varphi$  iff not  $\mathcal{M}, m, i \models \varphi$ .

Given  $\mathcal{M} = (M, \Sigma, \theta)$  and  $i$ , we say that  $m \in M$  satisfies  $\varphi \in \mathcal{L}$  if  $\mathcal{M}, m, i \models \varphi$ . We write  $\mathcal{M}, m, i \not\models \varphi$  if it is not the case that  $\mathcal{M}, m, i \models \varphi$ .

For  $\Phi \subseteq \mathcal{L}$ , we write  $\mathcal{M}, m, i \models \Phi$  if  $\mathcal{M}, m, i \models \varphi$  for all  $\varphi \in \Phi$ .

For  $\Phi \subseteq \mathcal{L}$  and  $\varphi \in \mathcal{L}$ , we write  $\Phi \models \varphi$  if  $\mathcal{M}, m, i \models \varphi$  whenever  $\mathcal{M}, m, i \models \Phi$ .

A formula or set of formulas is *satisfiable* if there exist an MP  $\mathcal{M}$ , an interpretation function  $i$  for  $\mathcal{M}$  and  $m \in \text{supp}(\mathcal{M})$  that satisfies it. We say that  $\varphi \in \mathcal{L}$  is *valid* and write  $\models \varphi$  if  $\emptyset \models \varphi$ , that is, if  $\neg \varphi$  is not satisfiable.

We now recall the axiomatization of ML for Markovian semantics that we have discussed in the previous part of our monograph. The system is a Hilbert-style system consisting of the axioms and rules of propositional logic and the axioms and rules listed in Table 9.2. The axioms and the rules are stated for arbitrary  $\varphi, \psi \in \mathcal{L}$  and arbitrary  $r, s \in \mathbb{Q}_0 = \mathbb{Q} \cap [0, 1]$ .

$$(A1): \vdash L_0 \varphi$$

$$(A2): \vdash L_r \top$$

$$(A3): \vdash L_r \varphi \rightarrow \neg L_s \neg \varphi, \quad r + s > 1$$

$$(A4): \vdash L_r(\varphi \wedge \psi) \wedge L_s(\varphi \wedge \neg \psi) \rightarrow L_{r+s} \varphi, \quad r + s \leq 1$$

$$(A5): \vdash \neg L_r(\varphi \wedge \psi) \wedge \neg L_s(\varphi \wedge \neg \psi) \rightarrow \neg L_{r+s} \varphi, \quad r + s \leq 1$$

$$(R1): \frac{\vdash \varphi \rightarrow \psi}{\vdash L_r \varphi \rightarrow L_r \psi}$$

$$(R2): \{L_{r_1 \dots r_n} \psi \mid r < s\} \vdash L_{r_1 \dots r_n} \psi$$

Table 9.2: The axioms of  $\mathcal{L}$

If  $\Phi \subseteq \mathcal{L}$  and  $\varphi \in \mathcal{L}$ , we write  $\Phi \vdash \varphi$  and say that  $\Phi$  *derives*  $\varphi$  if  $\varphi$  is provable from the axioms and the extra assumptions  $\Phi$ . We write  $\vdash \varphi$  if  $\emptyset \vdash \varphi$ . Obviously,  $\vdash$  is a deducibility relation in the sense of the definition in Preliminaries, section 3.7.

A formula or set of formulas is *consistent* if it cannot derive  $\perp$ . We say that  $\Phi \subseteq \mathcal{L}$  is *maximally consistent* if it is consistent and it has no proper consistent extensions.

The set  $\Phi$  of formulas is *filtered* if for all  $\varphi, \psi \in \Phi$  there exists  $\rho \in \Phi$  with  $\vdash \rho \rightarrow \varphi \wedge \psi$ .

We remind the reader that in the previous part of the monograph we have proved that the axiomatic system in Table 9.2 is sound and strongly-complete for the MP-semantics. Moreover, the logical equivalence induced by ML on the class of MPs coincides with bisimulation equivalence.

In what follows we will generalize these results in the form of a Stone duality.

## 9.4 Aumann Algebras

In this section we introduce an algebraic version of Markovian logic consisting of Boolean algebra with operators  $F_r$  for  $r \in \mathbb{Q}_0$  corresponding to the operators  $L_r$  of ML. We call this Aumann Algebra (AA) in honor of Robert Aumann, who has made fundamental contributions to probabilistic logic [2,3].

### 9.4.1 Definition of Aumann Algebras

**Definition 9.4.1** (Aumann algebra). *An Aumann algebra (AA) is a structure*

$$\mathbb{A} = (A, \rightarrow, \perp, \{F_r\}_{r \in \mathbb{Q}_0}, \leq)$$

where

- $(A, \rightarrow, \perp, \leq)$  is a Boolean algebra;
- for each  $r \in \mathbb{Q}_0$ ,  $F_r : A \rightarrow A$  is a unary operator; and
- the axioms in Table 9.3 hold for all  $a, b \in A, r, s, r_1, \dots, r_n \in \mathbb{Q}_0$ .

The Boolean operations  $\vee, \wedge, \neg$ , and  $\top$ , are defined from  $\rightarrow$  and  $\perp$  as usual.

Morphisms of Aumann algebras are Boolean algebra homomorphisms that commute with the operations  $F_r$ . The category of Aumann algebras and Aumann algebra homomorphisms is denoted by **AA**.

We abbreviate  $F_{r_1} \cdots F_{r_n} a$  by  $F_{r_1 \dots r_n} a$ .

The operator  $F_r$  is the algebraic counterpart of the logical modality  $L_r$ . The first two axioms state tautologies, while the third captures the way  $F_r$  interacts with negation. Axioms (AA4) and (AA5) assert finite additivity, while (AA6) asserts monotonicity.

(AA1): $\top \leq F_0 a$
(AA2): $\top \leq F_r \top$
(AA3): $F_r a \leq \neg F_s \neg a, r + s > 1$
(AA4): $F_r(a \wedge b) \wedge F_s(a \wedge \neg b) \leq F_{r+s} a, r + s \leq 1$
(AA5): $\neg F_r(a \wedge b) \wedge \neg F_s(a \wedge \neg b) \leq \neg F_{r+s} a, r + s \leq 1$
(AA6): $a \leq b \Rightarrow F_r a \leq F_r b$
(AA7): $(\bigwedge_{r < s} F_{r_1 \dots r_n r} a) = F_{r_1 \dots r_n s} a$

Table 9.3: Aumann algebra

The most interesting axiom is the infinitary axiom (AA7). It asserts that  $F_{r_1 \dots r_n s} a$  is the greatest lower bound of the set  $\{F_{r_1 \dots r_n r} a \mid r < s\}$  with respect to the natural order  $\leq$ . In SMPs, it will imply countable additivity.

Notice that (AA1)–(AA5) are the algebraic version of (A1)–(A5), (AA6) is the algebraic version of (R1) and (AA7) restate the algebraic property of (R2).

The following lemma establishes some basic consequences.

**Lemma 9.4.2.** *Let  $\mathbb{A} = (A, \rightarrow, \perp, \{F_r\}_{r \in \mathbb{Q}_0}, \leq)$  be an Aumann algebra. For all  $a, b \in A$  and  $r, s \in \mathbb{Q}_0$ ,*

- (i)  $F_r \perp = \perp$  for  $r > 0$ ;
- (ii) if  $r \leq s$ , then  $F_s a \leq F_r a$ ;
- (iii) if  $a \leq \neg b$  and  $r + s > 1$ , then  $F_r a \leq \neg F_s b$ .

*Proof.* 1. For  $s > 0$  there exists  $r \in [0, 1]$  with  $r + s > 1$ . (AA3) instantiated with  $a = \top$  gives us  $F_r \top \leq \neg F_s \perp$  and using (AA2) we obtain  $\top \leq \neg F_s \perp$ , thus  $F_s \perp \leq \perp$  and then  $F_s \perp = \perp$ .

2. We use (AA5) instantiated with  $b = \top$  and negate both sides to flip the inequality to get:

$$F_{r+s} a \leq \neg(\neg F_r(a \wedge \top) \wedge \neg F_s(a \wedge \perp)).$$

From de Morgan's law and (1) we get

$$F_{r+s} a \leq F_r(a) \vee F_s(\perp) = F_r a \vee \perp = F_r a.$$

3. Since  $a \leq \neg b$ , using (AA6) we get  $F_r a \leq F_r \neg b$ . Applying (AA3) and using the transitivity of  $\leq$  we obtain  $F_r a \leq \neg F_s b$ .  $\square$

As expected, the formulas of Markovian logic modulo logical equivalence form a free countable Aumann algebra. Define  $\equiv$  on formulas by:

$$\varphi \equiv \psi \text{ if } \vdash \varphi \rightarrow \psi \text{ and } \vdash \psi \rightarrow \varphi.$$

Let  $[\varphi]$  denote the equivalence class of  $\varphi$  modulo  $\equiv$ , and let

$$\mathcal{L}/\equiv = \{[\varphi] \mid \varphi \in \mathcal{L}\}.$$

By (R1), the modality  $L_r$  is well defined on  $\equiv$ -classes. The Boolean operators are also well defined by considerations of propositional logic.

**Theorem 9.4.3.** *The structure*

$$(\mathcal{L}/\equiv, \rightarrow, [\perp], \{L_r\}_{r \in \mathbb{Q}_0}, \leq)$$

is an Aumann algebra, where

$$[\varphi] \leq [\psi] \text{ iff } \vdash \varphi \rightarrow \psi.$$

*Proof.* The axioms of Boolean algebras are trivially satisfied.

Now we have to verify the axioms in Table 9.3 for  $F_r = L_r$ .

(AA1) in this setting corresponds to  $\vdash T \rightarrow L_0\varphi$ , which is equivalent to  $\vdash L_0\varphi$  and an instance of axiom (A1) of  $\mathcal{L}$ .

Similarly, (AA2), (AA3), (AA4) and (AA5) are easily seen to correspond to axioms (A2), (A3), (A4) and (A5) respectively. (AA6) and (AA7) are instances of the rules (R1) and (R2) respectively. Consequently all the axioms of Aumann algebra are verified.

Moreover,  $\mathcal{L}$  is countable. □

## 9.4.2 Extended Satisfiability Principles for Aumann Algebras

We saw in the previous section that the formulas of ML give rise to an Aumann algebra. We now define a satisfiability relation for AAs that generalizes the satisfiability relation for ML and prove the corresponding soundness result for MPs.

Let  $\mathcal{M} = (M, \Sigma, \theta)$  be a MP and let  $\llbracket \cdot \rrbracket$  be an interpretation of terms in the language of Aumann algebras as measurable sets in  $M$  in which the Boolean operators have their usual set-theoretic interpretation and for  $S \in \Sigma$  and  $t \in \mathbb{Q} \cap [0, 1]$ ,

$$\begin{aligned} \llbracket F_t \rrbracket(S) &= \{m \in M \mid \theta(m)(S) \geq t\} \\ \llbracket F_t a \rrbracket &= \llbracket F_t \rrbracket(\llbracket a \rrbracket). \end{aligned}$$

If  $\mathbb{A}$  is freely generated, then an interpretation function is defined by any function that associates to each generator of  $\mathbb{A}$  a subset of  $M$ .

**Lemma 9.4.4.**  $\llbracket F_t \rrbracket$  preserves meets of countable measurable chains; that is, if  $S_i$ ,  $i \in I$  is a countable chain of measurable sets, then

$$\llbracket F_t \rrbracket \left( \bigcap_{i \in I} S_i \right) = \bigcap_{i \in I} \llbracket F_t \rrbracket (S_i).$$

*Proof.* We have

$$\begin{aligned} \llbracket F_t \rrbracket \left( \bigcap_{i \in I} S_i \right) &= \{m \in M \mid \theta(m) \left( \bigcap_{i \in I} S_i \right) \geq t\} \\ &= \{m \in M \mid \inf_{i \in I} \theta(m)(S_i) \geq t\} \\ &= \{m \in M \mid \forall i \in I, \theta(m)(S_i) \geq t\} \\ &= \bigcap_{i \in I} \{m \in M \mid \theta(m)(S_i) \geq t\} \\ &= \bigcap_{i \in I} \llbracket F_t \rrbracket (S_i). \end{aligned}$$

□

We can prove now that the axioms of Aumann algebras are sound for Markov processes.

**Theorem 9.4.5 (Soundness).** Let  $\mathbb{A}$  be an Aumann algebra and  $a \in \mathbb{A}$ . If  $\top \leq a$ , then for any Markov process  $\mathcal{M} = (M, \Sigma, \theta)$  and any interpretation  $\llbracket \cdot \rrbracket$  of terms in the language of Aumann algebras as measurable sets in  $M$  with the properties listed above,  $\llbracket a \rrbracket = M$ .

*Proof.* As in the case of soundness for logics, we prove the soundness by showing that each axiom is satisfied by any Markov process.

In the following, we only prove it for (AA7).

Let  $\mathcal{F}$  be the set of terms of the form

$$\alpha^r = F_{t_1 \dots t_n r} a$$

for  $a \in A$  and  $0 \leq t_1, \dots, t_n, r \leq 1$ . We consider this formula parameterized by  $r$ ; that is, if  $\alpha^r = F_{t_1 \dots t_n r} a$ , then  $\alpha^s$  denotes  $F_{t_1 \dots t_n s} a$ .

The axiomatization of AAs includes all infinitary conditions of the form

$$\alpha^s = \bigwedge_{r < s} \alpha^r.$$

To prove the soundness of (AA7), it is sufficient to prove that

$$\llbracket \alpha^s \rrbracket = \llbracket \bigwedge_{r < s} \alpha^r \rrbracket.$$

We proceed by induction on  $n$ .

For the basis,

$$\begin{aligned} \llbracket F_s a \rrbracket &= \llbracket F_s \rrbracket(\llbracket a \rrbracket) = \{u \mid \theta(u)(\llbracket a \rrbracket) \geq s\} \\ &= \{u \mid \forall r < s \theta(u)(\llbracket a \rrbracket) \geq r\} = \bigcap_{r < s} \{u \mid \theta(u)(\llbracket a \rrbracket) \geq r\} \\ &= \bigcap_{r < s} \llbracket F_r a \rrbracket = \llbracket \bigwedge_{r < s} F_r a \rrbracket. \end{aligned}$$

For the induction step, if  $\llbracket \alpha^s \rrbracket = \llbracket \bigwedge_{r < s} \alpha^r \rrbracket$  and  $t > 0$ ,

$$\begin{aligned} \llbracket F_t \alpha^s \rrbracket &= \llbracket F_t \rrbracket(\llbracket \alpha^s \rrbracket) = \llbracket F_t \rrbracket(\llbracket \bigwedge_{r < s} \alpha^r \rrbracket) \\ &= \llbracket F_t \rrbracket(\bigcap_{r < s} \llbracket \alpha^r \rrbracket) = \bigcap_{r < s} \llbracket F_t \rrbracket(\llbracket \alpha^r \rrbracket) \quad \text{by Lemma 9.4.4} \\ &= \bigcap_{r < s} \llbracket F_t \alpha^r \rrbracket = \llbracket \bigwedge_{r < s} F_t \alpha^r \rrbracket. \end{aligned}$$

□

## 9.5 Stone Markov Processes

In our duality theory, we work with Markov processes constructed from certain zero-dimensional Hausdorff spaces. We call such structures *Stone–Markov processes* (SMPs), even if the underlying topology is not, in fact, compact.

### 9.5.1 MPs with Distinguished Base

We restrict our attention to Markov processes  $(M, \mathbb{A}, \theta)$ , where  $\mathbb{A}$  is a distinguished countable base of clopen sets that is closed under the set-theoretic Boolean operations and the operations

$$F_r(A) = \{m \mid \theta(m)(A) \geq r\}, \quad r \in \mathbb{Q} \cap [0, 1].$$

The measurable sets  $\Sigma$  are the Borel sets of the topology generated by  $\mathbb{A}$ .

Morphisms of such spaces are required to preserve the distinguished base; thus a morphism  $f : \mathcal{M} \rightarrow \mathcal{N}$  is a continuous function such that

- for all  $m \in M$  and  $B \in \Sigma_{\mathcal{N}}$ ,

$$\theta_{\mathcal{M}}(m)(f^{-1}(B)) = \theta_{\mathcal{N}}(f(m))(B);$$

- for all  $A \in \mathbb{A}_{\mathcal{N}}$ ,  $f^{-1}(A) \in \mathbb{A}_{\mathcal{M}}$ .

### 9.5.2 Saturation

Unlike Stone spaces, SMPs are not topologically compact, but we do postulate a completeness property that is a weak form of compactness, which we call *saturation*. One can saturate a given SMP by a completion procedure that is reminiscent of Stone–Čech compactification. Intuitively, one adds points to the structure without changing the represented algebra. An MP is *saturated* if it is maximal with respect to this operation. The term comes from a related concept of the same name in model theory [9, Ch. 5].

One can define the saturation by completing by a certain family of ultrafilters of called *good ultrafilters*. These are ultrafilters respecting the infinitary condition (AA7) in the definition of Aumann algebras (§9.4). All principal ultrafilters of an SMP are already good, and one must only add the rest. The details of this construction are given in §9.6.

However, one can give a more conceptual definition of saturation.

Formally, consider MP morphisms  $f : \mathcal{M} \rightarrow \mathcal{N}$  such that

- $f$  is a homeomorphism between  $\mathcal{M}$  and its image in  $\mathcal{N}$ ;
- the image  $f(\mathcal{M})$  is dense in  $\mathcal{N}$ ; and
- $f$  preserves the distinguished base in the forward direction as well as the backward; that is, if  $A \in \mathbb{A}_{\mathcal{M}}$ , then there exists  $B \in \mathbb{A}_{\mathcal{N}}$  such that  $A = f^{-1}(B)$ .

Call such a morphism a *strict embedding*. The collection of all  $\mathcal{N}$  such that there exists a strict embedding  $\mathcal{M} \rightarrow \mathcal{N}$  contains a final object, which is the colimit of the strict embeddings  $\mathcal{M} \rightarrow \mathcal{N}$ . This is the saturation of  $\mathcal{M}$ .

### 9.5.3 Definition of SMP

The saturated spaces underline what we will call Stone-Markov processes.

**Definition 9.5.1** (Stone–Markov Process). *A Markov process  $\mathcal{M} = (M, \mathbb{A}, \theta)$  with distinguished base is a Stone–Markov process (SMP) if it is saturated.*

*The morphisms of SMPs are just the morphisms of MPs with distinguished base as defined above.*

*The category of SMPs and SMP morphisms is denoted **SMP**.*

## 9.6 Stone Duality

In this section we describe the duality between SMPs and countable AAs. This is in the spirit of the classical Stone representation theorem [45], or, more precisely, the representation theorem of Jonsson and Tarski [30] for Boolean algebras with operators. Here the details are somewhat different, as we must deal with measure theory.

### 9.6.1 From AAs to SMPs

Firstly, we prove that we can construct an SMP from an Aumann algebra.

For this subsection, we fix an arbitrary countable Aumann algebra

$$\mathbb{A} = (A, \rightarrow, \perp, \{F_r\}_{r \in \mathbb{Q}_0}, \leq).$$

Let  $\mathcal{U}^*$  be the set of all ultrafilters of  $\mathbb{A}$ . The classical Stone construction gives a Boolean algebra of sets isomorphic to  $\mathbb{A}$  with elements

$$\begin{aligned} \langle a \rangle^* &= \{u \in \mathcal{U}^* \mid a \in u\}, \quad a \in \mathbb{A} \\ \text{and let } \langle \mathbb{A} \rangle^* &= \{\langle a \rangle^* \mid a \in A\}. \end{aligned}$$

The sets  $\langle a \rangle^*$  generate a Stone topology  $\tau^*$  on  $\mathcal{U}^*$ , and the sets  $\langle a \rangle^*$  are exactly the clopen sets of the topology.

Let  $\mathcal{F}$  be the set of elements of the form

$$\alpha^r = F_{t_1 \dots t_n r} a$$

for  $a \in A$  and  $t_1, \dots, t_n, r \in \mathbb{Q} \cap [0, 1]$ .

As before, we consider this term as parameterized by  $r$ ; that is, if  $\alpha^r = F_{t_1 \dots t_n r} a$ , then  $\alpha^s$  denotes  $F_{t_1 \dots t_n s} a$ .

The set  $\mathcal{F}$  is countable since  $A$  is. Axiom (AA7) asserts all infinitary conditions of the form

$$\alpha^s = \bigwedge_{r < s} \alpha^r. \tag{9.6.1}$$

for  $\alpha^s \in \mathcal{F}$ .

Let us call an ultrafilter  $u$  *bad* if it violates one of these conditions in the sense that for some  $\alpha^s \in \mathcal{F}$ ,  $\alpha^r \in u$  for all  $r < s$  but  $\alpha^s \notin u$ . Otherwise,  $u$  is called *good*.

Let  $\mathcal{U}$  be the set of good ultrafilters of  $\mathbb{A}$ .

Let  $\tau = \{B \cap \mathcal{U} \mid B \in \tau^*\}$  be the subspace topology on  $\mathcal{U}$ , and let

$$\begin{aligned} \langle a \rangle &= \{u \in \mathcal{U} \mid a \in u\} = \langle a \rangle^* \cap \mathcal{U} \\ \langle \mathbb{A} \rangle &= \{\langle a \rangle \mid a \in A\}. \end{aligned}$$

Then  $\tau$  is countably generated by the sets  $\langle a \rangle$  and all  $\langle a \rangle$  are clopen in the subspace topology.

**Lemma 9.6.1.** *If  $\bar{S}$  denotes the closure of  $S$  in  $\tau^*$ , then*

$$\langle \neg \alpha^s \rangle^* = \overline{\bigcup_{r < s} \langle \neg \alpha^r \rangle^*}.$$

*Proof.* ( $\supseteq$ ): From Lemma 9.4.2 and (AA6) we have that  $\neg\alpha^r \leq \neg\alpha^s$  for any  $r < s$ . Consequently, for any  $u \in \mathcal{U}^*$ ,  $u \ni \neg\alpha^r$  implies  $u \ni \neg\alpha^s$ .

( $\subseteq$ ): It is sufficient to prove that for every open set  $B \in \tau^*$ ,  
if  $B \cap (\neg\alpha^s) \neq \emptyset$ , then  $B \cap (\neg\alpha^r) \neq \emptyset$  for some  $r < s$ .

Proving this for  $B \in (\mathbb{A})^*$  is sufficient.

Let  $B = (a)^*$  and suppose that  $(a)^* \cap (\neg\alpha^s) \ni u$ . Then,

$$a \wedge \neg\alpha^s \in u, \text{ implying } a \wedge \neg\alpha^s \neq \perp.$$

Applying Rasiowa-Sikorski lemma we obtain that

there exists  $v \in \mathcal{U}$  such that  $a \wedge \neg\alpha^s \in v$ .

Consequently,  $a, \neg\alpha^s \in v$  and since  $v$  is a good ultrafilter, there exists  $r < s$  such that  $\neg\alpha^r \in v$ . Hence  $v \in (a)^* \cap (\neg\alpha^s) \neq \emptyset$ .  $\square$

The next lemma asserts that the set  $\mathcal{U} \setminus \mathcal{U}^*$  of bad ultrafilters is meager. We will use this to argue that  $\mathcal{U}$  is dense in  $\mathcal{U}^*$ , therefore no  $(a)$  vanishes as a result of dropping the bad points from  $(a)^*$ . It will follow that  $(\mathbb{A})$  and  $(\mathbb{A})^*$  are isomorphic as Boolean algebras.

**Lemma 9.6.2.** *The set  $\mathcal{U}^* \setminus \mathcal{U}$  is of first category in the Stone topology  $\tau^*$ . In particular,  $\mathcal{U}$  is dense in  $\mathcal{U}^*$ .*

*Proof.* We must prove that  $\mathcal{U}^* \setminus \mathcal{U}$  is a countable union of nowhere dense sets.

Since  $\mathcal{A}$  countable, the set  $\mathcal{F}$  is countable as well. Each bad ultrafilter  $u \in \mathcal{U}^* \setminus \mathcal{U}$  violates at least one constraint (9.6.1), thus

$$\mathcal{U}^* \setminus \mathcal{U} = \bigcup_{\alpha^s \in \mathcal{F}} U_{\alpha^s},$$

where

$$\begin{aligned} U_{\alpha^s} &= \{u \in \mathcal{U}^* \mid \alpha^s \notin u \text{ and } \forall r < s \alpha^r \in u\} \\ &= (\neg\alpha^s)^* \setminus \bigcup_{r < s} (\neg\alpha^r)^*. \end{aligned}$$

Now we argue that each  $U_{\alpha^s}$  is nowhere dense.

In  $\tau^*$ ,  $(\neg\alpha^s)^*$  is a closed set while  $\bigcup_{r < s} (\neg\alpha^r)^*$  is a countable union of open sets, hence open.

Applying Lemma 9.6.1,  $U_{\alpha^s}$  is the boundary of an open set, hence nowhere dense.  $\square$

In the remainder of this section, we show that the topological space  $(\mathcal{U}, \tau)$  is a zero-dimensional Hausdorff space with base  $(\mathbb{A})$  and that this space gives rise to an SMP  $(\mathcal{U}, (\mathbb{A}), \theta)$ . Note that unlike  $(\mathcal{U}^*, \tau^*)$ , the space  $(\mathcal{U}, \tau)$  is not compact, but it is saturated.

In order to prove these claims, we need to focus on the properties of the subspace topology.

Recall that for a topological space  $(X, \tau)$  and an arbitrary set  $Y \subseteq X$ , we can view  $Y$  as a topological space with the relative topology  $\tau' = \{u \cap Y \mid u \in \tau\}$ . Moreover, if  $u$  is an open (closed) set in  $\tau$ , then  $u \cap Y$  is open (closed) in  $\tau'$ , and if  $B \subseteq \tau$  is a base for  $\tau$ , then  $B' = \{u \cap Y \mid u \in B\}$  is a base for  $\tau'$ .

The next lemma is immediate from the definitions.

**Lemma 9.6.3.** *Any subspace of a zero-dimensional space is zero-dimensional, and any subspace of a Hausdorff space is Hausdorff.*

*Proof.* Let  $(X, \tau)$  space and  $(Y, \tau')$  a subspace with the relative topology.

Suppose that  $(X, \tau)$  is zero-dimensional.

For all  $u' \in \tau'$ , there exists  $u \in \tau$  such that  $u' = u \cap Y$ . Since  $(X, \tau)$  has a base of clopen sets, we have  $u = \bigcup_{i \in I} u_i$ , where  $(u_i)$  is an indexed collection of clopens in  $\tau$ . Then

$$u' = \left( \bigcup_{i \in I} u_i \right) \cap Y = \bigcup_{i \in I} (u_i \cap Y).$$

Since the  $u_i$  are clopen in  $\tau$ ,  $u_i \cap Y$  are clopen in  $\tau'$ , thus the clopens of  $\tau'$  form a base for the topology  $\tau'$ . Therefore  $(Y, \tau')$  is zero-dimensional.

Suppose that  $(X, \tau)$  is Hausdorff.

For all  $x, y \in Y$  there exist disjoint  $u, v \in \tau$  such that  $x \in u$  and  $y \in v$ . Then  $x \in u \cap Y$ ,  $y \in v \cap Y$ , and  $u \cap Y$  and  $v \cap Y$  are disjoint and in  $\tau'$ . As  $x, y$  were arbitrary,  $(Y, \tau')$  is Hausdorff.  $\square$

Since  $(\mathcal{U}, \tau)$  is a subspace of the Stone space  $(\mathcal{U}^*, \tau^*)$ , the next proposition is just an instance of Lemma 9.6.3.

**Proposition 9.6.4.** *The space  $(\mathcal{U}, \tau)$  is a zero-dimensional Hausdorff space.*

## 9.6.2 Construction of $\mathbb{M}(\mathbb{A})$

We can now form a Markov process

$$\mathbb{M}(\mathbb{A}) = (\mathcal{U}, \Sigma, \theta),$$

where  $\Sigma$  is the  $\sigma$ -algebra generated by  $(\mathbb{A})$ .

To define the measure  $\theta(u)$  for an ultrafilter  $u \in \mathcal{U}$ , we need to prove some additional results.

**Lemma 9.6.5.** *For all  $a \in A$  and  $u \in \mathcal{U}$ , the set*

$$\{r \in \mathbb{Q} \cap [0, 1] \mid F_r a \in u\}$$

*is nonempty and closed downward in the natural order on  $\mathbb{Q} \cap [0, 1]$ .*

*Proof.* The set contains at least 0 by (AA1).

Downward closure follows from Lemma 9.4.2(ii). □

It follows that  $\{r \in \mathbb{Q}_0 \mid \neg F_r a \in u\}$  is closed upward. Thus we can define the function  $\theta : \mathcal{U} \rightarrow [0, 1]$  by

$$\begin{aligned} \theta(u)(\langle a \rangle) &= \sup\{r \in \mathbb{Q} \cap [0, 1] \mid F_r a \in u\} \\ &= \inf\{r \in \mathbb{Q} \cap [0, 1] \mid \neg F_r a \in u\}. \end{aligned}$$

The aforementioned equality is proven in details for the Markovian logics in the previous part of this monograph – see Lemma 7.6.3 in Chapter 7.

Note that  $\theta(u)(\langle a \rangle)$  is not necessarily rational. In the following, we use Theorem 3.2.3 to show that  $\theta$  can be uniquely extended to a transition function. This will allow us to construct a Markov process on the space of good ultrafilters.

**Lemma 9.6.6.** *The set  $\langle \mathbb{A} \rangle$  is a field of sets, and for all  $u \in \mathcal{U}$ , the function  $\theta(u)$  is a finitely additive set function.*

*Proof.* That the set  $\langle \mathbb{A} \rangle$  is a field of sets is immediate from the classic Stone representation theorem and the fact that  $\langle \mathbb{A} \rangle$  is dense in  $\langle \mathbb{A} \rangle^*$ .

To show finite additivity, suppose  $a, b \in A$  and  $\langle a \rangle \cap \langle b \rangle = \emptyset$ . Then  $a \wedge b = 0$ . We wish to show that

$$\theta(u)(\langle a \vee b \rangle) = \theta(u)(\langle a \rangle) + \theta(u)(\langle b \rangle).$$

It suffices to show the inequality in both directions.

For  $\leq$ , by the definition of  $\theta$ , it suffices to show that

$$\begin{aligned} &\sup\{t \mid F_t(a \vee b) \in u\} \\ &\leq \inf\{r \mid \neg F_r a \in u\} + \inf\{s \mid \neg F_s b \in u\} \\ &= \inf\{r + s \mid \neg F_r a \in u \text{ and } \neg F_s b \in u\} \\ &= \inf\{r + s \mid \neg F_r a \wedge \neg F_s b \in u\}; \end{aligned}$$

that is, if  $F_t(a \vee b) \in u$  and  $\neg F_r a \wedge \neg F_s b \in u$ , then  $t \leq r + s$ .

But

$$\begin{aligned} \neg F_r a \wedge \neg F_s b &= \neg F_r((a \vee b) \wedge a) \wedge \neg F_s((a \vee b) \wedge \neg a) \\ &\leq \neg F_{r+s}(a \vee b) \quad \text{by (AA5),} \end{aligned}$$

thus  $\neg F_{r+s}(a \vee b) \in u$ , and  $t \leq r + s$  follows from the characterization of Lemma 9.6.5.

The inequality in the opposite direction is similar, using (AA4). We need to show

$$\inf\{t \mid \neg F_t(a \vee b) \in u\} \geq \sup\{r + s \mid F_r a \wedge F_s b \in u\};$$

that is, if  $\neg F_t(a \vee b) \in u$  and  $F_r a \wedge F_s b \in u$ , then  $t \geq r + s$ . But

$$\begin{aligned} F_r a \wedge F_s b &= F_r((a \vee b) \wedge a) \wedge F_s((a \vee b) \wedge \neg a) \\ &\leq F_{r+s}(a \vee b) \quad \text{by (AA4),} \end{aligned}$$

thus  $F_{r+s}(a \vee b) \in u$ , and again  $r + s \leq t$  follows from the characterization of Lemma 9.6.5.  $\square$

The following is the key technical lemma where we use the fact that we have removed the bad ultrafilters.

**Lemma 9.6.7.** *For  $u \in \mathcal{U}$ ,  $\theta(u)$  is continuous from above at  $\emptyset$  relatively to the field  $(\mathbb{A})$ .*

*Proof.* We prove that if  $u \in \mathcal{U}$  (it is a good ultrafilter) and  $b_0 \geq b_1 \geq \dots$  with  $\bigcap_i (b_i) = \emptyset$ , then

$$\inf_i \theta(u)((b_i)) = 0.$$

Consider the countable set  $\mathcal{F}$  of elements of the form

$$\alpha^r = F_{t_1 \dots t_n} r a$$

for  $a \in A$  and rational  $t_1, \dots, t_n, r \geq 0$ , parameterized by  $r$ .

If  $r < s$ , then  $\alpha^s \leq \alpha^r$ . Using (AA4),

$$\theta(u)((\alpha^r \wedge \neg \alpha^s)) \leq \theta(u)((\alpha^r)) - \theta(u)((\alpha^s)). \quad (9.6.2)$$

Since  $u$  is good,  $F_t \alpha^r \in u$  for all  $r < s$  iff  $F_t \alpha^s \in u$ , therefore

$$\theta(u)((\alpha^s)) = \inf_{r < s} \theta(u)((\alpha^r)). \quad (9.6.3)$$

Let  $\varepsilon > 0$  be an arbitrarily small positive number. For each  $\alpha \in \mathcal{F}$  and  $s \in \mathbb{Q} \cap [0, 1]$ , choose  $\varepsilon_\alpha^s > 0$  such that

$$\sum_{\alpha \in \mathcal{F}} \sum_{s \in \mathbb{Q}_0} \varepsilon_\alpha^s = \varepsilon.$$

By (9.6.2) and (9.6.3), we can choose  $r_\alpha^s < s$  such that

$$\theta(u)((\alpha^{r_\alpha^s} \wedge \neg \alpha^s)) \leq \theta(u)((\alpha^{r_\alpha^s})) - \theta(u)((\alpha^s)) \leq \varepsilon_\alpha^s.$$

The assumption  $\bigcap_i (b_i) = \emptyset$  implies that  $\bigcap_i (b_i)^*$  contains only bad ultrafilters.

The set of good ultrafilters is

$$\bigcap_{\alpha \in \mathcal{F}} \bigcap_{s \in \mathbb{Q}_0} \left( \bigcup_{r < s} (\neg \alpha^r)^* \cup (\alpha^s)^* \right). \quad (9.6.4)$$

Thus  $\bigcap_i (b_i) = \emptyset$  is equivalent to the condition

$$\left( \bigcap_{\alpha \in \mathcal{F}} \bigcap_{s \in \mathbb{Q}_0} \left( \bigcup_{r < s} (\neg \alpha^r)^* \cup (\alpha^s)^* \right) \right) \cap \bigcap_i (b_i)^* = \emptyset.$$

From this it follows that

$$\left( \bigcap_{\alpha \in \mathcal{F}} \bigcap_{s \in \mathbb{Q}_0} \left( (\neg \alpha^{r_\alpha^s})^* \cup (\alpha^s)^* \right) \right) \cap \bigcap_i (b_i)^* = \emptyset.$$

Since the space of ultrafilters is compact in the presence of the bad ultrafilters and  $(a)^*$  is a clopen for any  $a \in A$ , there exist finite sets  $C_0 \subseteq \mathcal{F}$  and  $S_0 \subseteq \mathbb{Q} \cap [0, 1]$  and  $j \in \mathbb{N}$  such that

$$\bigcap_{\alpha \in C_0} \bigcap_{s \in S_0} (\neg \alpha^{r_\alpha^s} \vee \alpha^s)^* \cap (b_j)^* = \emptyset,$$

or in other words,

$$\begin{aligned} (b_j)^* &\subseteq \bigcup_{\alpha \in C_0} \bigcup_{s \in S_0} (\alpha^{r_\alpha^s} \wedge \neg \alpha^s)^* \\ &= (\bigvee_{\alpha \in C_0} \bigvee_{s \in S_0} (\alpha^{r_\alpha^s} \wedge \neg \alpha^s))^* \end{aligned}$$

Thus in the Boolean algebra  $\mathbb{A}$ ,

$$b_j \leq \bigvee_{\alpha \in C_0} \bigvee_{s \in S_0} (\alpha^{r_\alpha^s} \wedge \neg \alpha^s). \quad (9.6.5)$$

Consequently,

$$\begin{aligned} \theta(u)((b_j)) &\leq \theta(u)((\bigvee_{\alpha \in C_0} \bigvee_{s \in S_0} (\alpha^{r_\alpha^s} \wedge \neg \alpha^s))) \\ &\leq \sum_{\alpha \in C_0} \sum_{s \in S_0} \theta(u)((\alpha^{r_\alpha^s} \wedge \neg \alpha^s)) \\ &\leq \sum_{\alpha \in C_0} \sum_{s \in S_0} \varepsilon_\alpha^s \leq \varepsilon. \end{aligned}$$

As  $\varepsilon > 0$  was arbitrary,  $\inf_i \theta(u)((b_i)) = 0$ . □

Since  $(\mathbb{A})$  is a field, the previous results and Theorem 3.2.3 imply that for all  $u \in \mathcal{U}$ , the set function  $\theta(u)$  can be uniquely extended to a measure on the  $\sigma$ -algebra  $\Sigma$  generated by  $(\mathbb{A})$ .

Now we are ready to prove that  $\mathbb{M}(\mathbb{A})$  is a Stone Markov process.

**Theorem 9.6.8.** *If  $\mathbb{A}$  is a countable Aumann algebra, then*

$$\mathbb{M}(\mathbb{A}) = (\mathcal{U}, (\mathbb{A}), \theta)$$

*is a Stone Markov process.*

*Proof.* We firstly prove that the space of good ultrafilters is analytic.

Since any second-countable Stone space is Polish, the set of all ultrafilters (good and bad) is Polish.

The good ultrafilters form a Borel set in the space of all ultrafilters—in fact, a  $G_{\sigma\delta}$  Borel set as given by (9.6.4)—and since any Borel set in a Polish space is analytic, we obtain that the space of good ultrafilters is analytic.

The space is saturated, since all possible good ultrafilters are present, and the set  $\{(a) \mid u \in (a)\}$  is just  $u$ .

To conclude that  $\mathbb{M}(\mathbb{A})$  is a Markov process, it remains to verify that  $\theta$  is a measurable function.

Let  $a \in A$ ,  $r \in \mathbb{R} \cap [0, 1]$ , and  $(r_i)_i \subseteq \mathbb{Q} \cap [0, 1]$  an increasing sequence with supremum  $r$ .

Let  $X = \{\mu \in \Delta(\mathcal{U}, \Sigma) \mid \mu((a)) \geq r\}$ .

It suffices to prove that  $\theta^{-1}(X) \in \Sigma$ . But

$$\begin{aligned} \theta^{-1}(X) &= \{u \in \mathcal{U} \mid \theta(u)((a)) \geq r\} \\ &= \bigcap_i \{u \in \mathcal{U} \mid \theta(u)((a)) \geq r_i\} \\ &= \bigcap_i (F_{r_i} a) \in \Sigma. \end{aligned}$$

□

Now we are ready to prove the algebraic version of a truth lemma for Aumann algebras.

**Lemma 9.6.9** (Extended Truth Lemma). *Let  $\mathbb{A}$  be a countable Aumann algebra and  $[\cdot]$  an interpretation of elements of  $\mathbb{A}$  as measurable sets in  $\mathcal{M}$  such that for any generator  $p$  of  $\mathbb{A}$ ,  $[[p]] = \{u \in \mathcal{U} \mid p \in u\}$ . Then, for arbitrary  $a \in A$ ,*

$$[[a]] = (a).$$

*Proof.* We prove it by induction on the structure of the element  $a$ .

The case  $a = p$ , where  $p$  is a generator of  $\mathbb{A}$ :

$$\llbracket a \rrbracket = \{u \in \mathcal{U} \mid u \ni p\} = \langle p \rangle.$$

The Boolean cases are trivial.

The case  $a = F_r b$ :

$$\begin{aligned} \llbracket F_r b \rrbracket &= \llbracket F_r \rrbracket(\llbracket b \rrbracket) = \{u \in \mathcal{U} \mid \theta(u)(\llbracket b \rrbracket) \geq r\} = \\ &= \{u \in \mathcal{U} \mid \theta(u)(\langle b \rangle) \geq r\} = \langle F_r b \rangle. \end{aligned}$$

□

### 9.6.3 From SMPs to AAs

Let  $\mathcal{M} = (M, \mathcal{B}, \theta)$  be a Stone Markov process with distinguished base  $\mathcal{B}$ . By definition,  $\mathcal{B}$  is a field of clopen sets closed under the operations

$$F_r(A) = \{m \in M \mid \theta(m)(A) \geq r\}.$$

**Theorem 9.6.10.** *The structure  $\mathcal{B}$  with the set-theoretic Boolean operations and the operations  $F_r$  for  $r \in \mathbb{Q} \cap [0, 1]$ , is a countable Aumann algebra.*

We denote this algebra by  $\mathbb{A}(\mathcal{M})$ .

*Proof.* We need to verify all the axioms of Aumann algebra.

(AA1) has the form  $M \subseteq F_0 c$  for arbitrary  $c \in \mathcal{B}$ . Trivially true, since

$$F_0 c = \{m \in M \mid \theta(m)(c) \geq 0\} = M.$$

(AA2) has the form  $M \subseteq F_r M$  that is trivially true, since  $\theta(m)(M) = 1$  implies

$$F_r M = \{m \in M \mid \theta(m)(M) \geq r\} = M.$$

(AA3) has the form  $F_r c \subseteq M \setminus F_s(M \setminus c)$  for arbitrary  $c \in \mathcal{B}$  and  $r + s > 1$ . This is equivalent to  $F_r c \cap F_s(M \setminus c) = \emptyset$  for  $s + t > 1$ .

Note that

$$F_s(M \setminus c) = \{m \in M \mid \theta(m)(M \setminus c) \geq s\} = \{m \in M \mid \theta(m)(c) \leq 1 - s\}.$$

If there exists  $m \in F_r c \cap F_s(M \setminus c)$ , then using the previous observation we should have  $r \leq \theta(m)(c) \leq 1 - s$ , implying  $r \leq 1 - s$ , but this contradicts the fact that  $r + s > 1$ .

(AA4) has the form

$$F_r(c \cap c') \cap F_s(c \cap (M \setminus c')) \subseteq F_{r+s}c.$$

Observe that

$$F_r(c \cap c') \cap F_s(c \cap (M \setminus c')) = \{m \in M \mid \theta(m)(c \cap c') \geq r \text{ and } \theta(m)(c \cap (M \setminus c')) \geq s\}.$$

Using additivity we obtain further that

$$F_r(c \cap c') \cap F_s(c \cap (M \setminus c')) = \{m \in M \mid \theta(m)(c) \geq r + s\} = F_{r+s}c.$$

(AA5) has the form

$$(M \setminus F_r(c \cap c')) \cap (M \setminus F_s(c \cap (M \setminus c'))) \subseteq M \setminus F_{r+s}c$$

and it can be proved similarly to the case of (AA4).

(AA6) has the form

$$c \subseteq c' \text{ implies } F_r c \subseteq F_r c'.$$

This follows from the monotonicity of  $\theta(m)$ .

(AA7) has the form

$$\bigcap_{r < s} F_{r_1, \dots, r_n, r} c = F_{r_1, \dots, r_n, s} c.$$

The proof is done inductively on  $n$ .

For  $n = 0$  we need to prove that  $\bigcap_{r < s} F_r c = F_s c$ . Observe that

$$\begin{aligned} \bigcap_{r < s} F_r c &= \{m \in M \mid \theta(m)(c) \leq r \text{ for all } r > s\} = \\ &= \{m \in M \mid \theta(m)(c) \leq s\} = F_s c. \end{aligned}$$

The inductive step uses (AA6). □

### 9.6.4 Duality

In this section we summarize the previous results in the form of the duality theorem.

**Theorem 9.6.11** (Duality Theorem).

(i) Any countable Aumann algebra  $\mathbb{A}$  is isomorphic to  $\mathbb{A}(\mathbb{M}(\mathbb{A}))$  via the map

$$\beta : \mathbb{A} \rightarrow \mathbb{A}(\mathbb{M}(\mathbb{A}))$$

defined for arbitrary  $a \in \mathbb{A}$  by

$$\beta(a) = \{u \in \text{supp}(\mathbb{M}(\mathbb{A})) \mid a \in u\} = \langle a \rangle.$$

(ii) Any Stone Markov process  $\mathcal{M} = (M, \mathbb{A}, \theta)$  is homeomorphic to  $\mathbb{M}(\mathbb{A}(\mathcal{M}))$  via the map

$$\alpha : \mathcal{M} \rightarrow \mathbb{M}(\mathbb{A}(\mathcal{M}))$$

defined for arbitrary  $m \in \mathcal{M}$  by

$$\alpha(m) = \{a \in \mathbb{A}(\mathcal{M}) \mid m \in a\}.$$

*Proof.* (i) The set  $\beta(a)$  is the set of good ultrafilters of  $\mathbb{A}$  that contain  $a$ ; that is,  $\beta(a) = \langle a \rangle$ .

By the classical Stone representation theorem,  $\mathbb{A}$  and  $\langle \mathbb{A} \rangle^*$  are isomorphic as Boolean algebras via the map  $a \mapsto \langle a \rangle^*$ .

By the Rasiowa–Sikorski lemma (Theorem 3.6.1) and Lemma 9.6.2, the good ultrafilters are dense in  $\langle \mathbb{A} \rangle^*$ , and since the space  $\mathcal{U}^*$  has a base of clopens,  $\langle \mathbb{A} \rangle^*$  and  $\langle \mathbb{A} \rangle$  are isomorphic as Boolean algebras via the map  $\langle a \rangle^* \mapsto \langle a \rangle$ .

Indeed, consider two elements  $a, b$  of  $\mathbb{A}$  such that  $\langle a \rangle^* \subseteq \langle b \rangle^*$ . This implies  $\langle a \rangle \subseteq \langle b \rangle$  by definition.

Reverse, if  $\langle a \rangle \subseteq \langle b \rangle$  and  $\langle a \rangle^* \not\subseteq \langle b \rangle^*$ , then  $\langle a \rangle^* \setminus \langle b \rangle^*$  is a non-empty clopen. By density,  $(\langle a \rangle^* \setminus \langle b \rangle^*) \cap \mathcal{U} \neq \emptyset$  implying  $\langle a \rangle \setminus \langle b \rangle \neq \emptyset$  - contradiction.

It remains to show that the operations  $F_r$  are preserved.

Let  $\mathcal{U} = \text{supp}(\mathbb{M}(\mathbb{A}))$ . For each  $r \in \mathbb{Q} \cap [0, 1]$ ,

$$\begin{aligned} \beta(F_r a) &= \{u \in \mathcal{U} \mid F_r a \in u\} \\ &= \{u \in \mathcal{U} \mid \theta(u)(\langle a \rangle) \geq r\} \\ &= \{u \in \mathcal{U} \mid \theta(u)(\beta(a)) \geq r\} \\ &= F_r(\beta(a)). \end{aligned}$$

(ii) The set  $\alpha(m)$  is the set of all elements of  $\mathbb{A}$  that contain  $m$ .

We first prove that this is a good ultrafilter of  $\mathbb{A}(\mathcal{M})$ . It is clearly an ultrafilter, as it is a principal ultrafilter of a set-theoretic Boolean algebra.

To show that it is good, we need to reason that if  $a \in \mathbb{A}$  and  $F_r a \in \alpha(m)$  for all  $r < s$ , then  $F_s a \in \alpha(m)$ . This follows immediately from the fact that

$$F_t a \in \alpha(m) \text{ iff } m \in F_t a \text{ iff } \theta(m)(a) \geq t.$$

The map  $\alpha$  is a strict embedding, since the two distinguished bases  $\mathbb{A}$  of  $\mathcal{M}$  and  $\langle \mathbb{A} \rangle$  of  $\mathbb{M}(\mathbb{A}(\mathcal{M}))$  are isomorphic.

This embedding must be a homeomorphism, since  $\mathcal{M}$  is saturated.  $\square$

### 9.6.5 Duality in Categorical Form

We present the previous results in a more categorical format. The categories of Aumann algebras (**AA**) and Stone Markov processes (**SMP**) were defined in §9.4 and §9.5, respectively.

We define two contravariant functors

$$\mathbb{A} : \mathbf{SMP} \rightarrow \mathbf{AA}^{\text{op}}$$

and

$$\mathbb{M} : \mathbf{AA} \rightarrow \mathbf{SMP}^{\text{op}}$$

The functor  $\mathbb{A}$  on an object  $\mathcal{M}$  produces the Aumann algebra  $\mathbb{A}(\mathcal{M})$  defined in Theorem 9.6.10.

On arrows  $f : \mathcal{M} \rightarrow \mathcal{N}$  we define

$$\mathbb{A}(f) = f^{-1} : \mathbb{A}(\mathcal{N}) \rightarrow \mathbb{A}(\mathcal{M})$$

It is well known that this is a Boolean algebra homomorphism. It is also easy to verify from the definition of morphisms in the category  $\mathbf{SMP}$  (Definition 9.5.1) that it is an Aumann algebra homomorphism.

To see this explicitly, let  $A \in \mathbb{A}_{\mathcal{N}}$ . We wish to show that

$$f^{-1}(F_r^{\mathcal{N}}(A)) = F_r^{\mathcal{M}}(f^{-1}(A)).$$

Using the fact that

$$\theta_{\mathcal{N}}(f(m))(A) = \theta_{\mathcal{M}}(m)(f^{-1}(A)),$$

we have

$$\begin{aligned} m \in f^{-1}(F_r^{\mathcal{N}}(A)) &\Leftrightarrow f(m) \in F_r^{\mathcal{N}}(A) \\ &\Leftrightarrow \theta_{\mathcal{N}}(f(m))(A) \geq r \\ &\Leftrightarrow \theta_{\mathcal{M}}(m)(f^{-1}(A)) \geq r \\ &\Leftrightarrow m \in F_r^{\mathcal{M}}(f^{-1}(A)). \end{aligned}$$

The functor  $\mathbb{M} : \mathbf{AA} \rightarrow \mathbf{SMP}^{\text{op}}$  on an object  $\mathbb{A}$  gives the Stone–Markov process  $\mathbb{M}(\mathbb{A})$  defined in Theorem 9.6.8.

On morphisms  $h : \mathbb{A} \rightarrow \mathbb{B}$ , it maps ultrafilters to ultrafilters by

$$\mathbb{M}(h) = h^{-1} : \mathbb{M}(\mathbb{B}) \rightarrow \mathbb{M}(\mathbb{A});$$

that is,

$$\mathbb{M}(h)(u) = h^{-1}(u) = \{A \in \mathbb{A}_{\mathcal{N}} \mid h(A) \in u\}.$$

Another way to view  $\mathbb{M}(h)$  is by composition, recalling that an ultrafilter can be identified with a homomorphism  $\bar{u} : \mathbb{A} \rightarrow \mathbb{2}$  by  $u = \{a \mid \bar{u}(a) = 1\}$ . In this view,

$$\mathbb{M}(h)(\bar{u}) = \bar{u} \circ h,$$

where  $\circ$  denotes function composition.

We know from classical Stone duality that this is continuous. We need to verify that it is a morphism.

It suffices to verify it on sets of the form  $\langle a \rangle$  as these generate the  $\sigma$ -algebra. Because  $h$  is a homomorphism, we calculate as follows:

$$\begin{aligned} \theta_{\mathcal{B}}(u)(\mathbb{M}(h)^{-1}(\langle a \rangle)) &= \sup\{r \mid F_r(h(a)) \in u\} \\ &= \sup\{r \mid h(F_r(a)) \in u\} \\ &= \{r \mid \bar{u}(h(F_r(a))) = 1\} \\ &= \{r \mid F_r(a) \in \mathbb{M}(h)(u)\} \\ &= \theta_{\mathcal{A}}(\mathbb{M}(h)(u)(\langle a \rangle)). \end{aligned}$$

**Theorem 9.6.12.** *The functors  $\mathbb{M}$  and  $\mathbb{A}$  define a dual equivalence of categories.*

$$\begin{array}{ccc} & \mathbb{A} & \\ \text{SMP} & \xrightarrow{\quad} & \mathbb{A}\mathbb{A}^{\text{op}} \\ & \xleftarrow{\quad} & \\ & \mathbb{M} & \end{array}$$

*Proof.* We need to show that we have a contravariant adjunction  $\mathbb{A} \dashv \mathbb{M}$  and that the unit and counit of the adjunction are isomorphisms.

$$\begin{array}{ccc} \mathbb{A}\mathbb{A}: \mathbb{A}(\mathcal{M}) & \longrightarrow & \mathbb{A} \\ \uparrow \mathbb{A} & & \downarrow \mathbb{M} \\ \text{SMP}: \mathcal{M} & \longleftarrow & \mathbb{M}(\mathbb{A}) \end{array}$$

The unit and counit are the natural transformations

$$\begin{aligned} \alpha' : \mathbb{M} \circ \mathbb{A} &\rightarrow \mathbb{I} \\ \beta' : \mathbb{A} \circ \mathbb{M} &\rightarrow \mathbb{I} \end{aligned}$$

respectively, with components

$$\begin{aligned} \alpha'_{\mathcal{M}} : \mathbb{M}(\mathbb{A}(\mathcal{M})) &\rightarrow \mathcal{M} \\ \beta'_{\mathbb{A}} : \mathbb{A}(\mathbb{M}(\mathbb{A})) &\rightarrow \mathbb{A}, \end{aligned}$$

the inverses of the  $\alpha$  and  $\beta$  defined in Theorem 9.6.11. Here we are using  $'$  to denote functional inverse so as not to conflict with our previous usage of  $^{-1}$  as

$$f^{-1}(X) = \{x \mid f(x) \in X\}.$$

We have already shown that these are isomorphisms.

To verify that we have an adjunction with unit  $\alpha'$  and counit  $\beta'$ , it suffices to verify the conditions

$$\begin{aligned}\beta' \mathbb{A} \circ \mathbb{A} \alpha' &= \mathbb{I} \\ \mathbb{M} \beta' \circ \alpha' \mathbb{M} &= \mathbb{I}.\end{aligned}$$

Since  $\alpha'$  and  $\beta'$  are invertible, these are equivalent to

$$\begin{aligned}\mathbb{A} \alpha' &= \beta \mathbb{A} \\ \mathbb{M} \beta' &= \alpha \mathbb{M}.\end{aligned}$$

Specializing at components  $\mathbb{A}$  and  $\mathcal{M}$  and using the definition of  $\mathbb{A}$  and  $\mathbb{M}$  on morphisms, it suffices to show

$$\begin{aligned}\alpha'_{\mathcal{M}}^{-1} &= \beta_{\mathbb{A}(\mathcal{M})} \\ \beta'_{\mathbb{A}}^{-1} &= \alpha_{\mathbb{M}(\mathbb{A})}.\end{aligned}$$

For the left-hand equation, as  $\alpha_{\mathcal{M}}$  is invertible, all elements of  $\mathbb{M}(\mathbb{A}(\mathcal{M}))$  are of the form  $\alpha_{\mathcal{M}}(m)$  for some  $m \in \mathcal{M}$ , and  $\alpha'_{\mathcal{M}}(\alpha_{\mathcal{M}}(m)) = m$ .

Thus for  $C \in \mathbb{A}(\mathbb{M}(\mathbb{A}(\mathcal{M})))$ ,

$$\begin{aligned}\alpha'_{\mathcal{M}}^{-1}(C) &= \{\alpha_{\mathcal{M}}(m) \mid \alpha'_{\mathcal{M}}(\alpha_{\mathcal{M}}(m)) \in C\} \\ &= \{\alpha_{\mathcal{M}}(m) \mid m \in C\} \\ &= \{\alpha_{\mathcal{M}}(m) \mid C \in \alpha_{\mathcal{M}}(m)\} \\ &= \{\alpha_{\mathcal{M}}(m) \mid \alpha_{\mathcal{M}}(m) \in \beta_{\mathbb{A}(\mathcal{M})}(C)\} \\ &= \beta_{\mathbb{A}(\mathcal{M})}(C).\end{aligned}$$

The argument for the right-hand equation is symmetric. □

## 9.7 Concluding Remarks

As promised, we have proved a duality theorem between Stone–Markov processes and Aumann algebras which subsumes and extends the completeness theorems that we have presented in the previous Part of the monograph. Our treatment improves on the existing axiomatizations.

The following novel features appear in our proof:

1. We must remove ultrafilters that fail to satisfy a key infinitary axiom, and we must show that this does not change the represented algebra by showing that these are “rare” in a topological sense (meager).

2. As a result, the usual compactness for Stone spaces fails, and we need a new concept, which we call saturation, instead.
3. We must establish the relevant measure theoretic properties of the Markov kernels that we construct from the algebras. This again uses the Baire category theorem in a crucial way.
4. We define our Markov processes with a distinguished base and use this to constrain the morphisms in order to achieve duality.

There are many variations one can imagine exploring. Perhaps the most interesting one is to consider more general measure spaces and work with different bisimulation notions [10] that apply more generally. Our treatment is not fully localic, and perhaps some of the topological subtleties of the present proof would disappear once we adopted a more localic point of view.

# Chapter 10

## Metrized Stone Duality for Markov Processes

### 10.1 Introduction

In the previous Chapter of this monograph we have presented a Stone-type duality developed for Markov processes defined on continuous state spaces, following [27]. The algebraic counterpart of the Markov processes were called Aumann algebras in honour of Aumann's work on probabilistic reasoning [3].

Aumann algebras capture, in algebraic form, the probabilistic Markovian logic, studied in Chapter 7 of this monograph. In this logic, bounds on probabilities enter into the modalities. As demonstrated in Chapter 7, this logic can be stripped down to a very spartan core – just the modalities and finite conjunction – and still characterize bisimulation for labelled Markov processes (see also [8, 11, 12]). However, to obtain the strong completeness properties that are implied by the duality theorems, one needs infinitary proof principles [21], since the Markovian logics are not compact.

One of the critiques [19] of logics and equivalences used for the treatment of probabilistic systems is that boolean logic is not robust with respect to small perturbations of the real-valued system parameters. Accordingly, a theory of metrics [13, 14, 39, 40] was developed and metric reasoning principles were advocated. In the present paper we extend our exploration of duality theory with an investigation into the role of metrics and exhibit a metric analogue of the duality theory. This opens the way for an investigation into quantitative aspects of *approximate reasoning*.

In the present Chapter we integrate quantitative information into the duality presented in the previous Chapter, by endowing Markov processes with a (pseudo)metric and Aumann algebras with a quantitative “norm-like” structure called a *metric diameter*. The interplay between the pseudometric and the boolean algebra is somewhat delicate and had to be carefully examined for the duality to emerge.

The final results, proven in [27], have easy proofs but the correct way to impose quantitative structure on Aumann algebras was elusive. The key idea is to axiomatize the notion of metric diameter on the Aumann algebra side. This is a concept more like a norm than a distance, but one can derive a metric from it. The idea comes from a paper by Banaschewski and Pultr [4] on Stone duality for metric spaces. However, our formulation is not quite the same as theirs.

## 10.2 Extending the Duality to Metrized Markov Processes

We add quantitative structure to both Markov processes and Aumann algebras.

We prove an extended version of the representation theorem for metrized Markov processes versus metrized Aumann algebras. This theorem states that starting from an arbitrary metrized Markov process, we can extend the Aumann algebra constructed in the previous Chapter to a metrized Aumann algebra that preserves the pseudometric and conversely. In other words the natural isomorphisms that arise in the duality presented in the previous Chapter will turn out to be isometries.

Moreover, the condition that defines the concept of metrized Markov process seems to be a further refinement of the concept of dynamic-continuous pseudometrics that we researched for in Chapter 8. In fact, the concept of dynamic-continuity, even if it succeeded to guarantee that certain sequences of MPs converge, it did not succeed to rule out undesirable pseudometrics such as the discrete one that measures to 0 the distance between bisimilar MPs and to 1 all the others. However, the condition in the definition of metrized MPs that we will present in what follows succeeds to complete the job.

## 10.3 Metrized Markov Processes

We equip Stone Markov processes with a pseudometric that measures distances between the states of the MP. The key condition that we impose is that for a particular state  $m$ , the diameters of the clopens containing  $m$  converges to 0.

**Definition 10.3.1** (Metrized Markov process). *A metrized Markov process is a tuple  $(\mathcal{M}, d)$ , where  $\mathcal{M} = (M, \mathbb{A}, \theta)$  is a Stone Markov process with  $\mathbb{A}$  its countable base of clopens and  $d : M \times M \rightarrow [0, 1]$  is a pseudometric on  $M$  satisfying for arbitrary  $m \in M$  the property*

$$(M) \quad \inf_{c \in \mathbb{A}, m \in c} \sup\{d(n, n') \mid n, n' \in c\} = 0.$$

The following lemma gives a number of conditions equivalent to (M). In particular, it shows the connection between the topology of the Stone Markov space and the pseudometric topology.

**Lemma 10.3.2.** *For a metrized MP  $(\mathcal{M}, d)$ , where  $\mathcal{M} = (M, \mathbb{A}, \theta)$ , the following are equivalent:*

- (i)  $\forall m, \inf_{c \in \mathbb{A}, m \in c} \sup\{d(n, n') \mid n, n' \in c\} = 0$
- (ii)  $\forall m, m' \inf_{c \in \mathbb{A}, m, m' \in c} \sup\{d(n, n') \mid n, n' \in c\} = d(m, m')$
- (iii)  $\forall m, \forall \varepsilon > 0 \exists c \in \mathbb{A} (m \in c \wedge \forall n, n' \in c d(n, n') < \varepsilon)$
- (iv)  $\forall m \forall \varepsilon > 0 \exists c \in \mathbb{A} (m \in c \wedge \forall n \in c d(m, n) < \varepsilon)$
- (v) *The topology generated by  $\mathbb{A}$  refines the pseudometric topology generated by  $d$ .*
- (vi) *The pseudometric  $d$  is continuous in both arguments with respect to the  $\mathbb{A}$ -topology.*

*Proof.* Note that (i) is just (M).

(i)  $\Leftrightarrow$  (iii) is immediate from the definitions.

For (iii)  $\Rightarrow$  (iv), we can substitute  $m, n$  for  $n, n'$  in (iii).

For (iv)  $\Rightarrow$  (iii), let  $m$  and  $\varepsilon > 0$  be arbitrary, and let  $c \in \mathbb{A}$  such that  $m \in c$  and for all  $n \in c$ ,  $d(m, n) < \varepsilon/2$  and  $d(n, m) < \varepsilon/2$ . Then for any  $n, n' \in c$ ,

$$d(n, n') \leq d(n, m) + d(m, n') < \varepsilon/2 + \varepsilon/2 = \varepsilon.$$

For (iv)  $\Leftrightarrow$  (v), let  $N_\varepsilon(m) = \{x \mid d(m, x) < \varepsilon\}$ . Then,

$$\begin{aligned} & \forall m \forall \varepsilon > 0 \exists c \in \mathbb{A} (m \in c \wedge \forall n \in c d(m, n) < \varepsilon \wedge d(n, m) < \varepsilon) \\ & \Leftrightarrow \forall m \forall \varepsilon > 0 \exists c \in \mathbb{A} (m \in c \wedge \forall n \in c n \in N_\varepsilon(m)) \\ & \Leftrightarrow \forall m \forall \varepsilon > 0 \exists c \in \mathbb{A} (m \in c \wedge c \subseteq N_\varepsilon(m)). \end{aligned}$$

The last statement says that every basic open neighborhood of the pseudometric topology contains a basic open neighborhood of the  $\mathbb{A}$ -topology, which says exactly that the  $\mathbb{A}$ -topology refines the pseudometric topology.

For (iii)  $\Rightarrow$  (vi), to show that  $d$  is continuous in its second argument, let  $m, x$  and  $\varepsilon > 0$  be arbitrary and let  $c \in \mathbb{A}$  such that  $x \in c$  and for all  $n, n' \in c$ ,  $d(n, n') < \varepsilon$ . Then for all  $y \in c$ ,

$$\begin{aligned} d(m, y) & \leq d(m, x) + d(x, y) < d(m, x) + \varepsilon \\ d(m, x) & \leq d(m, y) + d(y, x) < d(m, y) + \varepsilon \end{aligned}$$

so  $d(m, y) \in (d(m, x) - \varepsilon, d(m, x) + \varepsilon)$ .

That  $d$  is continuous in its first argument follows from symmetry.

For the other direction, suppose  $d$  is continuous in its second argument. Then for all  $m$  and  $\varepsilon > 0$ , the set  $N_\varepsilon(m)$  is open and contains  $m$ , thus there exists a basic open set  $c \in \mathbb{A}$  such that  $m \in c$  and  $c \subseteq N_\varepsilon(m)$ . Thus the  $\mathbb{A}$ -topology refines the pseudometric topology of  $d$ .

Statement (ii) implies (i) immediately by taking  $m' = m$  in (ii).

To show (i) implies (ii), let  $m, m'$  and  $\varepsilon > 0$  be arbitrary.

From (iv), we have  $c \in \mathbb{A}$  such that  $m \in c$  and from (iii) we have that for all  $n, n' \in c$ ,  $d(n, n') < \varepsilon/2$  and we have  $c' \in \mathbb{A}$  such that  $m' \in c'$  and for all  $n, n' \in c'$ ,  $d(n, n') < \varepsilon/2$ . We claim that for all  $n, n' \in c \cup c'$ ,

$$d(n, n') < d(m, m') + \varepsilon,$$

which will establish (ii).

If  $n, n' \in c$ , then

$$d(n, n') \leq d(n, m) + d(m, n') < \varepsilon/2 + \varepsilon/2 = \varepsilon \leq d(m, m') + \varepsilon.$$

If  $n, n' \in c'$ , the argument is the same, replacing  $m$  by  $m'$ . If  $n \in c$  and  $n' \in c'$ , then

$$\begin{aligned} d(n', n) = d(n, n') &\leq d(n, m) + d(m, m') + d(m', n') \\ &< \varepsilon/2 + d(m, m') + \varepsilon/2 = d(m, m') + \varepsilon. \end{aligned}$$

□

Having a concept of pseudometric at hand, it make sense to define what does it mean for two metrized MPs to be isometric.

**Definition 10.3.3** (Isometric Markov processes). *Given two metrized MPs  $(\mathcal{M}_i, d_i)$ , where  $\mathcal{M}_i = (M_i, \Sigma_i, \theta_i)$  for  $i = 1, 2$ , an isometry from  $\mathcal{M}_1$  to  $\mathcal{M}_2$  is a map  $f : M_1 \rightarrow M_2$  such that for arbitrary  $m, n \in M_1$ ,*

$$d_1(m, n) = d_2(f(m), f(n)).$$

## 10.4 Metrized Aumann Algebras

Now we introduce the metrized Aumann algebras. Despite their name, the metrized AAs are not directly equipped with a pseudometric structure, but with a concept of metric diameter. Later we will prove that the metric diameter can indeed define a pseudometric.

**Definition 10.4.1** (Metrized Aumann algebra). *A metrized Aumann algebra is a tuple  $(\mathbb{A}, |\cdot|)$ , where  $\mathbb{A} = (A, \rightarrow, \perp, \{F_r\}_{r \in \mathbb{Q}_0}, \leq)$  is an Aumann algebra and  $|\cdot| : A \rightarrow [0, 1]$  is a metric diameter on  $A$ , which is a map satisfying, for arbitrary  $a, b \in A$  and ultrafilter  $u$ , the following properties*

- (A1)  $|\perp| = 0$ ;
- (A2) if  $a \leq b$ , then  $|a| \leq |b|$ ;

- (A3) if  $a \wedge b \neq \perp$ , then  $|a \vee b| \leq |a| + |b|$ ;
- (A4)  $\inf\{|a| \mid a \in u\} = 0$ .

Similarly to the case of metrized MPs, the morphisms of Aumann algebras can be extended with a notion of isometry.

**Definition 10.4.2** (Isometric Aumann Algebras). *Given two metrizable Aumann Algebras  $(\mathbb{A}_i, | \cdot |_i)$  for  $i = 1, 2$ , an isometry from  $\mathbb{A}_1$  to  $\mathbb{A}_2$  is a map  $f : A_1 \rightarrow A_2$  such that for any  $a \in A_1$ ,*

$$|a|_1 = |f(a)|_2.$$

## 10.5 Metrized Stone Duality

With all these concepts in hands, we are ready to extend the Stone duality for MPs presented in the previous Chapter, to include metric structures and isometries.

Consider a metrizable MP  $(\mathcal{M}, d)$ , where  $\mathcal{M} = (M, \mathbb{A}, \theta)$ . As in the previous Chapter, let  $\mathbb{A}(\mathcal{M})$  be the AA constructed from  $\mathcal{M}$ .

We extend this construction so that  $\mathbb{A}(\mathcal{M})$  becomes a metrized AA.

For arbitrary  $a \in \mathbb{A}$ , let

$$|a|_d = \sup\{d(m, n) \mid m, n \in a\},$$

under the assumption that  $\sup \emptyset = 0$ .

**Lemma 10.5.1.**  *$(\mathbb{A}(\mathcal{M}), | \cdot |_d)$  is a metrized Aumann Algebra.*

*Proof.* We need to verify the axioms (A1) – (A4) in Definition 10.4.1.

**(A1)**  $|\perp|_d = 0$  follows from the assumption that  $\sup \emptyset = 0$ .

**(A2)** If  $c_1 \subseteq c_2$ , then  $\sup\{d(m, n) \mid m, n \in c_1\} \leq \sup\{d(m, n) \mid m, n \in c_2\}$ .

**(A3)** Suppose that  $c_1 \cap c_2 \neq \emptyset$ . Let  $z \in c_1 \cap c_2$ ,  $x \in c_1$  and  $y \in c_2$ . Then, the triangle inequality for  $d$  guarantees that

$$\begin{aligned} d(x, y) &\leq d(x, z) + d(z, y) \leq \\ \sup\{d(x, z) \mid x, z \in c_1\} &+ \sup\{d(z, y) \mid z, y \in c_2\} = \\ &|c_1| + |c_2|. \end{aligned}$$

Consequently,

$$\sup\{d(x, y) \mid x \in c_1, y \in c_2\} \leq |c_1| + |c_2|$$

and similarly,

$$\sup\{d(y, x) \mid x \in c_1, y \in c_2\} \leq |c_1| + |c_2|.$$

Since

$$\begin{aligned} |c_1 \cup c_2| &= \\ \max\{|c_1|, |c_2|, \sup\{d(x, y) \mid x \in c_1, y \in c_2\}, \sup\{d(x, y) \mid x \in c_2, y \in c_1\}\}, \\ |c_1 \cup c_2| &\leq |c_1| + |c_2|. \end{aligned}$$

**(A4)** Using the notation of the previous duality theorem 9.6.11 proven in the previous Chapter, we know that for any ultrafilter  $u$  of  $\mathbb{A}(\mathcal{M})$  and any clopen  $c \in \mathbb{A}$ , we have that

$$\alpha^{-1}(u) \in c \text{ iff } c \in u.$$

Then,

$$\inf\{|c|_d \mid c \in u\} = \inf\{|c|_d \mid \alpha^{-1}(u) \in c\}.$$

Since  $\alpha$  is a bijection,  $\alpha^{-1}(u) \in \mathcal{M}$  and  $|c|_d = \sup\{d(n, n') \mid n, n' \in c\}$ , using (M) we obtain

$$\inf_{c \in \mathbb{A}, c \ni \alpha^{-1}(u)} \sup\{d(n, n') \mid n, n' \in c\} = 0,$$

therefore

$$\inf\{|c|_d \mid c \in u\} = 0.$$

□

Consider a metrized Aumann algebra  $(\mathbb{A}, |\cdot|)$  and let  $\mathbb{M}(\mathbb{A})$  be the Markov process constructed from  $\mathbb{A}$ , as presented in the previous Chapter. We extend this construction so that  $\mathbb{M}(\mathbb{A})$  will become a metrized Markov process.

For arbitrary ultrafilters  $u, v$  of  $\mathbb{A}$ , let

$$\delta_{|\cdot|}(u, v) = \inf\{|a| \mid a \in u \cap v\}.$$

**Lemma 10.5.2.**  $(\mathbb{M}(\mathbb{A}), \delta_{|\cdot|})$  is a metrized MP.

*Proof.* First, we prove that  $\delta_{|\cdot|}$  is a pseudometric over the space of ultrafilters.

From (A4) we can simply infer that  $\delta_{|\cdot|}(u, u) = 0$ , while the symmetry of  $\delta_{|\cdot|}$  follows from the definition.

We now prove the triangle inequality: let  $u, v, w$  be three arbitrary ultrafilters. Let  $a \in u \cap v$ ,  $b \in u \cap w$  and  $c \in w \cap v$ . Obviously  $b \cup c \in u \cap v$ . Then,

$$\inf_{a \in u \cap v} |a| \leq |b \cup c|.$$

Since  $b \cap c \neq \emptyset$ , using (A3) we get  $|b \cup c| \leq |b| + |c|$  which guarantees that for any  $b \in u \cap w$  and any  $c \in w \cap v$ ,

$$\inf_{a \in u \cap v} |a| \leq |b| + |c|,$$

implying

$$\inf_{a \in u \cap v} |a| \leq \inf_{b \in u \cap w} |b| + \inf_{c \in w \cap v} |c|.$$

Hence,

$$\delta_{|\cdot|}(u, v) \leq \delta_{|\cdot|}(u, w) + \delta_{|\cdot|}(w, v)$$

which proves that  $\delta_{|\cdot|}$  is a pseudometric.

It remains to verify (M).

Since from theorem 9.6.11 we know that

$$a \in u \text{ iff } u \in \beta(a),$$

(M) follows directly from (A4). □

Finally, we extend the duality theorem 9.6.11 to include the metric structure. Essentially, we show that the isomorphisms  $\alpha$  and  $\beta$  of the duality theorem are isometries.

**Theorem 10.5.3** (Metriized Duality Theorem). *1. Any metriized countable Aumann algebra  $(\mathbb{A}, |\cdot|)$  is isomorphic to  $(\mathbb{A}(\mathbb{M}(\mathbb{A})), |\cdot|_{\delta_{|\cdot|}})$  via the map*

$$\beta : \mathbb{A} \rightarrow \mathbb{A}(\mathbb{M}(\mathbb{A}))$$

*defined for arbitrary  $a \in \mathbb{A}$  by*

$$\beta(a) = \{u \in \text{supp}(\mathbb{M}(\mathbb{A})) \mid a \in u\} = \langle a \rangle.$$

*Moreover,  $\beta$  is an isometry of metriized Aumann algebras, i.e., for arbitrary  $a \in \mathbb{A}$ ,*

$$|a| = |\beta(a)|_{\delta_{|\cdot|}}.$$

*2. Any metriized Stone Markov process  $(\mathcal{M}, d)$ , where  $\mathcal{M} = (M, \mathbb{A}, \theta)$  is homeomorphic to  $(\mathbb{M}(\mathbb{A}(\mathcal{M})), \delta_{|\cdot|_d})$  via the map*

$$\alpha : \mathcal{M} \rightarrow \mathbb{M}(\mathbb{A}(\mathcal{M}))$$

*defined for arbitrary  $m \in M$  by*

$$\alpha(m) = \{A \in \mathbb{A} \mid m \in A\}.$$

*Moreover,  $\alpha$  is an isometry of metriized Markov processes, i.e., for arbitrary  $m, n \in M$ ,*

$$d(m, n) = \delta_{|\cdot|_d}(\alpha(m), \alpha(n)).$$

*Proof.* We only need to prove the two isometries, since the rest has been proven in Theorem 9.6.11.

(i). The isometry of AAs. We need to prove that  $|a| = |\beta(a)|_{\delta_{|\cdot|}}$ .

Observe that

$$|\beta(a)|_{\delta_{|\cdot|}} = \sup_{u,v \in \beta(a)} \delta_{|\cdot|}(u,v) = \sup_{u,v \in \beta(a)} \inf_{a' \in u \cap v} |a'|.$$

Since  $\beta(a)$  is the set of all ultrafilters containing  $a$ ,  $a'$  quantifies over all elements that belong to the intersection of all ultrafilters containing  $a$ . But this intersection is nothing else but the principal filter  $\uparrow a$  of  $a$ . Hence, the previous equality became

$$|\beta(a)|_{\delta_{|\cdot|}} = \inf_{a' \in \uparrow a} |a'|.$$

Now the monotonicity stated by (A2) guarantees that

$$\inf_{a' \in \uparrow a} |a'| = |a|.$$

(ii). The isometries of MPs. We need to prove that  $d(m,n) = \delta_{|\cdot|_d}(\alpha(m), \alpha(n))$ .

From Lemma 10.3.2(ii) we know that

$$d(m,n) = \inf_{c \in \mathbb{A}, c \ni m,n} |c|_d.$$

From Theorem 9.6.11 we also know that

$$m, n \in c \text{ iff } c \in \alpha(m) \cap \alpha(n).$$

Consequently,

$$d(m,n) = \inf\{|c|_d \mid c \in \alpha(m) \cap \alpha(n)\} = \delta_{|\cdot|_d}(\alpha(m), \alpha(n)).$$

□

## 10.6 A canonic metric for Metrized Aumann Algebras

We have seen in Chapter 8 how useful can be to have a pseudometric over the space of logical formulas that agrees in certain sense with the pseudometric on MPs. In this section we show that we actually can define such a pseudometric in a canonic way from a structure of metrized Aumann algebra. It is exactly this canonicity that guarantees, via the duality, that the pseudometric on algebra side agrees with the one on MPs.

We have claimed earlier that the metric diameter on an Aumann algebra induces a pseudometric. We now demonstrate this.

Let  $(\mathbb{A}, |\cdot|)$  be a metrized Aumann algebra.

For arbitrary  $a, b \in \mathbb{A}$  and  $\varepsilon > 0$ , let

$$B_\varepsilon(b) = \bigcup \{ \beta(b') \mid b' \in \mathbb{A}, |b'| \leq \varepsilon, b \wedge b' \neq \perp \}.$$

Intuitively,  $B_\varepsilon(b)$  is a ball that contains all ultrafilters that are at distance at most  $\varepsilon$  from some ultrafilter containing  $b$ . This definition allows us to define a natural distance on  $\mathbb{A}$  by

$$d_{|\cdot|}(a, b) = \inf \{ \varepsilon > 0 \mid B_\varepsilon(b) \supseteq \beta(a) \text{ and } B_\varepsilon(a) \supseteq \beta(b) \}.$$

Intuitively, if in the light of the duality we think of the elements of  $\mathbb{A}$  as sets of ultrafilters, then the previous distance is just the Hausdorff pseudometric of the distance between ultrafilters.

To prove that the previous construction is not void, we show in the next lemma that for any non-zero element  $a \in \mathbb{A}$  the ball  $B_\varepsilon(a)$  is not empty for any  $\varepsilon$ .

**Lemma 10.6.1.** *If  $a \neq \perp$ , then for any  $\varepsilon > 0$  there exists  $a' \neq \perp$  such that  $a \wedge a' \neq \perp$  and  $|a'| \leq \varepsilon$ .*

*Proof.* Since  $a \neq \perp$ , applying Rasiowa-Sikorski lemma, there exists an ultrafilter  $u$  such that  $a \in u$ .

For any other  $a' \in u$ ,  $a \wedge a' \in u$ , hence  $a \wedge a' \neq \perp$ . Moreover, using (A4) there exists  $a' \in u$  such that  $|a'| \leq \varepsilon$ .  $\square$

Now we prove that  $d_{|\cdot|}$  is indeed a pseudometric: it is the Hausdorff pseudometric of the pseudometric  $\delta_{|\cdot|}$  on ultrafilters.

**Lemma 10.6.2.** *The function  $d_{|\cdot|}$  previously defined on the support set of a metrizable Aumann Algebra is a pseudometric. Moreover,*

$$d_{|\cdot|}(a, b) = \max \left\{ \sup_{u \in \beta(a)} \inf_{v \in \beta(b)} \delta_{|\cdot|}(u, v), \sup_{u \in \beta(a)} \inf_{v \in \beta(b)} \delta_{|\cdot|}(v, u) \right\}.$$

*Proof.* That  $d_{|\cdot|}(a, a) = 0$  follows directly from (A4) and its symmetry from the definition.

To prove the triangle inequality, we need first to prove some additional results.

For arbitrary  $a, b \in \mathbb{A}$ , let

$$d_1(a, b) = \inf \{ \varepsilon > 0 \mid B_\varepsilon(b) \supseteq \beta(a) \}$$

and

$$d_2(a, b) = \inf \{ \varepsilon > 0 \mid B_\varepsilon(a) \supseteq \beta(b) \}.$$

We prove now that

$$d_1(a, b) = \sup_{u \ni a} \inf_{v \ni b} \delta_{|\cdot|}(u, v).$$

Observe that

$$B_\varepsilon(b) = \bigcup \{ \beta(b') \in \mathbb{A} \mid b \wedge b' \neq \perp \text{ and } |b'| \leq \varepsilon \} = \{ u \mid \exists v \ni b, \delta_{|\cdot|}(u, v) \leq \varepsilon \}.$$

Since  $d_1(a, b) = \inf \{ \varepsilon \mid B_\varepsilon(b) \supseteq \beta(a) \}$ , we obtain that

$$d_1(a, b) = \inf \{ \varepsilon \mid \forall u \ni a, \exists v \ni b, \delta_{|\cdot|}(u, v) \leq \varepsilon \}.$$

It is not difficult to see now that

$$\inf \{ \varepsilon \mid \forall u \ni a, \exists v \ni b, \delta_{|\cdot|}(u, v) \leq \varepsilon \} = \sup_{u \ni a} \inf_{v \ni b} \delta_{|\cdot|}(u, v).$$

Similarly,

$$d_2(a, b) = \sup_{u \ni b} \inf_{v \ni a} \delta_{|\cdot|}(u, v).$$

With these results in hand, we can proceed to prove the triangle inequality for  $d_i$ ,  $i = 1, 2$ .

Consider arbitrary elements  $a, b, c \in \mathbb{A}$ . We can focus on the case when all of them are non-zero, the other cases being trivially true.

Consider arbitrary ultrafilters  $u \ni a$ ,  $v \ni b$  and  $w \ni c$ . We have that

$$\delta_{|\cdot|}(u, v) \leq \delta_{|\cdot|}(u, w) + \delta_{|\cdot|}(w, v).$$

From it we derive the following inequalities.

$$\begin{aligned} \delta_{|\cdot|}(u, v) &\leq \delta_{|\cdot|}(u, w) + \sup_{w' \in \beta(c)} \delta_{|\cdot|}(w', v) \\ \inf_{v' \in \beta(b)} \delta_{|\cdot|}(u, v') &\leq \delta_{|\cdot|}(u, w) + \sup_{w' \in \beta(c)} \delta_{|\cdot|}(w', v) \\ \inf_{v' \in \beta(b)} \delta_{|\cdot|}(u, v') &\leq \delta_{|\cdot|}(u, w) + \sup_{w' \in \beta(c)} \inf_{v'' \in \beta(b)} \delta_{|\cdot|}(w', v'') \end{aligned}$$

This is equivalent to

$$\inf_{v' \in \beta(b)} \delta_{|\cdot|}(u, v') \leq \delta_{|\cdot|}(u, w) + d_{|\cdot|}(c, b)$$

And from it we derive the following chain of inequalities.

$$\begin{aligned} \inf_{v' \in \beta(b)} \delta_{|\cdot|}(u, v') &\leq \sup_{u' \in \beta(a)} \delta_{|\cdot|}(u', w) + d_{|\cdot|}(c, b) \\ \inf_{v' \in \beta(b)} \delta_{|\cdot|}(u, v') &\leq \inf_{w' \in \beta(c)} \sup_{u' \in \beta(a)} \delta_{|\cdot|}(u', w') + d_{|\cdot|}(c, b) \\ \sup_{u \in \beta(a)} \inf_{v' \in \beta(b)} \delta_{|\cdot|}(u, v') &\leq \inf_{w' \in \beta(c)} \sup_{u' \in \beta(a)} d(u', w') + d_{|\cdot|}(c, b) \end{aligned}$$

This last inequality is equivalent to

$$d_{|\cdot|}(a, b) \leq d_{|\cdot|}(a, c) + d_{|\cdot|}(c, b).$$

And this shows that indeed  $d_{|\cdot|}$  is a pseudometric.  $\square$

## 10.7 Metric Duality in Categorical Form

In this section, we present the previous results in a more categorical format. The categories of metrized Aumann algebras (**MAA**) and metrized Markov processes (**MMP**) are defined as follows.

The objects of **MAA** are metrized AAs and their morphisms are expansive morphisms of AAs, i.e., morphisms  $f : \mathbb{A}_1 \rightarrow \mathbb{A}_2$  of Aumann algebras such that for any  $a \in \mathbb{A}_1$ ,

$$|a|_1 \leq |f(a)|_2.$$

The objects of **MMP** are metrized MPs and their morphisms are non-expansive morphisms of MPs, i.e., morphisms  $f : \mathcal{M}_1 \rightarrow \mathcal{M}_2$  of Stone Markov processes such that for any  $m, n \in \mathcal{M}_1$ ,

$$d_1(m, n) \geq d_2(f(m), f(n)).$$

We define two contravariant functors

$$\mathbb{A} : \mathbf{MMP} \rightarrow \mathbf{MAA}^{\text{op}}$$

and

$$\mathbb{M} : \mathbf{MAA} \rightarrow \mathbf{MMP}^{\text{op}}$$

The functor  $\mathbb{A}$  on an object  $\mathcal{M}$  produces the Aumann algebra  $\mathbb{A}(\mathcal{M})$  defined in Theorem 9.6.10.

On arrows  $f : \mathcal{M} \rightarrow \mathcal{N}$  we define

$$\mathbb{A}(f) = f^{-1} : \mathbb{A}(\mathcal{N}) \rightarrow \mathbb{A}(\mathcal{M}).$$

We have previously proved that this is an Aumann algebra homomorphism. To see that it is also expansive, consider a morphism  $f : \mathcal{M} \rightarrow \mathcal{N}$  such that for any  $m, n \in \mathcal{M}$ ,

$$d_{\mathcal{M}}(m, n) \geq d_{\mathcal{N}}(f(m), f(n)).$$

Observe that for arbitrary  $a \in \mathbb{A}(\mathcal{N})$ ,

$$|a|_{\mathbb{A}(\mathcal{N})} = |a|_{d_{\mathcal{N}}} \text{ and } |f^{-1}(a)|_{\mathbb{A}(\mathcal{M})} = |f^{-1}(a)|_{d_{\mathcal{M}}}$$

and in this context the previous inequality guarantees that

$$|a|_{\mathbb{A}(\mathcal{N})} \leq |f^{-1}(a)|_{\mathbb{A}(\mathcal{M})}.$$

The functor  $\mathbb{M} : \mathbf{MAA} \rightarrow \mathbf{MMP}^{\text{op}}$  on an object  $\mathbb{A}$  gives the Stone–Markov process  $\mathbb{M}(\mathbb{A})$  defined in Theorem 9.6.8.

On morphisms  $h : \mathbb{A} \rightarrow \mathbb{B}$ , it maps ultrafilters to ultrafilters by

$$\mathbb{M}(h) = h^{-1} : \mathbb{M}(\mathbb{B}) \rightarrow \mathbb{M}(\mathbb{A});$$

that is, for an arbitrary ultrafilter  $u$ ,

$$\mathbb{M}(h)(u) = h^{-1}(u) = \{A \in \mathbb{A}_{\mathcal{N}} \mid h(A) \in u\}.$$

Another way to view  $\mathbb{M}(h)$  is by composition, recalling that an ultrafilter can be identified with a homomorphism  $\bar{u} : \mathbb{A} \rightarrow \mathbb{2}$  by  $u = \{a \mid \bar{u}(a) = 1\}$ . In this view,

$$\mathbb{M}(h)(\bar{u}) = \bar{u} \circ h,$$

where  $\circ$  denotes function composition.

We have proven in the previous Chapter that this is a morphism of SMPs.

That it is also non-expansive can be demonstrated as follows.

Let  $h : \mathbb{A} \rightarrow \mathbb{B}$  be a morphism such that for arbitrary  $a \in \mathbb{A}$ ,

$$|a|_{\mathbb{A}} \leq |f(a)|_{\mathbb{B}}.$$

Using the previous results it is not difficult to verify that for arbitrary ultrafilters  $u, v$  of  $\mathbb{B}$ ,

$$\delta_{|\mathbb{B}}(u, v) \geq \delta_{|\mathbb{A}}(h^{-1}(u), h^{-1}(v)).$$

And since

$$\delta_{|\mathbb{B}}(u, v) = \delta_{\mathbb{M}(\mathbb{B})}(u, v)$$

and

$$\delta_{|\mathbb{A}}(h^{-1}(u), h^{-1}(v)) = \delta_{\mathbb{M}(\mathbb{A})}(h^{-1}(u), h^{-1}(v)),$$

We obtain that  $\mathbb{M}(h)$  is non-expansive.

**Theorem 10.7.1.** *The functors  $\mathbb{M}$  and  $\mathbb{A}$  define a dual equivalence of categories.*

$$\begin{array}{ccc} & \mathbb{A} & \\ \text{MMP} & \xrightarrow{\quad} & \text{MAA}^{\text{op}} \\ & \xleftarrow{\quad} & \\ & \mathbb{M} & \end{array}$$

## 10.8 Concluding Remarks

In this Chapter we have extended the duality theory presented in the previous Chapter to a quantitative setting.

It is important to note that the conditions we have imposed on the pseudometric relate the support topology on which the Stone Markov process is defined, to the open-ball topology induced by the pseudometric. This can be seen from the fact that the pseudometric topology is refined by the (non-compact but saturated) Stone topology.

We have defined our Stone Markov processes to be Hausdorff spaces, which was necessary for the duality theory. In effect, this means that the clopens separate points. In other words, one cannot have two states that satisfy exactly the same formulas. In view of the logical characterization of bisimulation, this implies that no two distinct states are bisimilar. That is, the process is already minimal with respect to bisimulation.

If we look at a broader class of Markov processes, then we would have possibly nontrivial bisimulations on the space. The Stone Markov processes would be a reflective subcategory with the reflector sending each Markov process to a version of the process with all the bisimulation equivalence classes collapsed to point; this would be a Stone Markov process.

How does the topology on a Stone Markov process “know” about the transition structure? Note that the base of clopens is required to be closed under the  $F_r$  operations, which are defined in terms of the transition function.

In this light, one can easily notice that the pseudometric associated to a metrized Markov process does in fact “the job” that we wanted to ensure in Chapter 8 when we defined the concept of dynamic-continuity. However, while the conditions in the definition of a dynamic-continuous pseudometric does not rule out trivial pseudometrics, such as the discrete one, the topological condition imposed in this Chapter does this. Indeed, it is not difficult to observe that the discrete pseudometric that measures to 0 the distance between bisimilar MPs and to 1 all the others, cannot be the support for a metrized Markov process since it trivially fail to satisfy (M). On the other hand in a pseudometrized MP a sequence of systems as the one in Figure 8.1 are convergent to the limit in the pseudometric topology.

The only work of which we are aware similar to this is a paper by Banaschewski and Pultr [4] called “A Stone duality for metric spaces.” They are not working with Markov processes, so there is nothing like the Aumann algebra structure there. They gave us the idea of using a metric diameter, but their axiomatization is different and the proofs that we developed do not resemble theirs.

The main impact of this work is to put quantitative reasoning about Markov processes on a firmer footing. It has been over a decade since metric analogues of bisimulation were developed, but they have not had the impact that they might have had. One reason is that with ordinary logical reasoning, one has a clear understanding of what completeness means, thus users of these logics have a good understanding of the power of the principles they are using. What does completeness mean for metric reasoning and approximate reasoning in general? The standard Stone-type duality theorem captures the concept of completeness; it is our hope that the present work will pave the way towards a similar understanding of approximate reasoning principles. There is much to be done, however. In the Chapter 8 we began investigating the relationship between the logic and metrics on Markov processes. The results of the present Chapter have strengthen and deepen these preliminary results.



# Bibliography

- [1] S. Abramsky, “Domain theory in logical form,” *Annals of Pure and Applied Logic*, vol. 51, pp. 1–77, 1991.
- [2] R. Aumann, “Interactive epistemology I: knowledge,” *International Journal of Game Theory*, vol. 28, pp. 263–300, 1999.
- [3] Robert Aumann. Interactive epistemology II: probability. *International Journal of Game Theory*, 28:301–314, 1999.
- [4] B. Banaschewski and A. Pultr. Stone duality for metric spaces. In R. A. G. Seely, editor, *Category Theory 1991*, volume 13 of *CMS Conference Proceedings*, pages 33–42. CMS, 1991.
- [5] N. Bezhanishvili, C. Kupke, and P. Panangaden, “Minimization via duality,” in *Logic, Language, Information and Computation - 19th International Workshop, WoLLIC 2012, Buenos Aires, Argentina, September 3-6, 2012. Proceedings*, ser. Lecture Notes in Computer Science, vol. 7456. Springer, 2012, pp. 191–205.
- [6] M. M. Bonsangue and A. Kurz. Duality for logics of transition systems. In *FoSSaCS*, pages 455–469, 2005.
- [7] F. Bonchi, M. Bonsangue, J. Rutten, and A. Silva, “Brzozowski’s algorithm (co)algebraically,” in *Logics and Program Semantics: Essays Dedicated to Dexter Kozen*, ser. Lecture Notes In Computer Science, R. Constable and A. Silva, Eds., vol. 7230. Springer-Verlag, 2012, pp. 12–23.
- [8] Luca Cardelli, Kim G. Larsen, and Radu Mardare. Continuous markovian logic - from complete axiomatization to the metric space of formulas. In *CSL*, pages 144–158, 2011.
- [9] C. C. Chang and H. J. Keisler, *Model Theory*. North-Holland, 1973.
- [10] V. Danos, J. Desharnais, F. Laviolette, and P. Panangaden, “Bisimulation and co-congruence for probabilistic systems,” *Information and Computation*, vol. 204, no. 4, pp. 503–523, 2006.

- [11] J. Desharnais, A. Edalat, and P. Panangaden. A logical characterization of bisimulation for labelled Markov processes. In *proceedings of the 13th IEEE Symposium On Logic In Computer Science, Indianapolis*, pages 478–489. IEEE Press, June 1998.
- [12] J. Desharnais, A. Edalat, and P. Panangaden. Bisimulation for labeled Markov processes. *Information and Computation*, 179(2):163–193, Dec 2002.
- [13] J. Desharnais, V. Gupta, R. Jagadeesan, and P. Panangaden. Metrics for labeled Markov systems. In *Proceedings of CONCUR99*, number 1664 in Lecture Notes in Computer Science. Springer-Verlag, 1999.
- [14] José Desharnais, Vineet Gupta, Radhakrishnan Jagadeesan, and Prakash Panangaden. A metric for labelled Markov processes. *Theoretical Computer Science*, 318(3):323–354, June 2004.
- [15] E.-E. Doberkat. *Stochastic Relations. Foundations for Markov Transition Systems*. Chapman and Hall, New York, 2007.
- [16] R. M. Dudley, *Real Analysis and Probability*. Wadsworth and Brookes/Cole, 1989.
- [17] R. Fagin and J. Y. Halpern. Reasoning about knowledge and probability. *Journal of the ACM*, 41(2):340–367, 1994.
- [18] M. Gehrke, S. Grigorieff, and J.-E. Pin, “Duality and equational theory of regular languages,” in *ICALP (2)*, 2008, pp. 246–257.
- [19] A. Giacalone, C. Jou, and S. Smolka. Algebraic reasoning for probabilistic concurrent systems. In *Proceedings of the Working Conference on Programming Concepts and Methods*, IFIP TC2, 1990.
- [20] R. Goldblatt, “On the role of the Baire category theorem in the foundations of logic,” *Journal of Symbolic logic*, pp. 412–422, 1985.
- [21] R. Goldblatt. Deduction systems for coalgebras over measurable spaces. *Journal of Logic and Computation*, 20(5):1069–1100, 2010.
- [22] Aviad Heifetz and Philippe Mongin. Probability logic for type spaces. *Games and Economic Behavior*, 35(1-2):31–53, April 2001.
- [23] B. Jacobs, “Probabilities, distribution monads, and convex categories,” *Theor. Comput. Sci.*, vol. 412, no. 28, pp. 3323–3336, 2011.
- [24] B. Jonsson and A. Tarski, “Boolean algebras with operators I,” *American Journal of Mathematics*, vol. 73, pp. 891–939, 1951.
- [25] P. Johnstone, *Stone Spaces*, ser. Cambridge Studies in Advanced Mathematics. Cambridge University Press, 1982, vol. 3.

- [26] D. Kozen. A probabilistic PDL. *Journal of Computer and Systems Sciences*, 30(2):162–178, 1985.
- [27] Dexter Kozen, Kim G. Larsen, Radu Mardare, and Prakash Panangaden. Stone duality for Markov processes. In *Proceedings of the 28th Annual IEEE Symposium On Logic In Computer Science*. IEEE Press, 2013.
- [28] K. G. Larsen and A. Skou. Bisimulation through probabilistic testing. *Information and Computation*, 94:1–28, 1991.
- [29] Kim G. Larsen, Radu Mardare, and Prakash Panangaden. Taking it to the limit: Approximate reasoning for markov processes. In *Proceedings of the 37th International Symposium on the Mathematical Foundations of Computer Science*, volume 7464 of *Lecture Notes In Computer Science*, pages 681–692, 2012.
- [30] Radu Mardare, Luca Cardelli, and Kim G. Larsen. Continuous markovian logics - axiomatization and quantified metatheory. *Logical Methods in Computer Science*, 8(4), 2012.
- [31] M. Mislove, J. Ouaknine, D. Pavlovic, and J. Worrell. Duality for labelled Markov processes. In I. Walukiewicz, editor, *Foundations of Software Science and Computation Structures, FOSSACS*, volume 2987 of *Lecture Notes In Computer Science*, pages 393–407, 2004.
- [32] Prakash Panangaden. *Labelled Markov Processes*. Imperial College Press, 2009.
- [33] G. D. Plotkin. Lecture notes on domain theory. Available from his home page as The Pisa Notes, 1983.
- [34] Gordon D. Plotkin. Dijkstra’s predicate transformers and Smyth’s power domains. In *Abstract Software Specifications*, volume 86 of *Lecture Notes in Computer Science*, pages 527–553. Springer, 1979.
- [35] H. Rasiowa and R. Sikorski, “A proof of the completeness theorem of gödel,” *Fund. Math*, vol. 37, pp. 193–200, 1950.
- [36] A. Silva, “Kleene coalgebra,” Ph.D. dissertation, University of Nijmegen, 2010.
- [37] M. Smyth. Powerdomains and predicate transformers. In J. Diaz, editor, *Proceedings of the International Colloquium On Automata Languages And Programming*, pages 662–676. Springer-Verlag, 1983. *Lecture Notes In Computer Science* 154.
- [38] M. H. Stone, “The theory of representations for boolean algebras,” *Trans. Amer. Math. Soc.*, vol. 40, pp. 37–111, 1936.

- [39] Franck van Breugel and James Worrell. An algorithm for quantitative verification of probabilistic systems. In K. G. Larsen and M. Nielsen, editors, *Proceedings of the Twelfth International Conference on Concurrency Theory - CONCUR'01*, number 2154 in Lecture Notes In Computer Science, pages 336–350. Springer-Verlag, 2001.
- [40] Franck van Breugel and James Worrell. Towards quantitative verification of probabilistic systems. In *Proceedings of the Twenty-eighth International Colloquium on Automata, Languages and Programming*. Springer-Verlag, July 2001.
- [41] C. Zhou. *A complete deductive system for probability logic with application to Harsanyi type spaces*. PhD thesis, Indiana University, 2007.
- [42] C. Zhou. and M. Ying, “Approximation of markov processes through filtration,” *Theoretical Computer Science*, p. in press, Jan 2012.

## **Part V**

# **Computational Aspects of Behavioural Distances**



In this Part of the Monograph we present some of the more practical research done in relation to behavioural distances for Markov processes. Unlike in the previous Parts, the Chapters in this Part are rather independent.

In Chapter 11 we approach the pseudometric introduced by Desharnais et al. [22] for discrete-time Markov chains (MCs). At the time we start working on this, van Breugel et al. [42] have presented a fixed point characterization of this pseudometric and several iterative algorithms have been developed in order to compute its approximation up to any degree of accuracy. In this Chapter, we propose an alternative approach, which allows us to compute the pseudometric exactly and efficiently in practice. This is inspired by the characterization of the undiscounted pseudometric using *couplings*. We aim at finding an optimal coupling using a greedy strategy that starts from an arbitrary coupling and repeatedly looks for new couplings that improve the discrepancy function. This strategy will eventually find an optimal coupling. We use it to support the design of an on-the-fly algorithm for computing the exact behavioural pseudometric that can be either applied to compute all the distances in the model or to compute only some particular distances. By using an on-the-fly approach, we avoid to exhaustively explore the state space. Instead, we only consider those fragments that are needed in the local computation. The efficiency of our algorithm has been evaluated on a significant set of randomly generated MCs. The results show that our algorithm performs orders of magnitude better than the corresponding iterative algorithms proposed, for instance in [14, 25]. Moreover, we provide empirical evidence for the fact that our algorithm enjoys good execution running times.

Our on-the-fly technique was proven to be sufficiently robust for even more complex models, such as the Markov Decision Processes with rewards (MDPs), for which a similar pseudometric was proposed in [25]. In Chapter 12 we present the BISIMDIST library, composed of two Mathematica packages which implement our on-the-fly algorithm for computing the bisimilarity distances both for MCs and MDPs. BISIMDIST is available at <http://people.cs.aau.dk/~mardare/tools.htm> together with simple tutorials presenting useful examples that show all the features of the library.

However, the case of MDPs is interesting in practice also because there are many examples reflecting the compositional features of such models. Realistic models are usually specified compositionally by means of operators that describe the interactions between the subcomponents. These specifications may thus suffer from an exponential growth of the state space. Moreover, when we consider approximate behaviors, the concept of distances requires *non-extensive* composition operators [22]. In Chapter 13 we study to which extent compositionality on MDPs can be exploited in the computation of the behavioral pseudometrics of [25], hence how the compositional structure of processes can be used in an approximated analysis of behaviors. This research has been inspired by our experience with probabilistic and stochastic process algebras and by our study of the interaction between the operational semantics of such an algebra and

the behavioural pseudometrics, detailed in Part I of this Monograph. As a result, we provide an algorithm to compute the bisimilarity pseudometric by exploiting both the on-the-fly state space exploration and the compositional structure of MDPs built over “safe” operators. Experimental results show that the compositional optimization yields a significant additional improvement on top of that obtained by the on-the-fly method.

In Chapter 14 we push further the idea of the on-the-fly algorithm for computing the behavioural distances, and we move from discrete-time models to continuous time models. Continuous-time Markov chains (CTMCs) are one of the most prominent models in performance and dependability analysis, constituting the underlying semantics of many modeling formalisms for real-time probabilistic systems. We extend the distance considered in the previous Chapters to CTMCs and we show that it can be computed in polynomial time in the size of the CTMC. This is obtained by reducing the problem of computing the distance to that of finding an optimal solution of a linear program that can be solved using the ellipsoid method. However, our linear program characterization has a number of constraints that is bounded by a polynomial in the size of the CTMC. This, in particular, allows us to avoid the use of the ellipsoid algorithm in favor of the simplex or the interior point methods. Nevertheless, we propose to follow an on-the-fly approach for computing the distance, by relying on the concept of coupling structure generalized for CTMCs. With respect to the previous algorithms, in this Chapter we have a series of innovations required by the fact that we need, in addition, to handle the information regarding the residence-time in a state of the system.

As we have already seen, the first proposals of behavioral distances in the literature are based on the Kantorovich metric and are branching-time. In Chapter 15 we consider a linear-time metric, motivated by the fact that in many applications, such as in systems biology, modeling/testing, and machine learning, the system to be modeled cannot be internally accessed, but only tested via observations performed over a set of random executions. In this respect, we introduce a very general class of models, the semi-Markov chains (SMCs), which are continuous-time probabilistic transition systems where the residence time on states is governed by generic distributions on the positive real line. SMCs subsume many probabilistic models including the discrete-time Markov chains and the continuous-time Markov Chains. Regarding the behavioural distance, we study the total variation between probability measures induced by an SMC over infinite timed traces, which corresponds to the largest possible difference between the probabilities that the measures assign to the same event. In applications, the events are specified either as metric temporal logic (MTL) formulas [2, 3], or languages accepted by timed automata (TAs) [1]. We prove that each of these classes of specifications characterize the total variation.

However, the problem of computing the total variation is known to be at least NP-hard already for the case of MCs [17, 34] and there is no proof that it is, in fact, decidable. In this context, in Chapter 15, we also prove that the problem of approximating the total variation distance with arbitrary precision is computable. This is done providing

two sequences that converge from below and above to the total variation distance. Our results are based on a duality that characterizes the total variation between two measures as the minimal discrepancy associated with their couplings. The computability of the converging sequences provides us with a computable procedure to approximate the total variation distance on SMCs with arbitrary precision.

This Part of the Monograph comprises results presented in the following articles.

- [I] G. Bacci, G. Bacci, K.G. Larsen, R. Mardare. *On-The-Fly Computation of Bisimilarity Distances*. Logical Methods in Computer Science, LMCS, to appear, 2015.
- [II] G. Bacci, G. Bacci, K.G. Larsen, R. Mardare. *On the Total Variation Distance of Semi-Markov Chains*. under submission 2014.
- [III] G. Bacci, G. Bacci, K.G. Larsen, R. Mardare. *Computing Behavioral Distances, Compositionally*. In Proc. of 38th International Symposium Mathematical Foundations of Computer Science, MFCS 2013, Lecture Notes in Computer Science, LNCS 8087, pages: 74-85, 2013.
- [IV] G. Bacci, G. Bacci, K.G. Larsen, R. Mardare. *The BisimDist Library: Efficient Computation of Bisimilarity Distances for Markovian Models*. In Proc. of 10th International Conference Quantitative Evaluation of Systems, QEST 2013, Lecture Notes in Computer Science, LNCS 8054, pages: 278-281, 2013.
- [V] G. Bacci, G. Bacci, K.G. Larsen, R. Mardare. *On-the-Fly Exact Computation of Bisimilarity Distances*. In Proc. of 19th International Conference Tools and Algorithms for the Construction and Analysis of Systems, TACAS 2013, Lecture Notes in Computer Science, LNCS 7795, pages:1-15, 2013.



# Chapter 11

## Behavioural Distances for Markov Chains

### 11.1 Introduction

Probabilistic bisimulation for Markov chains (MCs), introduced by Larsen and Skou [32], is the key concept for reasoning about the equivalence of probabilistic systems. However, when one focuses on quantitative behaviours it becomes obvious that such an equivalence is too “exact” for many purposes as it only relates processes with identical behaviours. In various applications, such as systems biology [41], games [13], planning [18] or security [10], we are interested in knowing whether two processes that may differ by a small amount in the real-valued parameters (probabilities) have “sufficiently” similar behaviours. This motivated the development of the metric theory for MCs, initiated by Desharnais et al. [22] and greatly developed and explored by van Breugel, Worrell and others [42,44]. It consists in proposing a *bisimilarity distance* (*pseudometric*), which measures the behavioural similarity of two MCs. The pseudometric proposed by Desharnais et al. is parametric in a *discount factor*  $\lambda \in (0,1]$  that controls the significance of the future in the measurement.

Since van Breugel et. al. have presented a fixed point characterization of the aforementioned pseudometric in [42], several iterative algorithms have been developed in order to compute its approximation up to any degree of accuracy [25,42,44]. Recently, Chen et. al. [14] proved that, for finite MCs with rational transition function, the bisimilarity pseudometrics can be computed exactly in polynomial time. The proof consists in describing the pseudometric as the solution of a linear program that can be solved using the *ellipsoid method*. Although the ellipsoid method is theoretically efficient, “*computational experiments with the method are very discouraging and it is in practice by no means a competitor of the, theoretically inefficient, simplex method*”, as stated in [40]. Unfortunately, in this case the simplex method cannot be used to speed up performances in practice, since the linear program to be solved may have an exponential number of constraints in the number of states of the MC.

In this Chapter, we propose an alternative approach to this problem, which allows us to compute the pseudometric exactly and efficiently in practice. This is inspired by the characterization of the undiscounted pseudometric using *couplings*, given in [14], which we extend to generic discount factors.

A coupling for a pair of states of a given MC is a function that describes a possible redistribution of the transition probabilities of the two states; it is evaluated by the *discrepancy function* that measures the behavioural dissimilarities between the two states. In [14] it is shown that the bisimilarity pseudometric for a given MC is the minimum among the discrepancy functions corresponding to all the couplings that can be defined for that MC; moreover, the bisimilarity pseudometric is itself a discrepancy function corresponding to an *optimal coupling*. This suggests that the problem of computing the pseudometric can be reduced to the problem of finding a coupling with the least discrepancy function.

Our approach aims at finding an optimal coupling using a greedy strategy that starts from an arbitrary coupling and repeatedly looks for new couplings that improve the discrepancy function. This strategy will eventually find an optimal coupling. We use it to support the design of an on-the-fly algorithm for computing the exact behavioural pseudometric that can be either applied to compute all the distances in the model or to compute only some designated distances. The advantage of using an on-the-fly approach consists in the fact that we do not need to exhaustively explore the state space, nor to construct entire couplings, but only those fragments that are needed in the local computation.

The efficiency of our algorithm has been evaluated on a significant set of randomly generated MCs. The results show that our algorithm performs orders of magnitude better than the corresponding iterative algorithms proposed, for instance in [14, 25]. Moreover, we provide empirical evidence for the fact that our algorithm enjoys good execution running times.

One of the main practical advantages of our approach consists in the fact that it can focus on computing only the distances between states that are of particular interest. This is useful in practice, for instance when large systems are considered and visiting the entire state space is expensive. A similar issue has been considered by Comanici et. al., in [19], who have noticed that for computing the approximated pseudometric one does not need to update the current value for all the pairs at each iteration, but it is sufficient to only focus on the pairs where changes are happening rapidly. Our approach goes much beyond this idea. Firstly, we are not only looking at approximations of the bisimilarity distance, but we develop an exact algorithm. Secondly, we provide a termination condition that can be checked locally, still ensuring that the local optimum corresponds to the global one.

In addition, our method can be applied to decide whether two states of an MC are probabilistic bisimilar, to identify the bisimilarity classes for a given MC or to solve lumpability problems.

It can also be used to address issues related to state space reduction for MCs: our

technique indicates sets of neighbour states that can be collapsed due to their similarity and it provides a tool to estimate the difference between the initial MC and the reduced one.

Last but not least, our approach can be used with approximation techniques as, for instance, to provide a least over-approximation of the behavioural distance given over-estimates of some particular distances. This can be further integrated with other approximate algorithms having the advantage of the on-the-fly state space exploration.

## 11.2 Preliminaries: The Transportation Problem

Central in this Chapter is the so-called (homogeneous) *transportation problem* considered in 1941 by Hitchcock and, independently, in 1947 by Koopmans. This problem can be intuitively described as: a homogeneous product is to be shipped in the amounts  $a_1, \dots, a_m$  respectively, from each of  $m$  shipping *origins* and received in amounts  $b_1, \dots, b_n$  respectively, by each of  $n$  shipping *destinations*. The cost of shipping a unit amount from the  $i$ -th origin to the  $j$ -th destination is  $c_{i,j}$  and is known for all combinations  $(i, j)$ . The problem is to determine an optimal *shipping schedule*, i.e. the amount  $x_{i,j}$  to be shipped over all routes  $(i, j)$ , which minimizes the total cost of transportation.

It can be easily formalized as a linear programming problem

$$\begin{aligned} & \text{minimize } \sum_{i=1}^m \sum_{j=1}^n c_{i,j} \cdot x_{i,j} \\ & \text{such that } \sum_{j=1}^n x_{i,j} = a_i && (i = 1, \dots, m) \\ & \sum_{i=1}^m x_{i,j} = b_j && (j = 1, \dots, n) \\ & x_{i,j} \geq 0 && (i = 1, \dots, m \text{ and } j = 1, \dots, n) \end{aligned}$$

The set of schedules feasible for a transportation problem, which is formalized as a conjunction of linear constraints, describes a (bounded) convex polytope in  $\mathbb{R}^2$ , often called transportation polytope.

There are several algorithms in literature which efficiently solve (not necessarily homogeneous) transportation problems. Among these we recall [20, 26].

## 11.3 Markov Chains and Bisimilarity Pseudometrics

In this section we give the definitions of (*discrete-time*) *Markov chains* (MCs) and *probabilistic bisimilarity* for MCs [32]. Then we recall the *bisimilarity pseudometric* of Desharnais et. al. [22], but rather than giving its first logical definition, we present its fixed point characterization given by van Breugel et. al. [42].

**Definition 11.3.1** (Markov chain). A (discrete-time) Markov chain is a tuple

$$\mathcal{M} = (S, A, \pi, \ell)$$

consisting of a countable nonempty set  $S$  of states, a nonempty set  $A$  of labels, a transition probability function  $\pi: S \times S \rightarrow [0, 1]$  such that,

$$\text{for arbitrary } s \in S, \quad \sum_{t \in S} \pi(s, t) = 1,$$

and a labelling function  $\ell: S \rightarrow A$ .  $\mathcal{M}$  is finite if its support set  $S$  is finite.

Given a finite MC  $\mathcal{M} = (S, A, \pi, \ell)$ , we identify the transition probability function  $\pi$  with its *transition matrix*  $(\pi(s, t))_{s, t \in S}$ . For  $s, t \in S$ , we denote by  $\pi(s, \cdot)$  and  $\pi(\cdot, t)$ , respectively, the probability distribution of exiting from  $s$  to any state and the sub-probability distribution of entering to  $t$  from any state.

The MC  $\mathcal{M}$  induces an *underlying (directed) graph*, denoted by  $\mathcal{G}(\mathcal{M})$ , where the states act as vertices and  $(s, t)$  is an edge in  $\mathcal{G}(\mathcal{M})$ , if and only if,  $\pi(s, t) > 0$ .

For a subset  $Q \subseteq S$ , we denote by  $R_{\mathcal{M}}(Q)$  the set of states reachable from some  $s \in Q$ , and by  $R_{\mathcal{M}}(s)$  we denote  $R_{\mathcal{M}}(\{s\})$ .

From a theoretical point of view, it is irrelevant whether the transition probability function of a given Markov Chain has rational values or not. However, for algorithmic purposes, in this paper we assume that for arbitrary  $s, t \in S$ ,  $\pi(s, t) \in \mathbb{Q} \cap [0, 1]$ .

For computational reasons, in the rest of the paper we restrict our investigation to finite Markov chains.

**Definition 11.3.2** (Probabilistic Bisimulation). Let  $\mathcal{M} = (S, A, \pi, \ell)$  be an MC. An *equivalence relation*  $R \subseteq S \times S$  is a *probabilistic bisimulation* if whenever  $s R t$ , then

$$(i) \quad \ell(s) = \ell(t) \text{ and,}$$

$$(ii) \quad \text{for each } R\text{-equivalence class } E, \quad \sum_{u \in E} \pi(s, u) = \sum_{u \in E} \pi(t, u).$$

Two states  $s, t \in S$  are *bisimilar*, written  $s \sim t$ , if they are related by some probabilistic bisimulation.

This definition is due to Larsen and Skou [32]. Intuitively, two states are bisimilar if they have the same label and their probability of moving by a single transition to any given equivalence class is always the same.

The notion of equivalence can be relaxed by means of a pseudometric, which measures how far apart two elements are from each other; whenever they are at zero distance, they are equivalent.

The bisimilarity pseudometric of Desharnais et. al. [22] on MCs enjoys the property that two states are at zero distance if and only if they are bisimilar. This pseudometric can be defined as the least fixed point of an operator based on the Kantorovich metric for comparing probability distributions, which makes use of the notion of *matching*.

**Definition 11.3.3** (Matching). *Let  $\mu, \nu: S \rightarrow [0, 1]$  be probability distributions on  $S$ . A matching for the pair  $(\mu, \nu)$  is a probability distribution  $\omega: S \times S \rightarrow [0, 1]$  on  $S \times S$  satisfying*

$$\forall u \in S. \sum_{s \in S} \omega(u, s) = \mu(u), \quad \forall v \in S. \sum_{s \in S} \omega(s, v) = \nu(v). \quad (11.3.1)$$

We call  $\mu$  and  $\nu$ , respectively, the left and the right marginals of  $\omega$ .

In the following, we denote by  $\mu \otimes \nu$  the set of all matchings for  $(\mu, \nu)$ .

**Remark 11.3.4.** Note that, for  $S$  finite, (11.3.1) describes the constraints of a homogeneous transportation problem (TP) [20, 26], where the vector  $(\mu(u))_{u \in S}$  specifies the amounts to be shipped and  $(\nu(v))_{v \in S}$  the amounts to be received.

Thus, a matching  $\omega$  for  $(\mu, \nu)$  induces a matrix  $(\omega(u, v))_{u, v \in S}$  to be thought as a shipping schedule belonging to the transportation polytope  $\mu \otimes \nu$ . Hereafter, we denote by  $TP(c, \nu, \mu)$  the TP with cost matrix  $(c(u, v))_{u, v \in S}$  and marginals  $\nu$  and  $\mu$ . ■

For  $\mathcal{M} = (S, A, \pi, \ell)$  an MC, and  $\lambda \in (0, 1]$  a discount factor, the operator

$$\Delta_\lambda^{\mathcal{M}}: [0, 1]^{S \times S} \rightarrow [0, 1]^{S \times S},$$

defined for  $d: S \times S \rightarrow [0, 1]$  and  $s, t \in S$ , is given as follows.

$$\Delta_\lambda^{\mathcal{M}}(d)(s, t) = \begin{cases} 1 & \text{if } \ell(s) \neq \ell(t) \\ \lambda \cdot \min_{\omega \in \pi(s, \cdot) \otimes \pi(t, \cdot)} \sum_{u, v \in S} d(u, v) \cdot \omega(u, v) & \text{if } \ell(s) = \ell(t) \end{cases}$$

In the above definition,  $\pi(s, \cdot) \otimes \pi(t, \cdot)$  is a closed polytope so that the minimum is well defined and it corresponds to the optimal value of  $TP(d, \pi(s, \cdot), \pi(t, \cdot))$ .

The set  $[0, 1]^{S \times S}$  is endowed with the partial order  $\sqsubseteq$  defined by

$$d \sqsubseteq d' \text{ iff } d(s, t) \leq d'(s, t) \text{ for all } s, t \in S.$$

This forms a complete lattice with bottom element  $\bar{0}$  and top element  $\bar{1}$ , defined as

$$\bar{0}(s, t) = 0 \text{ and } \bar{1}(s, t) = 1, \text{ for all } s, t \in S.$$

For  $D \subseteq [0, 1]^{S \times S}$ , the least upper bound  $\sqcup D$ , and greatest lower bound  $\sqcap D$  are given for all  $s, t \in S$  by

$$\begin{aligned} (\sqcup D)(s, t) &= \sup_{d \in D} d(s, t), \\ (\sqcap D)(s, t) &= \inf_{d \in D} d(s, t). \end{aligned}$$

In [42], for any  $\mathcal{M}$  and  $\lambda \in [0, 1]$ ,  $\Delta_\lambda^{\mathcal{M}}$  is proved to be monotonic, thus, by Tarski's fixed point theorem, it admits least and greatest fixed points.

**Definition 11.3.5** (Bisimilarity pseudometric). *Let  $\mathcal{M}$  be an MC and  $\lambda \in (0, 1]$  be a discount factor. Then, the  $\lambda$ -discounted bisimilarity pseudometric for  $\mathcal{M}$ , denoted by  $\delta_\lambda^{\mathcal{M}}$ , is the least fixed point of  $\Delta_\lambda^{\mathcal{M}}$ .*

Hereafter,  $\Delta_\lambda^{\mathcal{M}}$  and  $\delta_\lambda^{\mathcal{M}}$  will be denoted simply by  $\Delta_\lambda$  and  $\delta_\lambda$ , respectively, when the Markov chain  $\mathcal{M}$  is clear from the context.

## 11.4 Alternative Characterization of the Pseudometric

In [14], Chen et. al. proposed an alternative characterization of  $\delta_1$ , relating the pseudometric to the notion of *coupling*. In this section, we recall the definition of coupling, and generalize the characterization for generic discount factors.

**Definition 11.4.1** (Coupling). *Let  $\mathcal{M} = (S, A, \pi, \ell)$  be a finite MC.*

*The Markov chain  $\mathcal{C} = (S \times S, A \times A, \omega, l)$  is said a coupling for  $\mathcal{M}$  if, for all  $s, t \in S$ ,*

1.  $\omega((s, t), \cdot) \in \pi(s, \cdot) \otimes \pi(t, \cdot)$ , and
2.  $l(s, t) = (\ell(s), \ell(t))$ .

A coupling for  $\mathcal{M}$  can be seen as a probabilistic pairing of two copies of  $\mathcal{M}$  running synchronously, although not necessarily independently.

Couplings have been used to characterize weak ergodicity of arbitrary Markov chains [29], or to give upper bounds on convergence to stationary distributions [7, 35].

Given a coupling  $\mathcal{C} = (S \times S, A \times A, \omega, l)$  for  $\mathcal{M} = (S, A, \pi, \ell)$  we define the operator

$$\Gamma_\lambda^{\mathcal{C}}: [0, 1]^{S \times S} \rightarrow [0, 1]^{S \times S}$$

for  $d: S \times S \rightarrow [0, 1]$  and  $s, t \in S$ , as follows.

$$\Gamma_\lambda^{\mathcal{C}}(d)(s, t) = \begin{cases} 1 & \text{if } \ell(s) \neq \ell(t) \\ \lambda \cdot \sum_{u, v \in S} d(u, v) \cdot \omega((s, t), (u, v)) & \text{if } \ell(s) = \ell(t) \end{cases}$$

One can easily verify that, for any  $\lambda \in (0, 1]$ ,  $\Gamma_\lambda^{\mathcal{C}}$  is well-defined and order preserving. Hence, by Tarski's fixed point theorem,  $\Gamma_\lambda^{\mathcal{C}}$  admits a least fixed point, which we denote by  $\gamma_\lambda^{\mathcal{C}}$ .

In Section 14.5.1 we will see that, for any  $s, t \in S$ ,  $\gamma_1^{\mathcal{C}}(s, t)$  corresponds to the probability of reaching a state  $(u, v)$  with  $\ell(u) \neq \ell(v)$  starting from the state  $(s, t)$  in the underlying graph of  $\mathcal{C}$ . For this reason we will call  $\gamma_\lambda^{\mathcal{C}}$  the  $\lambda$ -discounted discrepancy of  $\mathcal{C}$  or simply the  $\lambda$ -discrepancy of  $\mathcal{C}$ .

**Lemma 11.4.2.** *Let  $\mathcal{M}$  be an MC,  $\mathcal{C}$  be a coupling for  $\mathcal{M}$ , and  $\lambda \in (0, 1]$  be a discount factor. If  $d = \Gamma_\lambda^{\mathcal{C}}(d)$  then  $\delta_\lambda \sqsubseteq d$ .*

*Proof.* Assume that  $\mathcal{M} = (S, A, \pi, \ell)$  and  $\mathcal{C} = (S \times S, A \times A, \omega, l)$ .

In order to prove  $\delta_\lambda \sqsubseteq d$ , it suffices to show that  $\Delta_\lambda(d) \sqsubseteq d$ . Indeed, by Tarski's fixed point theorem,  $\delta_\lambda$  is a lower bound of  $\{d \mid \Delta_\lambda(d) \sqsubseteq d\}$ .

Let  $s, t \in S$ . If  $\ell(s) \neq \ell(t)$ , then

$$\Delta_\lambda(d)(s, t) = 1 = \Gamma_\lambda^{\mathcal{C}}(d)(s, t) = d(s, t).$$

If  $\ell(s) = \ell(t)$ ,

$$\Delta_\lambda(d)(s, t) = \lambda \cdot \min_{\omega' \in \pi(s, \cdot) \otimes \pi(t, \cdot)} \sum_{u, v \in S} d(u, v) \cdot \omega'(u, v)$$

and

$$\Gamma_\lambda^{\mathcal{C}}(d)(s, t) = \lambda \cdot \sum_{u, v \in S} d(u, v) \cdot \omega((s, t), (u, v)).$$

Since  $\omega((s, t), \cdot) \in \pi(s, \cdot) \otimes \pi(t, \cdot)$  (Definition 14.4.1), we have that

$$\Delta_\lambda(d)(s, t) \leq \Gamma_\lambda^{\mathcal{C}}(d)(s, t) = d(s, t).$$

□

As a consequence of Lemma 13.5.2 we obtain the following characterization for  $\delta_\lambda$ , which generalizes [14, Theorem 8] for generic discount factors.

**Theorem 11.4.3** (Minimum coupling criterion). *Let  $\mathcal{M}$  be an MC and  $\lambda \in (0, 1]$  be a discount factor. Then,*

$$\delta_\lambda = \min\{\gamma_\lambda^{\mathcal{C}} \mid \mathcal{C} \text{ coupling for } \mathcal{M}\}.$$

*Proof.* For any fixed  $d \in [0, 1]^{S \times S}$  there exists a coupling  $\mathcal{C}$  for  $\mathcal{M}$  such that

$$\Gamma_\lambda^{\mathcal{C}}(d) = \Delta_\lambda(d).$$

Indeed, we can take as transition function for  $\mathcal{C}$ , the joint probability distribution  $\omega$  such that, for all  $s, t \in S$ ,

$$\sum_{u,v \in S} d(u,v) \cdot \omega((s,t), (u,v))$$

achieves the minimum value.

Let  $\mathcal{D}$  be a coupling for  $\mathcal{M}$  such that

$$\Gamma_\lambda^{\mathcal{D}}(\delta_\lambda) = \Delta_\lambda(\delta_\lambda).$$

By Definition 14.2.5,  $\Delta_\lambda(\delta_\lambda) = \delta_\lambda$ , therefore  $\delta_\lambda$  is a fixed point for  $\Gamma_\lambda^{\mathcal{D}}$ .

By Lemma 13.5.2,  $\delta_\lambda$  is a lower bound of the set of fixed points of  $\Gamma_\lambda^{\mathcal{D}}$ , therefore  $\delta_\lambda = \gamma_\lambda^{\mathcal{D}}$ .

By Lemma 13.5.2, we have also that, for any coupling  $\mathcal{C}$  of  $\mathcal{M}$ ,  $\delta_\lambda \sqsubseteq \gamma_\lambda^{\mathcal{C}}$ .

Therefore, given the set  $D = \{\gamma_\lambda^{\mathcal{C}} \mid \mathcal{C} \text{ coupling for } \mathcal{M}\}$ , it follows that  $\delta_\lambda \in D$  and  $\delta_\lambda$  is a lower bound for  $D$ .

Hence, by antisymmetry of  $\sqsubseteq$ ,  $\delta_\lambda = \min D$ . □

## 11.5 Exact Computation of Bisimilarity Distance

Inspired by the characterization given in Theorem 14.4.3, in this section we propose a procedure to exactly compute the bisimilarity pseudometric.

For  $\lambda \in (0, 1]$ , the set of couplings for  $\mathcal{M}$  can be endowed with the preorder  $\preceq_\lambda$  defined as

$$\mathcal{C} \preceq_\lambda \mathcal{D} \text{ if and only if } \gamma_\lambda^{\mathcal{C}} \sqsubseteq \gamma_\lambda^{\mathcal{D}}.$$

Theorem 14.4.3 suggests to look to all the couplings  $\mathcal{C}$  for  $\mathcal{M}$  in order to find an optimal one, i.e., minimal with respect to  $\preceq_\lambda$ . However, it is clear that the enumeration of all the couplings is unfeasible, therefore it is crucial to provide an efficient search strategy which prevents us to do that. Moreover, we also need an efficient method for computing the  $\lambda$ -discrepancy.

In Subsection 14.5.1 the problem of computing the  $\lambda$ -discrepancy of a coupling  $\mathcal{C}$  is reduced to the problem of computing reachability probabilities in  $\mathcal{C}$ . Then, Subsection 14.5.2 illustrates a greedy strategy that explores the set of couplings until an optimal one is eventually reached.

### 11.5.1 Computing the $\lambda$ -Discrepancy

In this section, we first recall the problem of computing the reachability probability for general MCs [7], then we instantiate it to compute the  $\lambda$ -discrepancy.

Let  $\mathcal{M} = (S, A, \pi, \ell)$  be an MC, and  $x_s$  denote the probability of reaching  $G \subseteq S$  from  $s \in S$ . The goal is to compute  $x_s$  for all  $s \in S$ .

The following holds

$$x_s = 1 \quad \text{if } s \in G, \quad x_s = \sum_{t \in S} x_t \cdot \pi(s, t) \quad \text{if } s \in S \setminus G, \quad (11.5.1)$$

that is, either  $G$  is already reached, or it can be reached by way of another state.

Equation (11.5.1) defines a linear equation system of the form  $\vec{x} = \vec{A}\vec{x} + \vec{b}$ , where

$$S_? = S \setminus G, \quad \vec{x} = (x_s)_{s \in S_?}, \quad \vec{A} = (\pi(s, t))_{s, t \in S_?}, \quad \text{and} \quad \vec{b} = \left( \sum_{t \in G} \pi(s, t) \right)_{s \in S_?}.$$

This linear equation system always admits a solution in  $[0, 1]^S$ . However, it may not be unique.

Since we are interested in the least solution, we address this problem by fixing each free variable to zero, so that we obtain a reduced system with a unique solution. This can be easily done by inspecting the graph  $\mathcal{G}(\mathcal{M})$ : all variables with zero probability of reaching  $G$  are detected by checking that they cannot be reached from any state in  $G$  in the reverse graph of  $\mathcal{G}(\mathcal{M})$ .

Regarding the  $\lambda$ -discrepancy for a coupling  $\mathcal{C}$ , if  $\lambda = 1$ , one can directly instantiate the aforementioned method with

$$G = \{(s, t) \in S \times S \mid \ell(s) \neq \ell(t)\} \quad \text{and} \quad S_? = (S \times S) \setminus G.$$

As for generic  $\lambda \in (0, 1]$ , the discrepancy  $\gamma_\lambda^{\mathcal{C}}$  can be formulated as the least solution in  $[0, 1]^{S \times S}$  of the linear equation system

$$\vec{x} = \lambda \vec{A} \vec{x} + \lambda \vec{b}. \quad (11.5.2)$$

**Remark 11.5.1.** If one is interested in computing the  $\lambda$ -discrepancy for a particular pair of states  $(s, t)$ , the method above can be applied on the least independent subsystem of Equation (11.5.2) containing the variable  $x_{(s, t)}$ . Moreover, assuming that for some pairs the  $\lambda$ -discrepancy is already known, the goal set can be extended with all those pairs with  $\lambda$ -discrepancy greater than zero. ■

## 11.5.2 Greedy Search Strategy for Computing an Optimal Coupling

In this subsection, we give a greedy strategy for moving toward an optimal coupling starting from a given one. Then we provide sufficient and necessary conditions for a coupling, ensuring that its associated  $\lambda$ -discrepancy coincides with  $\delta_\lambda$ .

Hereafter, we fix a coupling  $\mathcal{C} = (S \times S, A \times A, \omega, l)$  for  $\mathcal{M} = (S, A, \pi, \ell)$ .

Let  $s, t \in S$  and  $\mu$  be a matching for  $(\pi(s, \cdot), \pi(t, \cdot))$ .

We denote by  $\mathcal{C}[(s,t)/\mu]$  the coupling for  $\mathcal{M}$  with the same labeling function of  $\mathcal{C}$  and transition function  $\omega'$  defined by  $\omega'((u,v), \cdot) = \omega((u,v), \cdot)$ , for all  $(u,v) \neq (s,t)$ , and  $\omega'((s,t), \cdot) = \mu$ .

**Lemma 11.5.2.** *Let  $\mathcal{C}$  be a coupling for  $\mathcal{M} = (S, A, \pi, \ell)$ ,  $s, t \in S$ ,  $\omega' \in \pi(s, \cdot) \otimes \pi(t, \cdot)$ , and  $\mathcal{D} = \mathcal{C}[(s,t)/\omega']$ .*

$$\text{If } \Gamma_{\lambda}^{\mathcal{D}}(\gamma_{\lambda}^{\mathcal{C}})(s,t) < \gamma_{\lambda}^{\mathcal{C}}(s,t), \text{ then } \gamma_{\lambda}^{\mathcal{D}} \sqsubset \gamma_{\lambda}^{\mathcal{C}}.$$

*Proof.* It suffices to show that  $\Gamma^{\mathcal{D}}(\gamma_{\lambda}^{\mathcal{C}}) \sqsubset \gamma_{\lambda}^{\mathcal{C}}$ , i.e.,  $\gamma_{\lambda}^{\mathcal{C}}$  is a strict post-fixed point of  $\Gamma_{\lambda}^{\mathcal{D}}$ . Then, the thesis follows by Tarski's fixed point theorem.

Assume  $\bar{\omega}$  be the transition function of  $\mathcal{D}$  and let  $u, v \in S$ .

If  $\ell(u) \neq \ell(v)$ , then

$$\Gamma_{\lambda}^{\mathcal{D}}(\gamma_{\lambda}^{\mathcal{C}})(u,v) = 1 = \Gamma_{\lambda}^{\mathcal{C}}(\gamma_{\lambda}^{\mathcal{C}})(u,v) = \gamma_{\lambda}^{\mathcal{C}}(u,v).$$

Notice that, this also means that  $\ell(s) = \ell(t)$ , since  $\Gamma_{\lambda}^{\mathcal{D}}(\gamma_{\lambda}^{\mathcal{C}})(s,t) < \gamma_{\lambda}^{\mathcal{C}}(s,t)$ , by hypothesis.

If  $\ell(u) = \ell(v)$  and  $(u,v) \neq (s,t)$ , by definition of  $\mathcal{D}$ , we have that

$$\bar{\omega}((u,v), \cdot) = \omega((u,v), \cdot),$$

hence

$$\Gamma_{\lambda}^{\mathcal{C}}(\gamma_{\lambda}^{\mathcal{C}})(u,v) = \Gamma_{\lambda}^{\mathcal{D}}(\gamma_{\lambda}^{\mathcal{C}})(u,v).$$

This proves  $\Gamma^{\mathcal{D}}(\gamma_{\lambda}^{\mathcal{C}}) \sqsubset \gamma_{\lambda}^{\mathcal{C}}$ . □

Lemma 11.5.2 states that  $\mathcal{C}$  can be improved w.r.t.  $\preceq_{\lambda}$  by updating its transition function at  $(s,t)$ , if  $\ell(s) = \ell(t)$  and there exists  $\omega' \in \pi(s, \cdot) \otimes \pi(t, \cdot)$  such that

$$\sum_{u,v \in S} \gamma_{\lambda}^{\mathcal{C}}(u,v) \cdot \omega'(u,v) < \sum_{u,v \in S} \gamma_{\lambda}^{\mathcal{C}}(u,v) \cdot \omega((s,t), (u,v)).$$

Notice that, an optimal schedule  $\omega'$  for  $TP(\gamma_{\lambda}^{\mathcal{C}}, \pi(s, \cdot), \pi(t, \cdot))$  enjoys the above condition, so that, the update  $\mathcal{C}[(s,t)/\omega']$  improves  $\mathcal{C}$ . This gives us a strategy for moving toward  $\delta_{\lambda}$  by successive improvements on the couplings.

**Lemma 11.5.3.** *Let  $s, t \in S$ , and  $\gamma_1^{\mathcal{C}} = \Delta_1(\gamma_{\lambda}^{\mathcal{C}})$ . Then,*

$$\gamma_1^{\mathcal{C}}(s,t) = 1 \text{ iff } \delta_1(s,t) = 1.$$

*Proof.* ( $\Leftarrow$ ) Follows from Theorem 14.4.3.

( $\Rightarrow$ ) Assume  $\omega$  be the transition function of  $\mathcal{C}$ .

If  $\ell(s) \neq \ell(t)$  the thesis follows trivially.

Assume  $\ell(s) = \ell(t)$ .

$$\begin{aligned} 1 &= \gamma_1^{\mathcal{C}}(s, t) = \Gamma_1^{\mathcal{C}}(\gamma_1^{\mathcal{C}})(s, t) \\ &= \sum_{u, v \in S} \gamma_1^{\mathcal{C}}(u, v) \cdot \omega((s, t), (u, v)) \\ &\leq \sum_{u, v \in S} \omega((s, t), (u, v)) = 1 \end{aligned}$$

Thus, whenever  $\omega((s, t), (u, v)) > 0$ , we have that  $\gamma_1^{\mathcal{C}}(u, v) = 1$ . By hypothesis,  $\gamma_1^{\mathcal{C}} = \Delta_1(\gamma_1^{\mathcal{C}})$ , therefore

$$1 = \gamma_1^{\mathcal{C}}(s, t) = \min_{\omega' \in \pi(s, \cdot) \otimes \pi(t, \cdot)} \sum_{u, v \in S} \gamma_1^{\mathcal{C}}(u, v) \cdot \omega'(u, v).$$

Hence, there is no coupling that can improve the summation. Therefore, by Theorem 14.4.3,  $\delta_1(s, t) = 1$ . □

**Lemma 11.5.4.** *Let  $\mathcal{C}$  be a coupling for  $\mathcal{M}$ . For any  $\lambda \in (0, 1]$ ,*

$$\text{if } \gamma_\lambda^{\mathcal{C}} = \Delta_\lambda(\gamma_\lambda^{\mathcal{C}}), \text{ then } \delta_\lambda = \gamma_\lambda^{\mathcal{C}}.$$

*Proof.* By Definition 14.2.5, it suffices to prove that if  $\gamma_\lambda^{\mathcal{C}}$  is a fixed point for  $\Delta_\lambda$ , it is also the least one. We distinguish two cases: when  $\lambda < 1$  and  $\lambda = 1$ .

For  $\lambda < 1$ , [14, Theorem 6] states that  $\Delta_\lambda$  has a unique fixed point. By hypothesis  $\gamma_\lambda^{\mathcal{C}}$  is a fixed point for  $\Delta_\lambda$ , therefore it is also the least one.

For  $\lambda = 1$ , we proceed by contradiction.

Assume  $\delta_1 \neq \gamma_1^{\mathcal{C}}$  and  $\omega$  be the transition function of  $\mathcal{C}$ . By  $\delta_1 \neq \gamma_1^{\mathcal{C}}$  and Theorem 14.4.3, we have that  $\delta_1 \sqsubset \gamma_1^{\mathcal{C}}$ .

Let  $\Delta'' : [0, 1]^{S \times S} \rightarrow [0, 1]^{S \times S}$  be defined by

$$\Delta''(d)(s, t) = \begin{cases} 0 & \text{if } \gamma_1^{\mathcal{C}}(s, t) = 0 \\ \Delta_1(d)(s, t) & \text{otherwise} \end{cases}$$

Since  $\Delta_1$  is monotonic so is  $\Delta''$ , thus  $\Delta''$  admits a greatest fixed-point, say  $g$ . By  $\delta_1 \sqsubset \gamma_1^{\mathcal{C}}$  there exists  $s, t \in S$  such that  $\delta_1(s, t) < \gamma_1^{\mathcal{C}}(s, t)$ , so that  $\gamma_1^{\mathcal{C}}(s, t) \neq 0$ .

Suppose that

$$\{(s, t) \mid \gamma_1^{\mathcal{C}}(s, t) = 0\} = \sim.$$

Applying [14, Corollary 18],  $\Delta''$  has a unique fixed point which corresponds to  $\delta_1$ . By  $\gamma_1^{\mathcal{C}} = \Delta_1(\gamma_1^{\mathcal{C}})$ , we have that  $\gamma_1^{\mathcal{C}} = \Delta''(\gamma_1^{\mathcal{C}})$ , which contradicts the hypothesis that  $\delta_1 \neq \gamma_1^{\mathcal{C}}$ .

It can be shown that there exist  $s, t \in S$  such that  $\gamma_1^{\mathcal{C}}(s, t) \neq 0$ ,  $s \sim t$  (as proven above), and  $g(s, t) = 1$ .

Using Lemma 11.5.3 and  $\delta_1(s, t) = 0$ , we get that  $\gamma_1^{\mathcal{C}}(s, t) < 1$ . Hence,  $\gamma_1^{\mathcal{C}} \sqsubset g$ .

Let  $m$  and  $M$  be defined as follows

$$m = \max\{g(s, t) - \gamma_1^{\mathcal{C}}(s, t) \mid s, t \in S\}, \quad M = \{(s, t) \mid g(s, t) - \gamma_1^{\mathcal{C}}(s, t) = m\}.$$

Since  $\gamma_1^{\mathcal{C}} \sqsubset g$ ,  $m > 0$ .

We prove firstly two properties on  $M$ :

$$M \cap \{(s, t) \mid \ell(s) \neq \ell(t)\} = \emptyset \quad (11.5.3)$$

$$M \cap \{(s, t) \mid \gamma_1^{\mathcal{C}}(s, t) = 0\} = \emptyset \quad (11.5.4)$$

(11.5.3) follows since, for all  $\ell(u) \neq \ell(v)$ ,  $\gamma_1^{\mathcal{C}}(u, v) = 1 = g(u, v)$ , and  $m > 0$ .

(11.5.4) follows by definition of  $\Delta''$  and  $m > 0$ .

Let  $(s, t) \in M$ , then

$$\begin{aligned} m &= g(s, t) - \gamma_1^{\mathcal{C}}(s, t) \\ &= \Delta''(g)(s, t) - \Gamma_1^{\mathcal{C}}(\gamma_1^{\mathcal{C}})(s, t) \\ &= \Delta_1(g)(s, t) - \Gamma_1^{\mathcal{C}}(\gamma_1^{\mathcal{C}})(s, t) && \text{(by (11.5.4))} \\ &= \left( \min_{\omega' \in \pi(s, \cdot) \otimes \pi(t, \cdot)} \sum_{u, v \in S} g(u, v) \cdot \omega'(u, v) \right) - \sum_{u, v \in S} \gamma_1^{\mathcal{C}}(u, v) \cdot \omega((s, t), (u, v)) \\ &\leq \sum_{u, v \in S} g(u, v) \cdot \omega((s, t), (u, v)) - \sum_{u, v \in S} \gamma_1^{\mathcal{C}}(u, v) \cdot \omega((s, t), (u, v)) \\ &= \sum_{u, v \in S} (g(u, v) - \gamma_1^{\mathcal{C}}(u, v)) \cdot \omega((s, t), (u, v)). \end{aligned}$$

Since, for all  $u, v \in S$ ,

$$g(u, v) - \gamma_1^{\mathcal{C}}(u, v) \leq m$$

and

$$\sum_{u, v \in S} \omega((s, t), (u, v)) = 1,$$

we have that, whenever

$$\omega((s, t), (u, v)) > 0$$

guarantees

$$g(u, v) - \gamma_1^{\mathcal{C}}(u, v) = m.$$

Thus,  $\omega$  has the support contained in  $M$ . This means that, for all  $(s, t) \in M$ ,  $R_{\mathcal{C}}((s, t)) \subseteq M$ . Hence, using (11.5.3), we get that  $\gamma_1^{\mathcal{C}}(s, t) = 0$ , which contradicts (11.5.4).  $\square$

With these results in hand, we proceed now to give a sufficient and necessary condition for termination.

**Lemma 11.5.5.** *Let  $\mathcal{C}$  be a coupling for  $\mathcal{M}$ . For any  $\lambda \in (0, 1]$ , if  $\gamma_\lambda^{\mathcal{C}} \neq \delta_\lambda$ , then there exist  $s, t \in S$  and a coupling  $\mathcal{D} = \mathcal{C}[(s, t)/\omega']$  for  $\mathcal{M}$  such that*

$$\Gamma_\lambda^{\mathcal{D}}(\gamma_\lambda^{\mathcal{C}})(s, t) < \gamma_\lambda^{\mathcal{C}}(s, t).$$

*Proof.* We proceed by contraposition.

Suppose that for all  $s, t \in S$  and for all couplings  $\mathcal{D} = \mathcal{C}[(s, t)/\omega']$ ,

$$\Gamma_\lambda^{\mathcal{D}}(\gamma_\lambda^{\mathcal{C}})(s, t) \geq \gamma_\lambda^{\mathcal{C}}(s, t).$$

This means that  $\gamma_\lambda^{\mathcal{C}} = \Delta_\lambda(\gamma_\lambda^{\mathcal{C}})$ . Then the result follows from Lemma 11.5.4.  $\square$

The above result ensures that, unless  $\mathcal{C}$  is optimal w.r.t  $\triangleleft_\lambda$ , the hypothesis of Lemma 11.5.2 are satisfied, so that, we can further improve  $\mathcal{C}$  as aforesaid.

The next statement proves that this search strategy is correct.

**Theorem 11.5.6.** *Let  $\mathcal{C}$  be a coupling for  $\mathcal{M}$ . Then,  $\delta_\lambda = \gamma_\lambda^{\mathcal{C}}$  iff there is no coupling  $\mathcal{D}$  for  $\mathcal{M}$  such that*

$$\Gamma_\lambda^{\mathcal{D}}(\gamma_\lambda^{\mathcal{C}}) \sqsubset \gamma_\lambda^{\mathcal{C}}.$$

*Proof.* We prove that:

$$\delta_\lambda \neq \gamma_\lambda^{\mathcal{C}} \text{ iff there exists } \mathcal{D} \text{ such that } \Gamma_\lambda^{\mathcal{D}}(\gamma_\lambda^{\mathcal{C}}) \sqsubset \gamma_\lambda^{\mathcal{C}}.$$

( $\Rightarrow$ ) Assume  $\delta_\lambda \neq \gamma_\lambda^{\mathcal{C}}$ . By Lemma 14.5.3, there exist  $s, t \in S$  and  $\omega' \in \pi(s, \cdot) \otimes \pi(t, \cdot)$  such that

$$\lambda \cdot \sum_{u, v \in S} \gamma_\lambda^{\mathcal{C}}(u, v) \cdot \omega'(u, v) < \gamma_\lambda^{\mathcal{C}}(s, t).$$

As in the proof of Lemma 11.5.2, we have that  $\mathcal{D} = \mathcal{C}[(s, t)/\omega']$  satisfies  $\Gamma_\lambda^{\mathcal{D}}(\gamma_\lambda^{\mathcal{C}}) \sqsubset \gamma_\lambda^{\mathcal{C}}$ .

( $\Leftarrow$ ) Let  $\mathcal{D}$  be such that  $\Gamma_\lambda^{\mathcal{D}}(\gamma_\lambda^{\mathcal{C}}) \sqsubset \gamma_\lambda^{\mathcal{C}}$ . By Tarski's fixed point theorem,  $\gamma_\lambda^{\mathcal{D}} \sqsubset \gamma_\lambda^{\mathcal{C}}$ . Applying Theorem 14.4.3,  $\delta_\lambda \sqsubset \gamma_\lambda^{\mathcal{D}} \sqsubset \gamma_\lambda^{\mathcal{C}}$ , therefore,  $\delta_\lambda \neq \gamma_\lambda^{\mathcal{C}}$ .  $\square$

**Remark 11.5.7.** Note that, in general, there could be an infinite number of couplings for a given MC.

However, for each fixed  $d \in [0, 1]^{S \times S}$ , the linear function mapping  $\omega((s, t), \cdot)$  to  $\lambda \sum_{u, v \in S} d(u, v) \cdot \omega((s, t), (u, v))$  achieves its minimum at some vertex in the transportation polytope  $\pi(s, \cdot) \otimes \pi(t, \cdot)$ .

Since the number of such vertices are finite, using the optimal TP schedule for the update, ensures that the search strategy is always terminating.  $\blacksquare$

## 11.6 The On-the-Fly Algorithm for Exact Computation

In this section we provide an on-the-fly algorithm for exact computation of the bisimilarity distance  $\delta_\lambda$  for generic discount factors, by making full use of the greedy strategy presented in Section 14.5.2.

Let  $Q \subseteq S \times S$ . Assume that we want to compute  $\delta_\lambda(s, t)$ , for all  $(s, t) \in Q$ . The method proposed in Section 14.5.2 has the following key features:

1. the improvement of each coupling  $\mathcal{C}$  is obtained by a *local* update of its transition function at some state  $(u, v)$  in  $\mathcal{C}$ ;
2. the strategy does not depend on the choice of the state  $(u, v)$ ;
3. whenever a coupling  $\mathcal{C}$  is considered, the over-approximation  $\gamma_\lambda^{\mathcal{C}}$  of the distance can be computed by solving a system of linear equations.

Among them, only the last one requires a visit of the coupling. However, as noticed in Remark 11.5.1, the value  $\gamma_\lambda^{\mathcal{C}}(s, t)$  can be computed without considering the entire linear system of Equation (11.5.2), but only its smallest independent subsystem containing the variable  $x_{(s,t)}$ , which is obtained by restricting on the variables  $x_{(u,v)}$  such that  $(u, v) \in R_{\mathcal{C}}((s, t))$ . This subsystem can be further reduced, by Gaussian elimination, when some values for  $\delta_\lambda$  are known. The last observation suggests that, in order to compute  $\gamma_\lambda^{\mathcal{C}}(s, t)$ , we do not need to store the entire coupling, but it can be constructed on-the-fly.

The exact computation of the bisimilarity pseudometric is implemented by Algorithm 1. It takes as input a finite MC  $\mathcal{M} = (S, A, \pi, \ell)$ , a discount factor  $\lambda$ , and a query set  $Q$ .

We assume the following global variables to store:

- $\mathcal{C}$ , the current partial coupling;
- $d$ , the  $\lambda$ -discrepancy associated with  $\mathcal{C}$ ;
- *ToCompute*, the pairs of states for which the distance has to be computed;
- *Exact*, the pairs of states  $(s, t)$  such that  $d(s, t) = \delta_\lambda(s, t)$ ;
- *Visited*, the states of  $\mathcal{C}$  considered so far.

At the beginning (line 1) both the coupling  $\mathcal{C}$  and the discrepancy  $d$  are empty, there are no visited states, and no exact computed distances.

While there are still pairs left to be computed (line 2), we pick one (line 4), say  $(s, t)$ .

According to the definition of  $\delta_\lambda$ , if  $\ell(s) \neq \ell(t)$  then  $\delta_\lambda(s, t) = 1$ ; if  $s = t$  then  $\delta_\lambda(s, t) = 0$ , so that,  $d(s, t)$  is set accordingly, and  $(s, t)$  is added to *Exact* (lines 4–7).

Otherwise, if  $(s, t)$  was not previously visited, a matching  $\omega \in \pi(s, \cdot) \otimes \pi(t, \cdot)$  is guessed, and the routine *SetPair* updates the coupling  $\mathcal{C}$  at  $(s, t)$  with  $\omega$  (line 9), then the routine *Discrepancy* updates  $d$  with the  $\lambda$ -discrepancy associated with  $\mathcal{C}$  (line 10).

**Algorithm 1** On-the-Fly Bisimilarity Pseudometric**Input:** MC  $\mathcal{M} = (S, A, \pi, \ell)$ ; discount factor  $\lambda \in (0, 1)$ ; query  $Q \subseteq S \times S$ .

1.  $\mathcal{C} \leftarrow \text{empty}$ ;  $d \leftarrow \text{empty}$ ;  $Visited \leftarrow \emptyset$ ;  $Exact \leftarrow \emptyset$ ;  $ToCompute \leftarrow Q$ ;     —Init.—
2. **while**  $ToCompute \neq \emptyset$  **do**
3.   pick  $(s, t) \in ToCompute$
4.   **if**  $\ell(s) \neq \ell(t)$  **then**
5.      $d(s, t) \leftarrow 1$ ;  $Exact \leftarrow Exact \cup \{(s, t)\}$ ;  $Visited \leftarrow Visited \cup \{(s, t)\}$
6.   **else if**  $s = t$  **then**
7.      $d(s, t) \leftarrow 0$ ;  $Exact \leftarrow Exact \cup \{(s, t)\}$ ;  $Visited \leftarrow Visited \cup \{(s, t)\}$
8.   **else**     —if  $(s, t)$  is nontrivial—
9.     **if**  $(s, t) \notin Visited$  **then** pick  $\omega \in \pi(s, \cdot) \otimes \pi(t, \cdot)$ ;  $SetPair(\mathcal{M}, (s, t), \omega)$
10.      $Discrepancy(\lambda, (s, t))$      —update  $d$  as the  $\lambda$ -discrepancy for  $\mathcal{C}$ —
11.     **while**  $\exists (u, v) \in R_{\mathcal{C}}((s, t))$ .  $\mathcal{C}[(u, v)]$  not opt. for  $TP(d, \pi(u, \cdot), \pi(v, \cdot))$  **do**
12.        $\omega \leftarrow$  optimal schedule for  $TP(d, \pi(u, \cdot), \pi(v, \cdot))$
13.        $SetPair(\mathcal{M}, (u, v), \omega)$      —improve the current coupling—
14.        $Discrepancy(\lambda, (s, t))$      —update  $d$  as the  $\lambda$ -discrepancy for  $\mathcal{C}$ —
15.     **end while**
16.      $Exact \leftarrow Exact \cup R_{\mathcal{C}}((s, t))$      —add new exact distances—
17.     remove from  $\mathcal{C}$  all edges exiting from nodes in  $Exact$
18.   **end if**
19.    $ToCompute \leftarrow ToCompute \setminus Exact$      —remove exactly computed pairs—
20. **end while**
21. **return**  $d|_Q$      —return the distance for all pairs in  $Q$ —

According to the greedy strategy,  $\mathcal{C}$  is successively improved and  $d$  is consequently updated, until no further improvements are possible (lines 11–15). Each improvement is demanded by the existence of a better schedule for  $TP(d, \pi(u, \cdot), \pi(v, \cdot))$  (line 11). Note that, each improvement actually affects the current value of  $d(s, t)$ . This is done by restricting our attention only to the pairs that are reachable from  $(s, t)$  in  $\mathcal{G}(\mathcal{C})$ .

It is worth to note that  $\mathcal{C}$  is constantly updated, hence  $R_{\mathcal{C}}((s, t))$  may differ from one iteration to another.

When line 16 is reached, for each  $(u, v) \in R_{\mathcal{C}}((s, t))$ , we are guaranteed that  $d(u, v) = \delta_{\lambda}(s, t)$ , therefore  $R_{\mathcal{C}}((s, t))$  is added to  $Exact$ , and these values can be used in successive computations, so the edges exiting from these states are removed from  $\mathcal{G}(\mathcal{C})$ .

In line 19, the exact pairs computed so far are removed from  $ToCompute$ . Finally, if no more pairs need be considered, the exact distance on  $Q$  is returned (line 21).

Algorithm 1 calls the subroutines  $SetPair$  and  $Discrepancy$ , respectively, to construct or update the coupling  $\mathcal{C}$ , and to update the current over-approximation  $d$  during the computation. Now we explain how they work.

$SetPair$  (Algorithm 2) takes as input an MC  $\mathcal{M} = (S, A, \pi, \ell)$ , a pair of states  $(s, t)$ , and a matching  $\omega \in \pi(s, \cdot) \otimes \pi(t, \cdot)$ .

**Algorithm 2** *SetPair*( $\mathcal{M}, (s, t), \omega$ )**Input:** MC  $\mathcal{M} = (S, A, \pi, \ell)$ ;  $s, t \in S$ ;  $\omega \in \pi(s, \cdot) \otimes \pi(t, \cdot)$ 

- 
1.  $\mathcal{C}[(s, t)] \leftarrow \omega$  —update the coupling at  $(s, t)$  with  $\omega$ —
  2.  $Visited \leftarrow Visited \cup \{(s, t)\}$  —set  $(s, t)$  as visited—
  3. **for all**  $(u, v) \in \{(u', v') \mid \omega(u', v') > 0\} \setminus Visited$  **do** —for all demanded pairs—
  4.    $Visited \leftarrow Visited \cup \{(u, v)\}$
  5.   **if**  $u = v$  **then**  $d(u, v) \leftarrow 0$ ;  $Exact \leftarrow Exact \cup \{(u, v)\}$ ;
  6.   **if**  $\ell(u) \neq \ell(v)$  **then**  $d(u, v) \leftarrow 1$ ;  $Exact \leftarrow Exact \cup \{(u, v)\}$ ;
  7.   // propagate the construction
  8.   **if**  $(u, v) \notin Exact$  **then**
  9.     pick  $\omega' \in \pi(u, \cdot) \otimes \pi(v, \cdot)$  —guess a matching—
  10.     *SetPair*( $\mathcal{M}, (u, v), \omega'$ )
  11.   **end if**
  12. **end for**
- 

In lines 1–2 the transition function of the coupling  $\mathcal{C}$  is set to  $\omega$  at  $(s, t)$ , then  $(s, t)$  is added to *Visited*.

The on-the-fly construction of the coupling is recursively propagated to the successors of  $(s, t)$  in  $\mathcal{G}(\mathcal{C})$ .

During this construction, if some states with trivial distances are encountered,  $d$  and *Exact* are updated accordingly (lines 5–6).

*Discrepancy* (Algorithm 3) takes as input a discount factor  $\lambda$  and a pair of states  $(s, t)$ . It constructs the smallest (reduced) independent subsystem of Equation 11.5.2 having the variable  $x_{(s,t)}$  (lines 9–10).

As noticed in Remark 11.5.1, the least solution is computed by fixing  $d$  to zero for all the pairs which cannot be reached from any pair in *Exact* and such that its distance is greater than zero (lines 5–7).

Then, the discrepancy is computed and  $d$  is consequently updated.

Next, we present a simple example of how Algorithm 1 computes, showing the main features of our method:

1. the on-the-fly construction of the (partial) coupling, and
2. the restriction only to those variables which are demanded for the solution of the system of linear equations.

**Example 11.6.1** (On-the-fly computation). *Consider the undiscounted distance between states 1 and 4 for the {white, gray}-labeled MC depicted in Figure 11.1.*

*Algorithm 1 guesses an initial coupling  $\mathcal{C}_0$  with transition distribution  $\omega_0$ . This is done considering only the pairs of states which are needed: starting from  $(1, 4)$ , the distribution  $\omega_0((1, 4), \cdot)$  is guessed as in Figure 11.1, which demands for the exploration of  $(3, 4)$  and a guess  $\omega_0((3, 4), \cdot)$ .*

**Algorithm 3** *Discrepancy*( $\lambda, (s, t)$ )**Input:** discount factor  $\lambda \in (0, 1]$ ;  $s, t \in S$ 

1.  $Nonzero \leftarrow \emptyset$  —detect non-zero variables—
2. **for all**  $(u, v) \in R_C((s, t)) \cap Exact$  **such that**  $d(u, v) > 0$  **do**
3.    $Nonzero \leftarrow Nonzero \cup \{(u', v') \mid (u, v) \rightsquigarrow (u', v') \text{ in } \mathcal{G}^{-1}(C)\}$
4. **end for**
5. **for all**  $(u, v) \in R_C((s, t)) \setminus Nonzero$  **do** —set distance to zero—
6.    $d(u, v) \leftarrow 0$ ;  $Exact \leftarrow Exact \cup \{(u, v)\}$
7. **end for**
8. // construct the reduced linear system over nonzero variables
9.  $\vec{A} \leftarrow (C[(u, v)](u', v'))_{(u, v), (u', v') \in Nonzero}$
10.  $\vec{b} \leftarrow (\sum_{(u', v') \in Exact} d(u', v') \cdot C[(u, v)](u', v'))_{(u, v) \in Nonzero}$
11.  $\vec{x} \leftarrow$  solve  $\vec{x} = \lambda \vec{A} \vec{x} + \lambda \vec{b}$  —solve the reduced linear system—
12. **for all**  $(u, v) \in Nonzero$  **do** —update distances—
13.    $d(u, v) \leftarrow \vec{x}_{(u, v)}$
14. **end for**

Since no other pairs are demanded, the construction of  $C_0$  terminates. This gives the equation system:

$$\begin{cases} x_{1,4} = \frac{1}{3} \cdot \overbrace{x_{1,2}}^{=1} + \frac{1}{3} \cdot \overbrace{x_{2,3}}^{=1} + \frac{1}{6} \cdot x_{3,4} + \frac{1}{6} \cdot \overbrace{x_{3,6}}^{=1} = \frac{1}{6} \cdot x_{3,4} + \frac{5}{6} \\ x_{3,4} = \frac{1}{3} \cdot \overbrace{x_{1,2}}^{=1} + \frac{1}{6} \cdot \overbrace{x_{2,2}}^{=0} + \frac{1}{6} \cdot \overbrace{x_{2,3}}^{=1} + \frac{1}{3} \cdot \overbrace{x_{3,3}}^{=0} = \frac{1}{2}. \end{cases}$$

Note that the only variables appearing in the above equation system correspond to the pairs which have been considered so far. The least solution for it is given by  $d^{C_0}(1, 4) = \frac{11}{12}$  and  $d^{C_0}(3, 4) = \frac{1}{2}$ .

Now, these solutions are taken as the costs of a TP, from which we get an optimal transportation schedule  $\omega_1((1, 4), \cdot)$  improving  $\omega_0((1, 4), \cdot)$ .

The distribution  $\omega_1$  is used to update  $C_0$  to  $C_1 = C_0[(1, 4)/\omega_1]$  (depicted in Figure 11.1), obtaining the following new equation system:

$$x_{1,4} = \frac{1}{3} \cdot \overbrace{x_{2,2}}^{=0} + \frac{1}{3} \cdot \overbrace{x_{3,3}}^{=0} + \frac{1}{6} \cdot x_{1,4} + \frac{1}{6} \cdot \overbrace{x_{1,6}}^{=1} = \frac{1}{6} \cdot x_{1,4} + \frac{1}{6},$$

which has  $d^{C_1}(1, 4) = \frac{1}{5}$  as least solution.

Note that,  $(3, 4)$  is no more demanded, thus we do not need to update it.

Running again the TP on the improved over-approximation  $d^{C_1}$ , we discover that the coupling  $C_1$  cannot be further improved, hence we stop the computation, returning  $\delta_1(1, 4) = d^{C_1}(1, 4) = \frac{1}{5}$ .

It is worth noticing that Algorithm 1 does not explore the entire MC, not even all the reachable states from 1 and 4. The only edges in the MC which have been considered during the

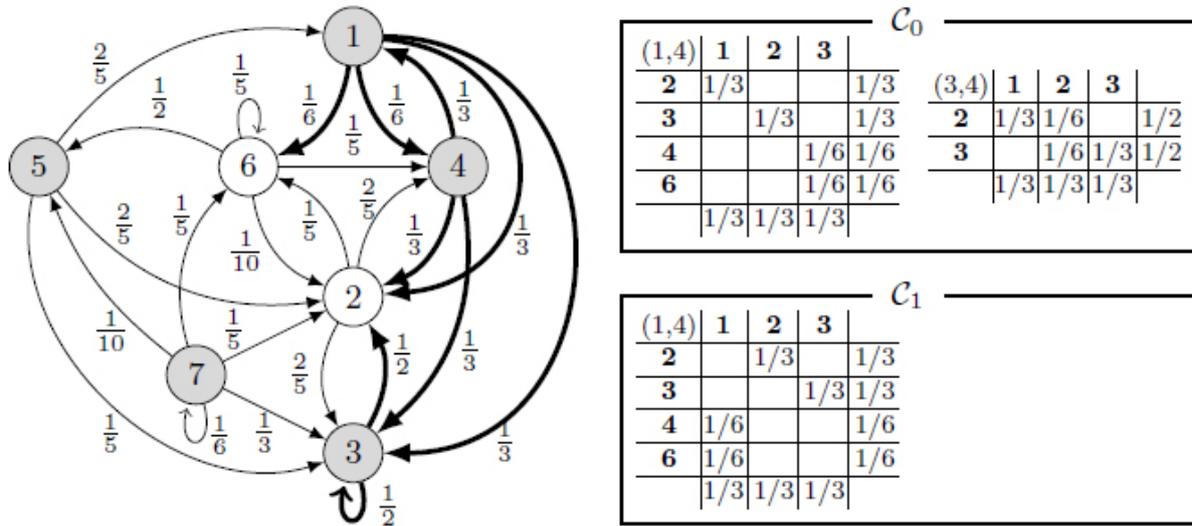


Figure 11.1: Execution trace for the computation of  $\delta_1(1,4)$  (details in Example 11.6.1).

computation are highlighted in Figure 11.1. ■

**Remark 11.6.2.** Notably, Algorithm 1 can also be used for computing over-approximated distances.

Indeed, assuming over-estimates for some particular distances are already known, they can be taken as inputs and used in our algorithm simply storing them in the variable  $d$  and treated as “exact” values. In this way, our method will return the least over-approximation of the distance agreeing with the given over-estimates.

This modification of the algorithm can be used to further decrease the exploration of the MC. Moreover, it can be employed in combination with other existing approximated algorithms, having the advantage of an on-the-fly state space exploration. ■

## 11.7 Experimental Results

In this section, we evaluate the performances of the on-the-fly algorithm on a collection of randomly generated MCs<sup>1</sup>. The detailed of these experiments can be seen in Table 11.3.

First, we compare the execution times of the on-the-fly algorithm with those of the iterative method proposed in [14] in the discounted case. Since the iterative method only allows for the computation of the distance for all state pairs at once, the comparison

<sup>1</sup> The tests have been made using a prototype implementation coded in Mathematica<sup>®</sup> (available at [http://people.cs.aau.dk/~mardare/projects/tools/mc\\_dist.zip](http://people.cs.aau.dk/~mardare/projects/tools/mc_dist.zip)) running on an Intel Core-i7 3.4 GHz processor with 12GB of RAM.

# States	On-the-Fly (exact)		Iterating (approximated)			Approximation
	Time (s)	# TPs	Time (s)	# Iterations	# TPs	Error
5	0.019675	1.19167	0.0389417	1.73333	26.7333	0.139107
6	0.05954	3.04667	0.09272	1.82667	38.1333	0.145729
7	0.13805	6.01111	0.204789	2.19444	61.7278	0.122683
8	0.255067	8.5619	0.364019	2.30476	83.0286	0.11708
9	0.499983	12.0417	0.673275	2.57917	114.729	0.111104
10	1.00313	18.7333	1.27294	3.11111	174.363	0.0946047
11	2.15989	25.9733	2.66169	3.55667	239.557	0.0959714
12	4.64225	34.797	5.52232	4.04242	318.606	0.0865612
13	6.73513	39.9582	8.06186	4.63344	421.675	0.0977743
14	6.33637	38.0048	7.18807	4.91429	593.981	0.118971
17	11.2615	47.0143	12.8048	5.88571	908.61	0.13213
19	26.6355	61.1714	29.6542	6.9619	1328.6	0.14013
20	34.379	66.4571	38.2058	7.5381	1597.92	0.142834

Table 11.1: Comparison between the on-the-fly algorithm and the iterative method.

is (in fairness) made with respect to runs of our on-the-fly algorithm with input query being the set of all state pairs. For each input instance, the comparison involves the following steps:

1. We run the on-the-fly algorithm, storing both execution time and the number of solved transportation problems,
2. Then, on the same instance, we execute the iterative method until the running time exceeds that of step 1. We report the execution time, the number of iterations, and the number of solved transportation problems.
3. Finally, we calculate the approximation error between the exact solution  $\delta_\lambda$  computed by our method at step 1 and the approximate result  $d$  obtained in step 2 by the iterative method, as  $\max_{s,t \in S} \delta_\lambda(s,t) - d(s,t)$ .

This has been made on a collection of MCs varying from 5 to 20 states.

For each  $n = 5, \dots, 20$ , we have considered 80 randomly generated MCs per out-degree, varying from 2 to  $n$ . Table 11.1 reports the average results of the comparison while Table 11.3 contains detailed data of all these experiments.

As can be seen, our use of a greedy strategy in the construction of the couplings leads to a significant improvement in the performances. We are able to compute the exact solution before the iterative method can under-approximate it with an error of  $\approx 0.1$ , which is a considerable error for a value in  $[0, 1]$ .

So far, we only examined the case when the on-the-fly algorithm is run on all state pairs at once. Now, we show how the performance of our method is improved even further when the distance is computed only for single pairs of states.

Table 11.2 shows the average execution times and number of solved transportation problems for (nontrivial) single-pair queries for randomly generated of MCs with num-

# States	out-degree = 3		$2 \leq \text{out-degree} \leq \# \text{ States}^*$	
	Time (s)	# TPs	Time (s)	# TPs
5	0.00594318	0.272727	0.011654	0.657012
6	0.0115532	0.548936	0.0304482	1.66696
7	0.0168408	0.980892	0.0884878	3.67706
8	0.0247971	1.34606	0.164227	5.30112
9	0.0259426	1.29074	0.394543	8.16919
10	0.0583405	2.03887	1.1124	13.0961
11	0.0766988	1.82706	2.22016	18.7228
12	0.0428891	1.62038	4.94045	26.0965
13	0.06082	1.88134	10.3606	35.1738
14	0.0894778	2.79441	20.1233	46.0775
20	0.35631	6.36833	1.5266	13.1367
30	4.66113	17.3167	74.8146	76.2642
50	27.2147	30.8217	2234.54	225.394

Table 11.2: Average performances of the on-the-fly algorithm on single-pair queries. In the first to columns the out-degree is 3; in the last two columns, the out-degree varies from 2 to # States. (\*) For 20, 30 and 50 states, out-degree is 4;

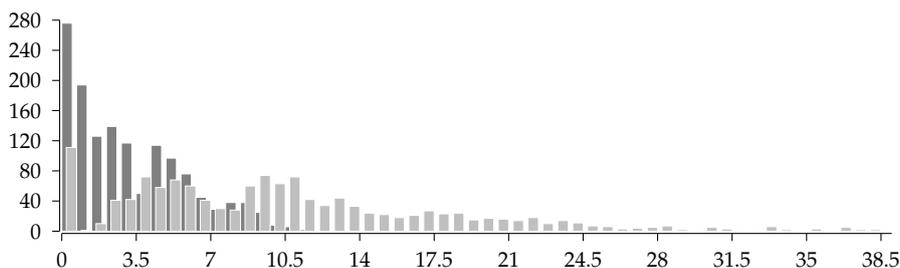


Figure 11.2: Distribution of the execution times (in seconds) for 1332 tests on randomly generated MCs with 14 states, out-degree 6 (darkest) and 8 (lightest).

ber of states varying from 5 to 50. In the first two columns we consider MCs with out-degree equal to 3, while the last two columns show the average values for out-degrees varying from 2 to the number of states of the MCs.

The results show that, when the out-degree of the MCs is low, our algorithm performs orders of magnitude better than in the general case. This is illustrated in Figure 11.2, where the distributions of the execution times for out-degree 6 and 8 are juxtaposed, in the case of MCs with 14 states. Each bar in the histogram represents the number of tests that terminate within the time interval indicated in the  $x$ -axis.

Notably, our method may perform better on large queries than on single-pairs queries. This is due to the fact that, although the returned value does not depend on the order the queried pairs are considered, a different order may speed up the performances. When the algorithm is run on more than a single pair, the way they are picked may increase the performances (e.g., compare the execution times in Tables 11.1 and 11.2 for MCs

with 14 states).

# States	out-degree	Time (s)	# TPs	# States	out-degree	Time (s)	# TPs
5	2	0.00400676	0.108108	12	2	0.0154774	0.455709
	3	0.00594318	0.272727		3	0.0428891	1.62038
	4	0.0139438	0.89375		4	0.250268	6.65647
	5	0.0219477	1.30233		5	0.642051	13.3423
6	2	0.00418537	0.0829268		6	2.76465	25.1563
	3	0.0115532	0.548936		7	2.35534	23.0476
	4	0.0181834	1.31878		8	3.33388	28.0218
	5	0.0503963	3.01843		9	8.58611	40.2267
	6	0.0659359	3.26496		10	9.53899	41.2665
7	2	0.00381973	0.108844		11	8.56025	43.165
	3	0.0168408	0.980892		12	18.5177	64.8665
	4	0.045375	2.64329		13	2	0.011122
	5	0.0953065	4.56774	3		0.06082	1.88134
	6	0.171368	6.73054	4		0.573837	8.97117
7	0.190202	6.6859	5	0.964231		14.4766	
8	2	0.00492955	0.0863636	6		2.78855	23.8415
	3	0.0247971	1.34606	7		6.00371	34.8411
	4	0.0574707	3.01171	8		9.11574	39.7318
	5	0.127983	5.57985	9		10.9838	51.6655
	6	0.194809	7.02206	10		15.0667	49.8645
	7	0.316305	9.49161	11		24.1127	65.2042
	8	0.426304	10.7912	12		22.8915	66.0113
	9	2	0.00858527	0.292636		13	32.8509
3		0.0259426	1.29074	14		2	0.0116639
4		0.0742248	3.33333		3	0.0894778	2.79441
5		0.17075	5.59023		4	0.549372	10.6123
6		0.448721	11.3662		5	1.61925	17.9549
7		0.554762	12.2854		6	3.19621	24.761
8		0.745632	13.875		7	7.1762	42.2557
9		1.11997	17.1407		8	10.674	50.3776
10		2	0.00921778		0.201481	9	20.8513
	3	0.0583405	2.03887		10	29.753	72.7825
	4	0.154308	4.8458		11	42.3095	85.4986
	5	0.298055	7.61401		12	56.4491	90.0443
	6	0.75986	13.313		13	75.946	99.7807
	7	0.899531	14.5403		20	2	0.0173417
	8	1.81921	22.0646	3		0.35631	6.36833
	9	2.62714	26.4177	4		1.5266	13.1367
	10	3.32188	26.5842	5		10.0704	38.8383
	11	2	0.0277172	0.495146		30	2
3		0.0766988	1.82706	3	4.66113		17.3167
4		0.299206	6.05806	4	74.8146		76.2642
5		0.72154	11.3061	50	2	0.093105	0.93
6		1.86622	17.7955		3	27.2147	30.8217
7		1.70401	17.1671		4	2234.54	225.394
8		3.19049	33.9219				
9		3.80195	31.5558				
10		4.52554	31.0671				
11		6.20838	36.8718				

Table 11.3: Detailed experimental results.

## 11.8 Conclusive Remarks

In this Chapter we have proposed an on-the-fly algorithm for computing exactly the bisimilarity distance between Markov chains, introduced by Desharnais et al. in [22].

Our algorithm represents an important improvement of the state of the art in this field where, before our contribution, the known tools were only concerned with computing approximations of the bisimilarity distances and they were, in general, based on iterative techniques. We demonstrate that, using on-the-fly techniques, we cannot only calculate exactly the bisimilarity distance, but the computation time is improved with orders of magnitude with respect to the corresponding iterative approaches. Moreover, our technique allows for the computation on a set of target distances that might be done by only investigating a significantly reduced set of states, and for further improvement of speed.

Our algorithm can be practically used to address a large spectrum of problems. For instance, it can be seen as a method to decide whether two states of a given MC are probabilistic bisimilar, to identify bisimilarity classes, or to solve lumpability problems. It is sufficiently robust to be used with approximation techniques as, for instance, to provide a least over-approximation of the behavioural distance given over-estimates of some particular distances. It can be integrated with other approximate algorithms, having the advantage of the efficient on-the-fly state space exploration.

Having a practically efficient tool to compute bisimilarity distances, opens the perspective of new applications. One of these is the state space reduction problem for MCs. Our technique can be used in this context, as an indicator for the sets of neighbor states that can be collapsed due to their similarity. It also provides a tool to estimate the difference between the initial MC and the reduced one, hence a tool for the approximation theory of Markov chains.

# Chapter 12

## Efficient Computation of Bisimilarity Distances

### 12.1 Introduction

As we have emphasized in the previous Chapters, probabilistic bisimulation of Larsen and Skou [32] is the standard equivalence for analyzing the behaviour of Markov chains. In [28], this notion has been extended to Markov Decision Processes with rewards (MDPs) with the intent of reducing the size of large systems to help the computation of optimal policies.

However, as in the case of MCs, also for MDPs, when the numerical values of probabilities are based on statistical sampling or subject to error estimates, any behavioral analysis based on a notion of equivalence is too fragile, as it only relates processes with identical behaviors. This is a common issue in applications such as systems biology [41], games [13], or planning [18].

Such problems motivated the study of *behavioral distances* (pseudometrics) for probabilistic systems, firstly developed for Markov chains [22, 42, 44] and later extended to MDPs [25]. These distances support approximate reasoning on probabilistic systems, providing a way to measure the behavioral similarity between states. They allow one to analyze models obtained as approximations of others, more accurate but less manageable, still ensuring that the obtained solution is close to the real one. For instance, in [11, 12] the pseudometric of [25] is used to compute (approximated) optimal policies for MDPs in applications for artificial intelligence. These arguments motivate the development of methods to efficiently compute behavioral distances also for MDPs.

In the previous Chapter we proposed an efficient on-the-fly algorithm for computing the behavioral pseudometrics of Desharnais et al. [22] on MCs. Our method has been inspired by an alternative characterization of the pseudometric given in [14], that relates the pseudometric to the least solutions of a set of equation systems induced by a collection transportation schedules. The pseudometric is computed by successive refinements of over-approximations of the actual distance using a greedy strategy that

always chooses a transportation schedule that better improves the current approximation. This strategy avoids the exhaustive exploration of the state space, and has the practical advantage that allows one to focus only on computing the distances between states that are of particular interest. Experimental results have shown that this technique performs, on average, orders of magnitude better than the corresponding iterative algorithms proposed in the literature, e.g., in [14].

This algorithm can be adapted, without difficulty, to compute the bisimilarity pseudometric introduced by Ferns et al. in [25] for MDPs with rewards. Being the similarity with the previous algorithm, we will not present the details in this monograph. However, an even more advanced algorithm for MDPs will be presented and discussed in the next Chapter.

In this Chapter we present the `BISIMDIST` library, composed of two Mathematica packages which implement our on-the-fly algorithm for computing the bisimilarity distances both for MCs and MDPs. `BISIMDIST` is available at <http://people.cs.aau.dk/~mardare/tools.htm> together with simple tutorials presenting use case examples that show all the features of the library.

## 12.2 The `BISIMDIST` Library

The `BISIMDIST` library consists of two Mathematica packages: `MCDIST` and `MDPDIST` providing data structures and primitives for creating, manipulating, and computing bisimilarity distances for MCs and MDPs respectively.

It also includes methods to identify bisimilarity classes and to solve lumpability problems.

**The `MCDIST` Package:** An MC with  $n$  states is represented as a term of the form

$$\text{MC}[\langle tm \rangle, \langle lbl \rangle],$$

where  $\langle tm \rangle$  is an  $n \times n$  probability transition matrix ( $\langle tm \rangle[[i, j]]$  denotes the probability of going from the state  $i$  to the state  $j$ ) and  $\langle lbl \rangle$  is a vector of strings of length  $n$  ( $\langle lbl \rangle[[i]]$  is the label associated with the state  $i$ ). Note that states are implicitly represented as indices  $1 \leq i \leq n$ .

The probability transition matrices can be defined explicitly as a matrix, or implicitly by listing only the transitions which have nonzero probability by means of the function `MCTm` (see Fig. 12.1).

Given a list `trRules` of rules of the form  $\{i, j\} \rightarrow p_{i,j}$ , the function `MCTm[trRules, n]` returns an  $n \times n$  matrix where each pair  $(i, j)$  is associated with the value  $p_{i,j}$ , otherwise 0.

An MC `mc` is displayed by calling `PlotMC[mc]`.

Given a sequence `mc1, ..., mck` of MCs, `JoinMC[mc1, ..., mck]` yields an MC representing their disjoint union. The indices representing the set of states are obtained shifting the indices of the states of the arguments according to their order in the sequence (e.g.

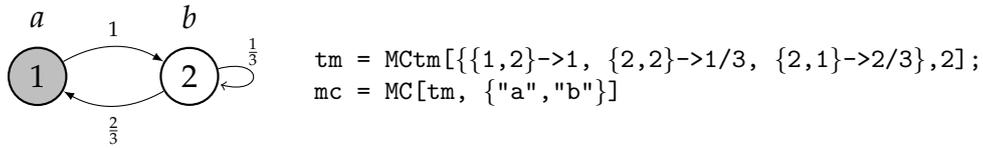


Figure 12.1: Encoding of a Markov Chain as a data term in BISIMDIST.

if  $mc_1$  has  $n$  states, the index corresponding to the  $i$ -th state of  $mc_2$  in  $\text{JoinMC}[mc_1, mc_2]$  is  $n + i$ ).

Given an MC  $mc$  with  $n$  states, a list  $Qpairs$  of pairs of indices  $1 \leq i, j \leq n$ , and a rational discount factor  $\lambda \in (0, 1]$ ,  $\text{BDistMC}[mc, \lambda, Qpairs]$  returns the list of all  $\lambda$ -discounted bisimilarity distances calculated between the pairs of states in  $Qpairs$  as list of rules of the form  $\{i, j\} \rightarrow d_{i,j}$ . The alias `All` is used for indicating the list of all pairs of states. `BDistMC` has the following options:

**Verbose:** (default `False`) displays all intermediate approximations steps;

**ConsistencyCheck:** (default `True`) checks that the term  $mc$  is a proper MC;

**Estimates:** (default `None`) takes a list of rules of the form  $\{i, j\} \rightarrow d_{i,j}$  and computes the least over-approximation of the bisimilarity distance assuming  $d_{i,j}$  to be the actual distance between the states  $i$  and  $j$ .

The package `MCDIST` provides also the functions `BisimClassesMC`, which calculates the bisimilarity classes of an MC, and `BisimQuotientMC` that, for a given an MC, yields its quotient w.r.t. probabilistic bisimilarity.

**The MDPDIST Package:** An MDP with  $n$  states and  $m$  action labels is represented as a term of the form

$$\text{MPD}[\langle tm \rangle, \langle rw \rangle, \langle act \rangle],$$

where  $\langle tm \rangle$  is an  $n \times m \times n$  labelled probability transition matrix ( $\langle tm \rangle[[i, a, j]]$  is the probability of going from the state  $i$  to the state  $j$ , known that the action  $a$  as been chosen),  $\langle rw \rangle$  is a  $n \times m$  real-valued matrix representing a reward function, and  $\langle act \rangle$  is a string-valued list of length  $m$  specifying the names of the action labels. States and action labels are implicitly encoded as indices.

Probability transition matrices of size  $n \times m \times n$  can be defined by giving the nonzero transition probabilities as a list `trRules` of rules of the form  $\{i, a, j\} \rightarrow p_{i,a,j}$  and calling `MDPtm[trRules, n, m]`.

Analogously,  $n \times m$  reward matrices can be defined by calling

$$\text{MDPrm}[\langle rwRules \rangle, n, m],$$

where  $\langle rwRules \rangle$  is a list of rules of the form  $\{i, a\} \rightarrow r_{i,a}$ .

The `MDPDIST` package is provided with an interface similar to `MCDIST` with analogous semantics: `PlotMDP`, `JoinMDP`, `BDistMDP`, `BisimClassesMDP`, and `BisimQuotientMDP`.

Table 12.1: Comparison between the on-the-fly and the iterative methods on MCs.

# States	On-the-Fly (exact)	Iterative (approximated)		Approximation
	Time (sec)	Time (sec)	# Iterations	Error
10	1.003	1.272	3.111	0.0946
12	4.642	5.522	4.042	0.0865
14	6.336	7.188	4.914	0.1189
20	34.379	38.205	7.538	0.1428

### 12.3 Conclusive Remarks

BISIMDIST is a research tool still undergoing development. While not yet mature enough to handle industrial case studies, the on-the-fly algorithm for computing the bisimilarity distance performs, on average, better than the iterative method proposed in [14].

Table 12.1 reports the average execution times of the on-the-fly algorithm run with discount factor  $\lambda = 1/2$  on a collection of randomly generated MCs. We executed the iterative method on the same input instances, interrupting it as soon as it exceeded the running time of our method. The on-the-fly approach leads to a significant improvement in the performances: it yields the exact solution before the iterative method can under-approximate it with an error of  $\approx 0.1$ , which is a non-negligible error for a value in the interval  $[0, 1]$ .

The BISIMDIST library provides primitives that aid the analysis on probabilistic systems by reasoning in terms of approximate behaviors. In the next Chapter, reporting the results published in [4], we further improve the efficiency of the implemented on-the-fly algorithm on MDPs, also in relation to the addition of primitives for handling algebraic operations over probabilistic systems, such as synchronous/asynchronous parallel composition.

# Chapter 13

## Compositional Distance for Markov Decision Processes

### 13.1 Introduction

In the previous two Chapters we proposed and implemented an efficient on-the-fly method for the exact computation of behavioral pseudometrics for Markov chains and Markov decision processes with rewards. In this Chapter we will further extend this algorithm, for the case of Markov decision processes, by taking into account the compositionality and the modularity of systems.

Realistic models are usually specified compositionally by means of operators that describe the interactions between the subcomponents. These specifications may thus suffer from an exponential growth of the state space, e.g. the parallel composition of  $n$  subsystems with  $m$  states may cause the main system to have  $m^n$  states. To cope with this problem, algorithms like [14, 18, 25] that need to investigate the entire state space of the system are not sufficient; the same can be said about our algorithm, presented in the previous two Chapters for MCs and MDPs, which even if it is more efficient than the others because it avoids the entire state space exploration using on-the-fly techniques, it is not sufficiently efficient when compositionality is an issue.

Classically, the exact behavior of systems can be analyzed compositionally if the considered behavioral equivalence (e.g. bisimilarity) is a congruence w.r.t. the composition operators. When the behavior of processes is approximated by means of behavioral distances, congruence is generalized by the notion of *non-extensiveness* of the composition operators, that describes the relation between the distances of the subcomponents to that of the composite system [22].

In this Chapter we study to which extent compositionality on MDPs can be exploited in the computation of the behavioral pseudometrics of [25], hence how the compositional structure of processes can be used in an approximated analysis of behaviors. This research has been inspired by our experience with probabilistic and stochastic process

algebras and by our study of the interaction between the operational semantics of such an algebra and the behavioural pseudometrics, detailed in the first Part of this monograph.

To this end, we introduce a general notion of composition operator on MDPs and characterize a class of operators, called *safe*, that are guaranteed to be non-extensive. This class is shown to cover a wide range of known operators (e.g. synchronous and asynchronous parallel composition), moreover its defining property provides an easy systematic way to check non-extensiveness.

We provide an algorithm to compute the bisimilarity pseudometric by exploiting both the on-the-fly state space exploration in the spirit of the previous two Chapters [5], and the compositional structure of MDPs built over safe operators. Experimental results show that the compositional optimization yields a significant additional improvement on top of that obtained by the on-the-fly method. In the best cases, the exploitation of compositionality achieves a reduction of computation time by a factor of 10, and for least significant cases the reduction is that of a factor of 2.

## 13.2 Markov Decision Processes and Behavioral Metrics

In this section we recall the definitions of *finite discrete-time Markov Decision Process with rewards* (MDP), and of *bisimulation relation* on MDPs [28]. Then we recall the definition of *bisimilarity pseudometric* introduced in [25], which measures behavioral similarities between states.

**Basic notations:** Recall that a *probability distribution* over a finite set  $S$  is a function  $\mu: S \rightarrow [0, 1]$  such that

$$\sum_{s \in S} \mu(s) = 1.$$

We denote by  $\Delta(S)$  the set of probability distributions over  $S$ .

Given  $\mu, \nu \in \Delta(S)$ , a distribution  $\omega \in \Delta(S \times S)$  is a *matching* for  $(\mu, \nu)$  if for all  $u, v \in S$ ,

$$\sum_{s \in S} \omega(u, s) = \mu(u) \quad \text{and} \quad \sum_{s \in S} \omega(s, v) = \nu(v).$$

We denote by  $\Pi(\mu, \nu)$  the set of matchings for  $(\mu, \nu)$ .

For a (pseudo)metric  $d: S \times S \rightarrow [0, \infty)$  over a finite set  $S$ , the *Kantorovich (pseudo)metric* is defined by

$$\mathcal{T}_d(\mu, \nu) = \min_{\omega \in \Pi(\mu, \nu)} \sum_{u, v \in S} \omega(u, v) d(u, v),$$

for arbitrary  $\mu, \nu \in \Delta(S)$ <sup>1</sup>.

---

<sup>1</sup> Since  $S$  is finite,  $\Pi(\mu, \nu)$  describes a *bounded* transportation polytope [20], hence the minimum in the definition of  $\mathcal{T}_d(\mu, \nu)$  exists and can be achieved at some vertex.

**Definition 13.2.1** (Markov Decision Process). *A Markov Decision Process is a tuple*

$$\mathcal{M} = (S, A, \tau, \rho)$$

*consisting of a finite nonempty set  $S$  of states, a finite nonempty set  $A$  of actions, a transition function  $\tau: S \times A \rightarrow \Delta(S)$ , and a reward function  $\rho: S \times A \rightarrow \mathbb{R}^{\geq 0}$ .*

The operational behavior of an MDP  $\mathcal{M} = (S, A, \tau, \rho)$  is as follows.

- The process in the state  $s_0 \in S$  chooses nondeterministically an action  $a \in A$  and it changes the state to  $s_1 \in S$ , with probability  $\tau(s_0, a)(s_1)$ .
- The choice of  $a$  in  $s_0$  is rewarded by  $\rho(s_0, a)$ .
- The *executions* are transition sequences  $w = (s_0, a_0)(s_1, a_1) \dots$ ; the challenge is to find *strategies* for choosing the actions in order to maximize the reward

$$R_\lambda(w) = \lim_{n \rightarrow \infty} \sum_{i=0}^n \lambda^i \rho(s_i, a_i),$$

where  $\lambda \in (0, 1)$  is a *discount factor*.

- A *strategy* is given by a function  $\pi: S \rightarrow \Delta(A)$ , called *policy*, where  $\pi(s_0)(a)$  is the probability of choosing the action  $a$  at state  $s_0$ . Each policy  $\pi$  induces a probability distribution over executions defined, for an arbitrary  $w = (s_0, a_0)(s_1, a_1) \dots$ , by

$$P^\pi(w) = \lim_{n \rightarrow \infty} \prod_{i=0}^n \pi(s_i)(a_i) \cdot \tau(s_i, a_i)(s_{i+1}).$$

- The *value* of  $s \in S$  according to  $\pi$ , written  $V_\lambda^\pi(s)$ , is the expected value of  $R_\lambda$  w.r.t.  $P^\pi$  on the measurable cylinder set of the executions starting from  $s$ .
- The mapping  $V_\lambda^\pi: S \rightarrow \mathbb{R}^{\geq 0}$  is the *value function according to  $\pi$* . The value functions induce a preorder on policies defined by

$$\pi \preceq \pi' \text{ iff } V_\lambda^\pi(s) \leq V_\lambda^{\pi'}(s), \text{ for all } s \in S.$$

- A policy  $\pi^*$  is *optimal* for an MDP  $\mathcal{M}$  if it is maximal w.r.t.  $\preceq$  among all policies for  $\mathcal{M}$ .
- Given  $\mathcal{M}$ , there always exists an optimal policy  $\pi^*$ , but it might not be unique; it has a unique value function  $V_\lambda^{\pi^*}$  satisfying, for all  $s \in S$ , the following system of equations known as the *Bellman optimality equations*:

$$V_\lambda^{\pi^*}(s) = \max_{a \in A} (\rho(s, a) + \lambda \sum_{t \in S} \tau(s, a)(t) \cdot V_\lambda^{\pi^*}(t)).$$

The MDPs are equated by stochastic bisimilarity in order to describe similar behaviours. Givan et al. [28] investigated several notions of state equivalence and determined that the most appropriate one is the one defined as follows.

**Definition 13.2.2** (Stochastic Bisimulation). *Let  $\mathcal{M} = (S, A, \tau, \rho)$  be an MDP. An equivalence relation  $R \subseteq S \times S$  is a stochastic bisimulation if whenever  $(s, t) \in R$  then, for all  $a \in A$ ,*

- i.  $\rho(s, a) = \rho(t, a)$  and,
- ii. for all  $R$ -equivalence classes  $C$ ,  $\tau(s, a)(C) = \tau(t, a)(C)$ .

Two states  $s, t \in S$  are stochastic bisimilar, written  $s \sim_{\mathcal{M}} t$ , if they are related by some stochastic bisimulation on  $\mathcal{M}$ .

To cope with the problem of measuring how similar two MDPs are, Ferns et al. [25] defined a bisimilarity pseudometrics that measure the behavioural similarity of two non-bisimilar MDPs. This is defined as the least fixed point of a transformation operator on functions in  $[0, 1]^{S \times S}$ .

Let  $\mathcal{M} = (S, A, \tau, \rho)$  be an MDP and  $\lambda \in (0, 1)$  be a discount factor.

The set  $[0, 1]^{S \times S}$  of  $[0, \infty)$ -valued maps on  $S \times S$  equipped with the point-wise partial order defined by

$$d \sqsubseteq d' \text{ iff } d(s, t) \leq d'(s, t), \text{ for all } s, t \in S,$$

forms an  $\omega$ -complete partial order with bottom the constant zero-function  $\bar{0}$ , and greatest lower bound given by

$$\left( \prod_{i \in \mathcal{N}} d_i \right)(s, t) = \inf_{i \in \mathcal{N}} d_i(s, t), \text{ for all } s, t \in S.$$

We define a fixed point operator  $F_{\lambda}^{\mathcal{M}}$  on  $[0, 1]^{S \times S}$ , for  $d: S \times S \rightarrow [0, \infty)$  and  $s, t \in S$ , as follows:

$$F_{\lambda}^{\mathcal{M}}(d)(s, t) = \max_{a \in A} (|\rho(s, a) - \rho(t, a)| + \lambda \cdot \mathcal{T}_d(\tau(s, a), \tau(t, a))).$$

$F_{\lambda}^{\mathcal{M}}$  is monotonic [25], thus, by Tarski's fixed point theorem, it admits a least fixed point. This fixed point is the *bisimilarity pseudometric*.

**Definition 13.2.3** (Bisimilarity pseudometric). *Let  $\mathcal{M}$  be an MDP and  $\lambda \in (0, 1)$  be a discount factor, then the  $\lambda$ -discounted bisimilarity pseudometric for  $\mathcal{M}$ , denoted by  $\delta_{\lambda}^{\mathcal{M}}$ , is the least fixed point of  $F_{\lambda}^{\mathcal{M}}$ .*

The pseudometric  $\delta_{\lambda}^{\mathcal{M}}$  enjoys the property that two states are at zero distance if and only if they are bisimilar. Moreover, in [19] it has been proved, using Banach's fixed point theorem, that for  $\lambda \in (0, 1)$ ,  $F_{\lambda}^{\mathcal{M}}$  has a *unique* fixed point.

### 13.3 Compositional Operators for MDPs

In this section we give a general definition of a composition operator on MDPs that subsumes most of the known composition operators such as the synchronous, asynchronous, and CCS-like parallel compositions. We introduce the notion of *safeness* for an operator and prove that it implies non-extensiveness. Recall that, non-extensiveness corresponds to the quantitative analogue of congruence when one aims to reason with behavioral distances, as advocated e.g. in [22,27].

**Definition 13.3.1** (Composition Operator). Let  $\mathcal{M}_i = (S_i, A_i, \tau_i, \rho_i)$ ,  $i = 1..n$ , be MDPs. A composition operator on  $\mathcal{M}_1, \dots, \mathcal{M}_n$  is a tuple  $op = (A, op_\tau, op_\rho)$  consisting of a nonempty set  $A$  of actions and the following operations

- **on transitions functions:**  $op_\tau: \prod_{i=1}^n \Delta(S_i)^{S_i \times A_i} \rightarrow \Delta(S)^{S \times A}$ ,
- **on reward functions:**  $op_\rho: \prod_{i=1}^n \mathbb{R}^{S_i \times A_i} \rightarrow \mathbb{R}^{S \times A}$ .

where,  $S = \prod_{i=1}^n S_i$  denotes the cartesian product of  $S_i$ ,  $i = 1..n$ .

We denote by  $op(\mathcal{M}_1, \dots, \mathcal{M}_n)$  the composite MDP  $(S, A, op_\tau(\tau_1, \dots, \tau_n), op_\rho(\rho_1, \dots, \rho_n))$ .

Below we present two examples of the most known parallel composition operators for two fixed MDPs  $\mathcal{M}_X = (X, A_X, \tau_X, \rho_X)$  and  $\mathcal{M}_Y = (Y, A_Y, \tau_Y, \rho_Y)$ .

**Example 13.3.2. Synchronous Parallel Composition** can be given as a binary composition operator  $| = (A_X \cap A_Y, |\tau, |\rho)$ , where

$$(\tau_X |_\tau \tau_Y)((x, y), a)(u, v) = \tau_X(x, a)(u) \cdot \tau_Y(y, a)(v),$$

$$(\rho_X |_\rho \rho_Y)((x, y), a) = \rho_X(x, a) + \rho_Y(y, a).$$

The process  $\mathcal{M}_X | \mathcal{M}_Y$  reacts iff  $\mathcal{M}_X$  and  $\mathcal{M}_Y$  can react synchronously. Actions are rewarded by summing up the rewards of the components. ■

**Example 13.3.3. CCS-like Parallel Composition** can be defined by the composition operator  $\parallel = (A_X \cup A_Y, \parallel_\tau, \parallel_\rho)$ , where

$$(\tau_X \parallel_\tau \tau_Y)((x, y), a)(u, v) = \begin{cases} \tau_X(x, a)(u) & \text{if } a \notin A_Y \text{ and } v = y \\ \tau_Y(y, a)(v) & \text{if } a \notin A_X \text{ and } u = x \\ \tau_X(x, a)(u) \cdot \tau_Y(y, a)(v) & \text{if } a \in A_X \cap A_Y \\ 0 & \text{otherwise} \end{cases}$$

$$(\rho_X \parallel_\rho \rho_Y)((x, y), a) = \begin{cases} \rho_X(x, a) & \text{if } a \notin A_Y \\ \rho_Y(y, a) & \text{if } a \notin A_X \\ \rho_X(x, a) + \rho_Y(y, a) & \text{if } a \in A_X \cap A_Y \end{cases}$$

In the process  $\mathcal{M}_X \parallel \mathcal{M}_Y$ , the components synchronize on the same action, otherwise they proceed asynchronously. Asynchronous parallel composition can be defined as above, requiring that the MDPs have disjoint set of actions. ■

## 13.4 Non-Extensive Compositional Operators

To introduce the concept of non-extensiveness for a composition operator, that is central for this research, some additional notations are required.

Consider the sets  $X_i$ , the functions  $d_i: X_i \times X_i \rightarrow [0, \infty)$ , for  $i = 1..n$ , and  $p \in [1, \infty]$ . We define the  $p$ -norm function

$$\|d_1, \dots, d_n\|_p: \prod_{i=1}^n X_i \times \prod_{i=1}^n X_i \rightarrow [0, \infty)$$

as follows.

$$\|d_1, \dots, d_n\|_p((x_1, \dots, x_n), (y_1, \dots, y_n)) = (\sum_{i=1}^n d_i(x_i, y_i)^p)^{\frac{1}{p}} \quad \text{for } p < \infty,$$

$$\|d_1, \dots, d_n\|_\infty((x_1, \dots, x_n), (y_1, \dots, y_n)) = \max_{1 \leq i \leq n} d_i(x_i, y_i).$$

Note that, if  $(X_i, d_i)$  are (pseudo)metric spaces,  $\|d_1, \dots, d_n\|_p$  is a (pseudo)metric on  $\prod_{i=1}^n X_i$ , known in the literature as the  $p$ -product (pseudo)metric.

**Definition 13.4.1.** Let  $p \in [1, \infty]$ . A composition operator  $op$  on MDPs  $\mathcal{M}_1, \dots, \mathcal{M}_n$  is  $p$ -non-extensive if

$$\delta_\lambda^{op(\mathcal{M}_1, \dots, \mathcal{M}_n)} \sqsubseteq \|\delta_\lambda^{\mathcal{M}_1}, \dots, \delta_\lambda^{\mathcal{M}_n}\|_p.$$

A composition operator is non-extensive if it is  $p$ -non-extensive for some  $p$ .

Non-extensiveness for a composition operator ensures that bisimilarity is a congruence with respect to it — a direct consequence of Theorem 4.5 in [25].

**Lemma 13.4.2.** Let  $\mathcal{M}_i = (S_i, A_i, \tau_i, \rho_i)$ , for  $i = 1..n$ , be MDPs and  $s_i, t_i \in S_i$ . Assume that  $op$  is a  $p$ -non-extensive composition operator on  $\mathcal{M}_1, \dots, \mathcal{M}_n$ . Then,

$$i) \text{ if } p < \infty, \quad \delta_\lambda^{op(\mathcal{M}_1, \dots, \mathcal{M}_n)}((s_1, \dots, s_n), (t_1, \dots, t_n)) \leq \left( \sum_{i=1}^n \delta_\lambda^{\mathcal{M}_i}(s_i, t_i)^p \right)^{\frac{1}{p}}$$

$$ii) \text{ if } p = \infty, \quad \delta_\lambda^{op(\mathcal{M}_1, \dots, \mathcal{M}_n)}((s_1, \dots, s_n), (t_1, \dots, t_n)) \leq \max_{i=1}^n \delta_\lambda^{\mathcal{M}_i}(s_i, t_i).$$

*Proof.* Since  $s_i \sim_{\mathcal{M}_i} t_i$ , applying and Theorem 4.5 in [25], we obtain that  $\delta_\lambda^{\mathcal{M}_i}(s_i, t_i) = 0$ . Further, using the  $p$ -non-extensiveness of  $op$  and the definition of  $p$ -product metric, we obtain the desired results. □

**Corollary 13.4.3.** Let  $\mathcal{M}_i = (S_i, A_i, \tau_i, \rho_i)$ , for  $i = 1..n$ , be MDPs and  $s_i, t_i \in S_i$ . Assume that  $op$  is a  $p$ -non-extensive composition operator on  $\mathcal{M}_1, \dots, \mathcal{M}_n$ . If  $s_i \sim_{\mathcal{M}_i} t_i$  for all  $i = 1..n$ , then

$$(s_1, \dots, s_n) \sim_{op(\mathcal{M}_1, \dots, \mathcal{M}_n)} (t_1, \dots, t_n).$$

*Proof.* It follows by Lemma 13.4.2 and Theorem 4.5 in [25].  $\square$

In general, proving non-extensiveness for a composition operator on MDPs is not a simple task, since one needs to consider the pseudometrics  $\delta_\lambda^{\mathcal{M}_i}$  which are defined as the least fixed point of  $F_\lambda^{\mathcal{M}_i}$ . A simpler sufficient condition that ensures non-extensiveness is the *safeness*.

**Definition 13.4.4.** Let  $\mathcal{M}_i = (S_i, A_i, \tau_i, \rho_i)$ , for  $i = 1..n$ , be MDPs and  $p \in [1, \infty]$ . A composition operator  $op$  on  $\mathcal{M}_1, \dots, \mathcal{M}_n$  is  $p$ -safe if, for any  $d_i$  pseudometric on  $S_i$ , such that  $d_i \sqsubseteq F_\lambda^{\mathcal{M}_i}(d_i)$ , the following condition is satisfied

$$F_\lambda^{op(\mathcal{M}_1, \dots, \mathcal{M}_n)}(\|d_1, \dots, d_n\|_p) \sqsubseteq \|F_\lambda^{\mathcal{M}_1}(d_1), \dots, F_\lambda^{\mathcal{M}_n}(d_n)\|_p.$$

A composition operator on MDPs is safe if it is  $p$ -safe for some  $p \in [1, \infty]$ .

**Theorem 13.4.5.** Any safe composition operator on MDPs is non-extensive.

*Proof.* Assume  $op$  is a safe composition operator on the MDPs  $\mathcal{M}_1, \dots, \mathcal{M}_n$ , and let  $\mathcal{M} = op(\mathcal{M}_1, \dots, \mathcal{M}_n)$ .

Then, for some  $p \in [1, \infty]$  the following condition is satisfied

$$F_\lambda^{\mathcal{M}}(\|\delta_\lambda^{\mathcal{M}_1}, \dots, \delta_\lambda^{\mathcal{M}_n}\|_p) \sqsubseteq \|F_\lambda^{\mathcal{M}_1}(\delta_\lambda^{\mathcal{M}_1}), \dots, F_\lambda^{\mathcal{M}_n}(\delta_\lambda^{\mathcal{M}_n})\|_p.$$

By definition  $F_\lambda^{\mathcal{M}_1}(\delta_\lambda^{\mathcal{M}_1}) = \delta_\lambda^{\mathcal{M}_1}$ , therefore  $\|\delta_\lambda^{\mathcal{M}_1}, \dots, \delta_\lambda^{\mathcal{M}_n}\|_p$  is a prefix point of  $F_\lambda^{\mathcal{M}}$ . Applying Tarski's fixed point theorem it follows that

$$\delta_\lambda^{\mathcal{M}} \sqsubseteq \|\delta_\lambda^{\mathcal{M}_1}, \dots, \delta_\lambda^{\mathcal{M}_n}\|_p.$$

$\square$

The examples of compositional operators that we have presented in the previous section are all 1-safe, hence non-extensive.

**Proposition 13.4.6.** The composition operators of Examples 13.3.2–13.3.3 are 1-safe.

*Proof.* Let  $\mathcal{M}_X = (X, A_X, \tau_X, \rho_X)$ ,  $\mathcal{M}_Y = (Y, A_Y, \tau_Y, \rho_Y)$  be MDPs, and  $d_1 \sqsubseteq F_\lambda^{\mathcal{M}_X}(d_1)$ ,  $d_2 \sqsubseteq F_\lambda^{\mathcal{M}_Y}(d_2)$ .

1. Assume  $\mathcal{M} = \mathcal{M}_X \mid \mathcal{M}_Y$ ,  $\tau = \tau_X \mid_{\tau} \tau_Y$ , and  $\rho = \rho_X \mid_{\rho} \rho_Y$ .

We have to show that

$$F_{\lambda}^{\mathcal{M}}(\|d_1, d_2\|_1) \sqsubseteq \|F_{\lambda}^{\mathcal{M}_X}(d_1), F_{\lambda}^{\mathcal{M}_Y}(d_2)\|_1.$$

Note that, for  $d \in A_X \cap A_Y$ ,

$$\omega_X \in \Pi(\tau_X(x, d), \tau_X(x', d))$$

and

$$\omega_Y \in \Pi(\tau_Y(y, d), \tau_Y(y', d)).$$

The function  $\omega: (X \times Y)^2 \rightarrow [0, 1]$ , defined by

$$\omega((u, v), (u', v')) = \omega_X(u, u') \cdot \omega_Y(v, v'),$$

is a probability distribution on  $(X \times Y)^2$ .

Moreover, the definition of  $\mid$  guarantees that

- i.  $\omega \in \Pi(\tau((x, y), d), \tau((x', y'), d))$ ;
- ii.  $\rho((x, y), d) = \rho_X(x, d) + \rho_Y(y, d)$  and  $\rho((x', y'), d) = \rho_X(x', d) + \rho_Y(y', d)$ .

Let  $x, x' \in X, y, y' \in Y, c \in A_X \cap A_Y$  and  $\tilde{\omega} \in \Pi(\tau((x, y), c), \tau((x', y'), c))$  be such that

$$F_{\lambda}^{\mathcal{M}}(\|d_1, d_2\|_1)((x, y), (x', y')) = |\rho((x, y), c) - \rho((x', y'), c)| + \lambda \sum_{\substack{u, u' \in X \\ v, v' \in Y}} (d_1(u, u') + d_2(v, v')) \cdot \tilde{\omega}((u, v), (u', v')),$$

i.e.,  $\tilde{\omega}$  is the matching that minimizes the last summand, and  $c$  is the action that maximizes the whole expression (see the definition of  $F_{\lambda}^{\mathcal{M}}$ ).

To ease the notation, let  $P$  and  $Q$  denote the following expressions:

$$P = |\rho_X(x, c) - \rho_X(x', c)| + \lambda \sum_{\substack{u, u' \in X \\ v, v' \in Y}} d_1(u, u') \cdot \tilde{\omega}((u, v), (u', v')),$$

$$Q = |\rho_Y(y, c) - \rho_Y(y', c)| + \lambda \sum_{\substack{u, u' \in X \\ v, v' \in Y}} d_2(v, v') \cdot \tilde{\omega}((u, v), (u', v')).$$

Using (ii) and  $|(A - B) + (C - D)| \leq |A - B| + |C - D|$ , it is easy to see that

$$F_{\lambda}^{\mathcal{M}}(\|d_1, d_2\|_1)((x, y), (x', y')) \leq P + Q.$$

To prove the thesis it suffices to show that

$$P \leq F_{\lambda}^{\mathcal{M}_X}(d_1)(x, x') \quad \text{and} \quad Q \leq F_{\lambda}^{\mathcal{M}_Y}(d_2)(y, y').$$

Indeed

$$F_\lambda^{\mathcal{M}_X}(d_1)(x, x') + F_\lambda^{\mathcal{M}_Y}(d_2)(y, y') = \|F_\lambda^{\mathcal{M}_X}(d_1), F_\lambda^{\mathcal{M}_Y}(d_2)\|_1((x, y), (x', y')).$$

Here we show only  $P \leq F_\lambda^{\mathcal{M}_X}(d_1)(x, x')$ ; the other inequality can be proven similarly.

Let  $a \in A_X$  and  $\tilde{\omega}_X \in \Pi(\tau_X(x, a), \tau_X(x', a))$  be such that

$$F_\lambda^{\mathcal{M}_X}(d_1)(x, x') = |\rho_X(x, a) - \rho_X(x', a)| + \lambda \sum_{u, u' \in X} d_1(u, u') \tilde{\omega}_X(u, u'),$$

**Case  $c = a$ :** By (i) we have that for any  $\omega_Y \in \Pi(\tau_Y(y, a), \tau_Y(y', a))$

$$\begin{aligned} P &\leq |\rho_X(x, a) - \rho_X(x', a)| + \lambda \sum_{\substack{u, u' \in X \\ v, v' \in Y}} d_1(u, u') \tilde{\omega}_X(u, u') \omega_Y(v, v') \\ &= |\rho_X(x, a) - \rho_X(x', a)| + \lambda \sum_{u, u' \in X} d_1(u, u') \tilde{\omega}_X(u, u') \\ &= F_\lambda^{\mathcal{M}_X}(d_1)(x, x'). \end{aligned}$$

**Case  $c \neq a$ :** By (i) we have that for any  $\omega_X \in \Pi(\tau_X(x, c), \tau_X(x', c))$  and  $\omega_Y \in \Pi(\tau_Y(y, c), \tau_Y(y', c))$

$$\begin{aligned} P &\leq |\rho_X(x, c) - \rho_X(x', c)| + \lambda \sum_{\substack{u, u' \in X \\ v, v' \in Y}} d_1(u, u') \omega_X(u, u') \omega_Y(v, v') \\ &= |\rho_X(x, c) - \rho_X(x', c)| + \lambda \sum_{u, u' \in X} d_1(u, u') \omega_X(u, u') \\ &\leq F_\lambda^{\mathcal{M}_X}(d_1)(x, x'). \end{aligned}$$

2. Assume that  $\mathcal{M} = \mathcal{M}_X \parallel \mathcal{M}_Y$ ,  $\tau = \tau_X \parallel_\tau \tau_Y$ , and  $\rho = \rho_X \parallel_\rho \rho_Y$ .

We need to prove that

$$F_\lambda^{\mathcal{M}}(\|d_1, d_2\|_1) \sqsubseteq \|F_\lambda^{\mathcal{M}_X}(d_1), F_\lambda^{\mathcal{M}_Y}(d_2)\|_1.$$

Let  $x, x' \in X, y, y' \in Y, c \in A_X \cup A_Y$  and  $\tilde{\omega} \in \Pi(\tau((x, y), c), \tau((x', y'), c))$  be such that

$$\begin{aligned} F_\lambda^{\mathcal{M}}(\|d_1, d_2\|_1)((x, y), (x', y')) &= |\rho((x, y), c) - \rho((x', y'), c)| + \\ &\quad \lambda \sum_{\substack{u, u' \in X \\ v, v' \in Y}} (d_1(u, u') + d_2(v, v')) \cdot \tilde{\omega}((u, v), (u', v')). \end{aligned} \quad (13.4.1)$$

**Case  $c \notin A_Y$ :** Notice firstly that, by definition of  $\parallel$ , for  $v, v' \in Y$  and arbitrary  $\omega \in \Pi(\tau((x, y), c), \tau((x', y'), c))$  the following condition is satisfied

- i.  $v \neq y$  or  $v' \neq y' \implies \omega((u, v), (u', v')) = 0$ ;
- ii.  $\omega((\cdot, y), (\cdot, y')) \in \Pi(\tau_X(x, c), \tau_X(x', c))$ ;
- iii.  $\rho((x, y), c) = \rho_X(x, c)$ , and  $\rho((x', y'), c) = \rho_X(x', c)$ .

Given (i)–(iii), it is easy to see that (13.4.1) can be rewritten as follows:

$$F_\lambda^{\mathcal{M}}(\|d_1, d_2\|_1)((x, y), (x', y')) = |\rho_X(x, c) - \rho_X(x', c)| + \lambda \sum_{u, u' \in X} d_1(u, u') \cdot \tilde{\omega}((u, y), (u', y')) + \lambda d_2(y, y').$$

By (i) and (ii), it also holds that

$$F_\lambda^{\mathcal{M}^X}(d_1)(x, x') \geq |\rho_X(x, c) - \rho_X(x', c)| + \lambda \sum_{u, u' \in X} d_1(u, u') \cdot \tilde{\omega}((u, y), (u', y')).$$

From the above considerations, we have

$$\begin{aligned} F_\lambda^{\mathcal{M}}(\|d_1, d_2\|_1)((x, y), (x', y')) &\leq F_\lambda^{\mathcal{M}^X}(d_1)(x, x') + \lambda d_2(y, y') \\ &\leq F_\lambda^{\mathcal{M}^X}(d_1)(x, x') + \lambda F_\lambda^{\mathcal{M}^Y}(d_2)(y, y') \\ &\leq F_\lambda^{\mathcal{M}^X}(d_1)(x, x') + F_\lambda^{\mathcal{M}^Y}(d_2)(y, y') \\ &= \|F_\lambda^{\mathcal{M}^X}(d_1), F_\lambda^{\mathcal{M}^Y}(d_2)\|_1((x, y), (x', y')). \end{aligned}$$

**Case  $c \notin A_X$ :** Similarly to the previous case.

**Case  $c \in A_X \cap A_Y$ :** Similarly to 1. □

## 13.5 Alternative Characterization of the Pseudometric

In this section we give an alternative characterization of  $\delta_\lambda^{\mathcal{M}}$  based on the notion of *coupling*, that allows us to transfer the results previously proven for Markov chains in [5] and presented in this monograph in Chapter 11, to MDPs. Then, we show how to relate this characterization to the concept of non-extensiveness for compositional operators on MDPs.

**Definition 13.5.1 (Coupling).** Let  $\mathcal{M} = (S, A, \tau, \rho)$  be an MDP. A coupling for  $\mathcal{M}$  is a pair  $\mathcal{C} = (\rho, \omega)$ , where  $\omega: (S \times S) \times A \rightarrow \Delta(S \times S)$  is such that, for any  $s, t \in S$  and  $a \in A$ ,

$$\omega((s, t), a) \in \Pi(\tau(s, a), \tau(t, a)).$$

Given a coupling  $\mathcal{C} = (\rho, \omega)$  for  $\mathcal{M}$  and a discount factor  $\lambda \in (0, 1)$ , we define the operator  $\Gamma_\lambda^{\mathcal{C}}: [0, 1]^{S \times S} \rightarrow [0, 1]^{S \times S}$ , for  $d \in [0, 1]^{S \times S}$  and  $s, t \in S$ , by

$$\Gamma_\lambda^{\mathcal{C}}(d)(s, t) = \max_{a \in A} (|\rho(s, a) - \rho(t, a)| + \lambda \sum_{u, v \in S} d(u, v) \cdot \omega((s, t), a)(u, v)).$$

Note that, any coupling  $\mathcal{C} = (\rho, \omega)$  for  $\mathcal{M}$  induces an MDP  $\mathcal{C}^* = (S \times S, A, \omega, \rho^*)$ , defined for any  $s, t \in S$  and  $a \in A$  by

$$\rho^*((s, t), a) = |\rho(s, a) - \rho(t, a)|,$$

and  $\Gamma_\lambda^{\mathcal{C}}$  corresponds to the *Bellman optimality operator* on  $\mathcal{C}^*$ .

This operator is monotonic and has a unique fixed point, hereafter denoted by  $\gamma_\lambda^{\mathcal{C}}$ , corresponding to the value function for  $\mathcal{C}^*$ , as demonstrated in [37, §6.2].

There exist certain dependences between  $\gamma_\lambda^{\mathcal{C}}$  and the bisimilarity pseudometric  $\delta_\lambda^{\mathcal{M}}$  that will be proven next.

**Lemma 13.5.2.** *Let  $\mathcal{C}$  be a coupling for the DMC  $\mathcal{M}$ . Then,  $\delta_\lambda^{\mathcal{M}} \sqsubseteq \gamma_\lambda^{\mathcal{C}}$ .*

*Proof.* Assume that  $\mathcal{M} = (S, A, \tau, \rho)$  and  $\mathcal{C} = (\rho, \omega)$ .

In order to prove  $\delta_\lambda^{\mathcal{M}} \sqsubseteq \gamma_\lambda^{\mathcal{C}}$  using Tarski's fixed point theorem, it suffices to show that  $F_\lambda^{\mathcal{M}}(\gamma_\lambda^{\mathcal{C}}) \sqsubseteq \gamma_\lambda^{\mathcal{C}}$ .

Let  $s, t \in S$ . Since  $\omega((s, t), a) \in \Pi(\tau(s, a), \tau(t, a))$  we have

$$\begin{aligned} F_\lambda^{\mathcal{M}}(\gamma_\lambda^{\mathcal{C}})(s, t) &= \max_{a \in A} (|\rho(s, a) - \rho(t, a)| + \lambda \mathcal{T}_{\gamma_\lambda^{\mathcal{C}}}(\tau(s, a), \tau(t, a))) \\ &\leq \max_{a \in A} (|\rho(s, a) - \rho(t, a)| + \lambda \sum_{u, v \in S} \gamma_\lambda^{\mathcal{C}}(u, v) \omega((s, t), a)(u, v)) \\ &= \Gamma_\lambda^{\mathcal{C}}(\gamma_\lambda^{\mathcal{C}})(s, t). \end{aligned}$$

Since  $\gamma_\lambda^{\mathcal{C}} = \Gamma_\lambda^{\mathcal{C}}(\gamma_\lambda^{\mathcal{C}})$ , we obtain  $F_\lambda^{\mathcal{M}}(\gamma_\lambda^{\mathcal{C}}) \sqsubseteq \gamma_\lambda^{\mathcal{C}}$ .  $\square$

Next we see that the bisimilarity pseudometric  $\delta_\lambda^{\mathcal{M}}$  can be characterized as the minimum  $\gamma_\lambda^{\mathcal{C}}$  among all the couplings  $\mathcal{C}$  for  $\mathcal{M}$ .

**Theorem 13.5.3.** *Let  $\mathcal{M}$  be an MDP. Then,*

$$\delta_\lambda^{\mathcal{M}} = \min\{\gamma_\lambda^{\mathcal{C}} \mid \mathcal{C} \text{ coupling for } \mathcal{M}\}.$$

*Proof.* Let  $\mathcal{M} = (S, A, \tau, \rho)$ .

For any fixed  $d \in [0, 1]^{S \times S}$  there exists a coupling  $\mathcal{C} = (\rho, \omega)$  for  $\mathcal{M}$  such that

$$\Gamma_\lambda^{\mathcal{C}}(d) = F_\lambda^{\mathcal{M}}(d).$$

This can be defined by taking  $\omega$  such that, for all  $s, t \in S$ ,

$$\mathcal{T}_d(\tau(s, a), \tau(t, a)) = \sum_{u, v \in S} d(u, v) \cdot \omega((s, t), a)(u, v).$$

Let  $\mathcal{D}$  be a coupling for  $\mathcal{M}$  such that

$$\Gamma_\lambda^{\mathcal{D}}(\delta_\lambda^{\mathcal{M}}) = F_\lambda^{\mathcal{M}}(\delta_\lambda^{\mathcal{M}}).$$

By definition,  $F_\lambda^{\mathcal{M}}(\delta_\lambda^{\mathcal{M}}) = \delta_\lambda^{\mathcal{M}}$ , therefore  $\delta_\lambda^{\mathcal{M}}$  is a fixed point for  $\Gamma_\lambda^{\mathcal{D}}$ .

Using the uniqueness of the fixed point of  $\Gamma_\lambda^{\mathcal{D}}$ , we get  $\delta_\lambda^{\mathcal{M}} = \gamma_\lambda^{\mathcal{D}}$ .

Let  $D = \{\gamma_\lambda^{\mathcal{C}} \mid \mathcal{C} \text{ coupling for } \mathcal{M}\}$ . We have  $\delta_\lambda^{\mathcal{M}} \in D$  and, applying Lemma 13.5.2,  $\delta_\lambda^{\mathcal{M}}$  is a lower bound for  $D$ . Hence,  $\delta_\lambda^{\mathcal{M}} = \min D$ .  $\square$

Theorem 13.5.3 allows us to transfer the compositional reasoning on couplings. To this end, we introduce the notion of composition operator on couplings.

**Definition 13.5.4.** Let  $\mathcal{M}_i = (S_i, A_i, \tau_i, \rho_i)$ , for  $i = 1..n$ , be MDPs. A coupling composition operator for  $\mathcal{M}_1, \dots, \mathcal{M}_n$  is a tuple  $op^* = (A, op_\rho^*, op_\omega^*)$  consisting of a nonempty set  $A$ , and the following operations, where  $S = \prod_{i=1}^n S_i$ .

- $op_\rho^*: \prod_{i=1}^n \mathbb{R}^{S_i \times A_i} \rightarrow \mathbb{R}^{S \times A}$ ,
- $op_\omega^*: \prod_{i=1}^n \Delta(S_i \times S_i)^{S_i \times S_i \times A_i} \rightarrow \Delta(S \times S)^{S \times S \times A}$ .

Let  $\mathcal{C}_i = (\rho_i, \omega_i)$  be a coupling for  $\mathcal{M}_i$ ,  $i = 1..n$ . We denote by  $op^*(\mathcal{C}_1, \dots, \mathcal{C}_n)$  the composite coupling  $(op_\rho^*(\rho_1, \dots, \rho_n), op_\omega^*(\omega_1, \dots, \omega_n))$ . Moreover,  $op^*$  is called *lifting* of a composition operator  $op$  on  $\mathcal{M}_1, \dots, \mathcal{M}_n$ , if for all  $i = 1..n$  and  $\mathcal{C}_i$  coupling for  $\mathcal{M}_i$ ,  $op^*(\mathcal{C}_1, \dots, \mathcal{C}_n)$  is a coupling for  $op(\mathcal{M}_1, \dots, \mathcal{M}_n)$ .

It is not always possible to find coupling composition operators that lift a composition operator on MDPs. Nevertheless, the composite operators presented in Examples 13.3.2–13.3.3 can be lifted on couplings. We show in the next example how this can be done for the CCS-like parallel composition. For the other examples the construction is similar.

**Example 13.5.5.** The composition operator of Example 13.3.3 can be lifted on couplings by the operator  $\|\cdot\|^* = (A_X \cup A_Y, \|\rho, \|\omega)$

$$\begin{aligned} & (\omega_X \|\omega_Y)((x, y), (x', y'), a)((u, v), (u', v')) = \\ & = \begin{cases} \omega_X((x, x'), a)(u, u') & \text{if } a \notin A_Y, (v, v') = (y, y') \\ \omega_Y((y, y'), a)(v, v') & \text{if } a \notin A_X, (u, u') = (x, x') \\ \omega_X((x, x'), a)(u, u') \cdot \omega_Y((y, y'), a)(v, v') & \text{if } a \in A_X \cap A_Y \\ 0 & \text{otherwise} \end{cases} \end{aligned}$$

Note how the definition above mimics the one in Example 13.3.3. ■

Next we adapt the concept of safeness to coupling composition operators.

**Definition 13.5.6.** Let  $\mathcal{M}_i = (S_i, A_i, \tau_i, \rho_i)$  be MDPs,  $i = 1..n$  and  $p \in [1, \infty]$ . A coupling composition operator  $op^*$  on  $\mathcal{M}_1, \dots, \mathcal{M}_n$  is  $p$ -safe if whenever  $\mathcal{C}_i$  is a coupling for  $\mathcal{M}_i$  and  $d_i: S_i \times S_i \rightarrow [0, \infty)$  is such that  $d_i \sqsubseteq \Gamma_\lambda^{\mathcal{C}_i}(d_i)$ , for all  $i = 1..n$ , the following assertion is satisfied.

$$\Gamma_\lambda^{op^*(\mathcal{C}_1, \dots, \mathcal{C}_n)}(\|d_1, \dots, d_n\|_p) \sqsubseteq \|\Gamma_\lambda^{\mathcal{C}_1}(d_1), \dots, \Gamma_\lambda^{\mathcal{C}_n}(d_n)\|_p.$$

A coupling composition operator is safe if it is  $p$ -safe for some  $p \in [1, \infty]$ .

As done for Proposition 13.4.6, the lifting in Example 13.5.5 can be shown to be 1-safe.

The non-extensiveness for an operator is ensured if it admits a lifting composition operator on couplings that is safe, as proven by the following theorem.

**Theorem 13.5.7.** *Let  $op^*$  be a coupling composition operator that lifts a composition operator  $op$  on  $\mathcal{M}_1, \dots, \mathcal{M}_n$ . If  $op^*$  is safe, then  $op$  is non-extensive.*

*Proof.* Applying Theorem 13.5.3, there exists a coupling  $\mathcal{C}_i$  for  $\mathcal{M}_i$ , for each  $1 \leq i \leq n$ , such that  $\delta_\lambda^{\mathcal{M}_i} = \gamma_\lambda^{\mathcal{C}_i}$ .

The hypothesis of the theorem ensures us that there exists  $1 \leq p \leq \infty$ , such that

$$\Gamma_\lambda^{op^*(\mathcal{C}_1, \dots, \mathcal{C}_n)}(\|\gamma_\lambda^{\mathcal{C}_1}, \dots, \gamma_\lambda^{\mathcal{C}_n}\|_p) \sqsubseteq \|\Gamma_\lambda^{\mathcal{C}_1}(\gamma_\lambda^{\mathcal{C}_1}), \dots, \Gamma_\lambda^{\mathcal{C}_n}(\gamma_\lambda^{\mathcal{C}_n})\|_p.$$

Since, for  $1 \leq i \leq n$ ,

$$\Gamma_\lambda^{\mathcal{C}_i}(\gamma_\lambda^{\mathcal{C}_i}) = \gamma_\lambda^{\mathcal{C}_i} = \delta_\lambda^{\mathcal{M}_i},$$

we have that  $\|\delta_\lambda^{\mathcal{M}_1}, \dots, \delta_\lambda^{\mathcal{M}_n}\|_p$  is a pre-fixed point of  $\Gamma_\lambda^{op^*(\mathcal{C}_1, \dots, \mathcal{C}_n)}$ . Therefore, applying Tarski's fixed point theorem,

$$\gamma_\lambda^{op^*(\mathcal{C}_1, \dots, \mathcal{C}_n)} \sqsubseteq \|\delta_\lambda^{\mathcal{M}_1}, \dots, \delta_\lambda^{\mathcal{M}_n}\|_p.$$

By hypothesis  $op^*$  is a lifting of  $op$ , that is  $op^*(\mathcal{C}_1, \dots, \mathcal{C}_n)$  is a coupling for  $({}^{\text{op}}\mathcal{M}_1, \dots, \mathcal{M}_n)$  and Lemma 13.5.2 ensures us that

$$\delta_\lambda^{({}^{\text{op}}\mathcal{M}_1, \dots, \mathcal{M}_n)} \sqsubseteq \gamma_\lambda^{op^*(\mathcal{C}_1, \dots, \mathcal{C}_n)}.$$

Therefore,  $\delta_\lambda^{({}^{\text{op}}\mathcal{M}_1, \dots, \mathcal{M}_n)} \sqsubseteq \|\delta_\lambda^{\mathcal{M}_1}, \dots, \delta_\lambda^{\mathcal{M}_n}\|_p$ .  $\square$

## 13.6 Exact Computation of Bisimilarity Distance

Inspired by the characterization provided in Theorem 13.5.3, in this section we propose a procedure to exactly compute the bisimilarity pseudometric. This extends to MDPs the method proposed in [5] and described in Chapter 11 for Markov chains. We also show how this strategy can be optimized to cope well with composite MDPs.

For a discount factor  $\lambda \in (0, 1)$ , the set of couplings for  $\mathcal{M}$  can be endowed with the preorder  $\preceq_\lambda$ , defined by

$$\mathcal{C} \preceq_\lambda \mathcal{D} \text{ iff } \gamma_\lambda^{\mathcal{C}} \sqsubseteq \gamma_\lambda^{\mathcal{D}}.$$

Theorem 13.5.3 suggests to look for a coupling for  $\mathcal{M}$  which is minimal w.r.t.  $\preceq_\lambda$ . The enumeration of all the couplings is clearly unfeasible, therefore it is crucial to provide an efficient search strategy which prevents us to do that.

**A Greedy Search Strategy:** We propose a greedy strategy that explores the set of couplings until an optimal one is eventually reached.

Let  $\mathcal{M} = (S, A, \tau, \rho)$  be an MDP and  $\mathcal{C} = (\rho, \omega)$  be a coupling for  $\mathcal{M}$ .

Given  $s, t \in S$ ,  $a \in A$ , and  $\mu \in \Pi(\tau(s, a), \tau(t, a))$ , we denote by  $\mathcal{C}[(s, t), a/\mu]$  the coupling  $(\rho, \omega')$  for  $\mathcal{M}$ , where  $\omega'$  is such that

$$\omega'((s, t), a) = \mu \quad \text{and} \quad \omega'((s', t'), a') = \omega((s', t'), a')$$

for all  $s', t' \in S$  and  $a' \in A$  with  $((s', t'), a') \neq ((s, t), a)$ .

**Lemma 13.6.1.** *Let  $\mathcal{M} = (S, A, \tau, \rho)$  be an MDP,  $\mathcal{C}$  be a coupling for  $\mathcal{M}$ ,  $s, t \in S$ ,  $a \in A$ ,  $\mu \in \Pi(\tau(s, a), \tau(t, a))$ , and  $\mathcal{D} = \mathcal{C}[(s, t), a/\mu]$ .*

$$\text{If } \Gamma_\lambda^{\mathcal{D}}(\gamma_\lambda^{\mathcal{C}})(s, t) < \gamma_\lambda^{\mathcal{C}}(s, t), \text{ then } \gamma_\lambda^{\mathcal{D}} \sqsubset \gamma_\lambda^{\mathcal{C}}.$$

*Proof.* Applying Tarski's fixed point theorem, it suffices to show that

$$\Gamma^{\mathcal{D}}(\gamma_\lambda^{\mathcal{C}}) \sqsubset \gamma_\lambda^{\mathcal{C}}.$$

Assume  $\mathcal{C} = (\rho, \omega)$ ,  $\mathcal{D} = (\rho, \bar{\omega})$ , and let  $s', t' \in S$ .

If  $(s', t') \neq (s, t)$ , by definition of  $\mathcal{D}$ , we have that for all  $a' \in A$ ,  $\bar{\omega}((s', t'), a') = \omega((s', t'), a')$ , hence

$$\Gamma_\lambda^{\mathcal{C}}(\gamma_\lambda^{\mathcal{C}})(s', t') = \Gamma_\lambda^{\mathcal{D}}(\gamma_\lambda^{\mathcal{C}})(s', t').$$

Since, by hypothesis,  $\Gamma_\lambda^{\mathcal{D}}(\gamma_\lambda^{\mathcal{C}})(s, t) < \gamma_\lambda^{\mathcal{C}}(s, t)$ , we have  $\Gamma^{\mathcal{D}}(\gamma_\lambda^{\mathcal{C}}) \sqsubset \gamma_\lambda^{\mathcal{C}}$ .  $\square$

The lemma above states that  $\mathcal{C}$  can be improved w.r.t.  $\preceq_\lambda$  by locally updating it as  $\mathcal{C}[(s, t), a/\mu]$ , with a matching  $\mu \in \Pi(\tau(s, a), \tau(t, a))$  such that

$$\sum_{u, v \in S} \gamma_\lambda^{\mathcal{C}}(u, v) \cdot \mu(u, v) < \sum_{u, v \in S} \gamma_\lambda^{\mathcal{C}}(u, v) \cdot \omega((s, t), a)(u, v),$$

where  $a \in A$  is the action that maximizes  $\Gamma^{\mathcal{C}}(\gamma_\lambda^{\mathcal{C}})_\lambda(s, t)$ .

A matching  $\mu$  satisfying the condition above can be obtained as a solution of a Transportation Problem [20] with cost matrix  $(\gamma_\lambda^{\mathcal{C}}(u, v))_{u, v \in S}$  and marginals  $\tau(s, a)$  and  $\tau(t, a)$ , hereafter denoted by

$$TP(\gamma_\lambda^{\mathcal{C}}, \tau(s, a), \tau(t, a)).$$

This gives us a strategy for moving toward  $\delta_\lambda^{\mathcal{M}}$  by successive improvements on the couplings.

Now we give a necessary and sufficient condition for termination.

**Lemma 13.6.2.** *Let  $\mathcal{M} = (S, A, \tau, \rho)$  be an MDP and  $\mathcal{C}$  be a coupling for  $\mathcal{M}$ . If  $\gamma_\lambda^{\mathcal{C}} \neq \delta_\lambda^{\mathcal{M}}$ , then there exist  $s, t \in S$ ,  $a \in A$ , and  $\mu \in \Pi(\tau(s, a), \tau(t, a))$  such that*

$$\Gamma_\lambda^{\mathcal{D}}(\gamma_\lambda^{\mathcal{C}})(s, t) < \gamma_\lambda^{\mathcal{C}}(s, t),$$

where  $\mathcal{D} = \mathcal{C}[(s, t), a/\mu]$ .

*Proof.* If for all  $s, t \in S$ ,  $a \in A$ , and  $\mu \in \Pi(\tau(s, a), \tau(t, a))$ ,

$$\Gamma_\lambda^{\mathcal{D}}(\gamma_\lambda^{\mathcal{C}})(s, t) \geq \gamma_\lambda^{\mathcal{C}}(s, t),$$

then  $\gamma_\lambda^{\mathcal{C}} = F_\lambda^{\mathcal{M}}(\gamma_\lambda^{\mathcal{C}})$ . Since  $F_\lambda^{\mathcal{M}}$  has a unique fixed point,  $\gamma_\lambda^{\mathcal{C}} = \delta_\lambda^{\mathcal{M}}$ .  $\square$

The above result ensures that, unless  $\mathcal{C}$  is optimal w.r.t  $\sqsubseteq_\lambda$ , the hypotheses of Lemma 13.6.1 are satisfied, so that, we can further improve  $\mathcal{C}$  following the same strategy. The next statement proves that this search strategy is correct.

**Theorem 13.6.3.**  $\delta_\lambda^{\mathcal{M}} = \gamma_\lambda^{\mathcal{C}}$  iff there exists no coupling  $\mathcal{D}$  for  $\mathcal{M}$  s.t.  $\Gamma_\lambda^{\mathcal{D}}(\gamma_\lambda^{\mathcal{C}}) \sqsubset \gamma_\lambda^{\mathcal{C}}$ .

*Proof.* We prove that  $\delta_\lambda \neq \gamma_\lambda^{\mathcal{C}}$  if and only if there exists a coupling  $\mathcal{D}$  for  $\mathcal{M}$  such that  $\Gamma_\lambda^{\mathcal{D}}(\gamma_\lambda^{\mathcal{C}}) \sqsubset \gamma_\lambda^{\mathcal{C}}$ .

( $\Rightarrow$ ) Assume  $\delta_\lambda \neq \gamma_\lambda^{\mathcal{C}}$ . From Lemma 13.6.2, there exist  $s, t \in S$ ,  $a \in A$  and  $\mu \in \Pi(\tau(s, a), \tau(t, a))$  s.t.

$$\Gamma_\lambda^{\mathcal{D}}(\gamma_\lambda^{\mathcal{C}})(s, t) < \gamma_\lambda^{\mathcal{C}}(s, t),$$

where  $\mathcal{D} = \mathcal{C}[(s, t), a/\mu]$ .

As in the proof of Lemma 13.6.1, we have  $\Gamma_\lambda^{\mathcal{D}}(\gamma_\lambda^{\mathcal{C}}) \sqsubset \gamma_\lambda^{\mathcal{C}}$ .

( $\Leftarrow$ ) Let  $\mathcal{D}$  be such that  $\Gamma_\lambda^{\mathcal{D}}(\gamma_\lambda^{\mathcal{C}}) \sqsubset \gamma_\lambda^{\mathcal{C}}$ . Using Tarski's fixed point theorem  $\gamma_\lambda^{\mathcal{D}} \sqsubset \gamma_\lambda^{\mathcal{C}}$ . By Lemma 13.5.2,  $\delta_\lambda \sqsubseteq \gamma_\lambda^{\mathcal{D}} \sqsubset \gamma_\lambda^{\mathcal{C}}$ .  $\square$

**Remark 13.6.4.** In general, there could be an infinite number of couplings  $(\rho, \omega)$ . However, for each fixed  $d \in [0, 1]^{S \times S}$ , the linear function mapping  $\omega((s, t), a)$  to  $\sum_{u, v \in S} d(u, v) \cdot \omega((s, t), a)(u, v)$  achieves its minimum at some vertex of the transportation polytope  $P = \Pi(\tau(s, a), \tau(t, a))$ . Since the number of such vertices is finite, using the optimal transportation schedule (which is a vertex in  $P$ ) for the update ensures that the search strategy is always terminating.  $\blacksquare$

**Compositional Heuristic:** Assume we want to compute the bisimilarity distance for a composite MDP  $\mathcal{M} = op(\mathcal{M}_1, \dots, \mathcal{M}_n)$ . The greedy strategy described above moves toward an optimal coupling for  $\mathcal{M}$  starting from an arbitrary one. Clearly, the better is the initial coupling the fewer are the steps to the optimal one. The following result gives a heuristic for choosing such a coupling when  $op$  admits a safe lifting coupling composition operator.

**Proposition 13.6.5.** *Let  $op$  be a composition operator on  $\mathcal{M}_1, \dots, \mathcal{M}_n$ , and  $op^*$  be a  $p$ -safe coupling composition operator that lifts  $op$ . Then,*

- (i)  $\gamma_\lambda^{op^*(\mathcal{C}_1, \dots, \mathcal{C}_n)} \sqsubseteq \|\gamma_\lambda^{\mathcal{C}_1}, \dots, \gamma_\lambda^{\mathcal{C}_n}\|_p$ , for any  $\mathcal{C}_i$  coupling for  $\mathcal{M}_i$ ;
- (ii)  $\delta_\lambda^{op(\mathcal{M}_1, \dots, \mathcal{M}_n)} \sqsubseteq \gamma_\lambda^{op^*(\mathcal{D}_1, \dots, \mathcal{D}_n)} \sqsubseteq \|\delta_\lambda^{\mathcal{M}_1}, \dots, \delta_\lambda^{\mathcal{M}_n}\|_p$ , where  $\mathcal{D}_i$  is a coupling for  $\mathcal{M}_i$  which is minimal w.r.t.  $\preceq_\lambda$ .

*Proof.* We prove the two items separately.

- (i) By hypothesis  $op^*$  is a  $p$ -safe, hence for all  $\mathcal{C}_i$  coupling for  $\mathcal{M}_i$ , we have

$$\Gamma_\lambda^{op^*(\mathcal{C}_1, \dots, \mathcal{C}_n)}(\|\gamma_\lambda^{\mathcal{C}_1}, \dots, \gamma_\lambda^{\mathcal{C}_n}\|_p) \sqsubseteq \|\gamma_\lambda^{\mathcal{C}_1}, \dots, \gamma_\lambda^{\mathcal{C}_n}\|_p.$$

Applying Tarski's fixed point theorem,

$$\gamma_\lambda^{op^*(\mathcal{C}_1, \dots, \mathcal{C}_n)} \sqsubseteq \|\gamma_\lambda^{\mathcal{C}_1}, \dots, \gamma_\lambda^{\mathcal{C}_n}\|_p.$$

- (ii) Since  $\mathcal{D}_i$  are minimal w.r.t.  $\preceq_\lambda$ , Theorem 13.5.3 guarantees that we have  $\delta_\lambda^{\mathcal{M}_i} = \gamma_\lambda^{\mathcal{D}_i}$ . Using (i), it holds

$$\gamma_\lambda^{op^*(\mathcal{D}_1, \dots, \mathcal{D}_n)} \sqsubseteq \|\delta_\lambda^{\mathcal{M}_1}, \dots, \delta_\lambda^{\mathcal{M}_n}\|_p.$$

Since  $op^*$  is a lifting of  $op$ , Lemma 13.5.2 implies  $\delta_\lambda^{op(\mathcal{M}_1, \dots, \mathcal{M}_n)} \sqsubseteq \gamma_\lambda^{op^*(\mathcal{D}_1, \dots, \mathcal{D}_n)}$ .

□

Proposition 13.6.5(ii) suggests to start from the coupling  $op^*(\mathcal{D}_1, \dots, \mathcal{D}_n)$ , i.e., the one given as the composite of the optimal couplings  $\mathcal{D}_i$  for the subcomponents  $\mathcal{M}_i$ . This ensures that the first over-approximation of  $\delta_\lambda^{\mathcal{M}}$ , that is  $\gamma_\lambda^{op^*(\mathcal{D}_1, \dots, \mathcal{D}_n)}$ , is at least as good as the upper bound given by non-extensiveness of  $op$ .

**Algorithm 4** On-the-Fly Bisimilarity Pseudometric

---

**Input:** MDP  $\mathcal{M} = (S, A, \tau, \rho)$ ; discount factor  $\lambda \in (0, 1)$ ; query  $Q \subseteq S \times S$ .

1.  $\mathcal{C} \leftarrow (\rho, \text{empty}); d \leftarrow \text{empty}; \text{visited} \leftarrow \emptyset; \text{exact} \leftarrow \emptyset; \text{toComp} \leftarrow Q;$      —Initialize—
  2. **while**  $\exists (s, t) \in \text{toComp}$  **do**
  3.     **for all**  $a \in A$  **do** guess  $\mu \in \Pi(\tau(s, a), \tau(t, a)); \text{UpdateC}(\mathcal{M}, (s, t), a, \mu)$
  4.      $d \leftarrow \text{BellmanOpt}(\lambda, \mathcal{C}, d)$      —update the current estimate—
  5.     **while**  $\mathcal{C}[(u, v), a]$  is not optimal for  $TP(d, \tau(u, a), \tau(v, a))$  **do**
  6.          $\mu \leftarrow \text{optimal schedule for } TP(d, \tau(u, a), \tau(v, a))$
  7.          $\text{UpdateC}(\mathcal{M}, (u, v), a, \mu)$      —improve the current coupling—
  8.          $d \leftarrow \text{BellmanOpt}(\lambda, \mathcal{C}, d)$      —update the current estimate—
  9.     **end while**
  10.      $\text{exact} \leftarrow \text{exact} \cup \text{visited}$      —add new exact distances—
  11.      $\text{toComp} \leftarrow \text{toComp} \setminus \text{exact}$      —remove exactly computed pairs—
  12. **end while**
  13. **return**  $d \upharpoonright_Q$      —return the distance restricted to the pairs in  $Q$ —
- 

## 13.7 A Compositional On-the-Fly Algorithm

In this section we propose an on-the-fly algorithm for computing the bisimilarity distance making full use of the greedy strategy presented in Section 13.6. Then, we describe how to optimize the computation on composite MDPs.

Let  $\mathcal{M} = (S, A, \tau, \rho)$  be an MDP,  $Q \subseteq S \times S$ , and assume we want to compute  $\delta_\lambda^{\mathcal{M}}$  restricted to  $Q$ , denoted by  $\delta_\lambda^{\mathcal{M}} \upharpoonright_Q$ .

Our strategy has the following features:

- when a coupling  $\mathcal{C}$  is considered,  $\gamma_\lambda^{\mathcal{C}}$  can be computed solving the Bellman optimality equation system associated with it;
- the current coupling  $\mathcal{C}$  can be improved by a *local* update  $\mathcal{C}[(u, v), a / \mu]$  that satisfies the hypotheses of Lemma 14.5.2.

Note that,  $\gamma_\lambda^{\mathcal{C}} \upharpoonright_Q$  can be computed considering only the smallest independent subsystem containing the variables associated with the pairs in  $Q$ . Therefore, we do not need to store the entire coupling, but we can construct it on-the-fly.

The computation of  $\delta_\lambda^{\mathcal{M}} \upharpoonright_Q$  is implemented by Algorithm 4. We assume the following global variables to store:

- $\mathcal{C}$ , the current partial coupling;
- $d$ , the current partial over-approximation of  $\delta_\lambda^{\mathcal{M}}$ ;
- $\text{toComp}$ , the pairs of states for which the distance has to be computed;
- $\text{exact}$ , the pairs of states  $(s, t)$  such that  $d(s, t) = \delta_\lambda^{\mathcal{M}}(s, t)$ ;

**Algorithm 5**  $UpdateC(\mathcal{M}, (s, t), a, \mu)$ **Input:** MDP  $\mathcal{M} = (S, A, \tau, \rho)$ ;  $s, t \in S$ ;  $a \in A$ ,  $\mu \in \Pi(\tau(s, a), \tau(t, a))$ 

1.  $\mathcal{C} \leftarrow \mathcal{C}[(s, t), a/\mu]$  —update the coupling—
2.  $visited \leftarrow visited \cup \{(s, t)\}$  —set  $(s, t)$  as visited—
3. **for all**  $(u, v) \in \{(u', v') \mid \mu(u', v') > 0\} \setminus visited$  **do** —for all demanded pairs—
4.    $visited \leftarrow visited \cup \{(u, v)\}$
5.   // propagate the construction
6.   **for all**  $a \in A$  **do** guess  $\mu' \in \Pi(\tau(u, a), \tau(v, a))$ ;  $UpdateC(\mathcal{M}, (u, v), a, \mu')$
7. **end for**

- visited, the pair of states considered so far.

At the beginning  $\mathcal{C}$  and  $d$  are empty, there are no visited states and no exact distances.

While there are pairs  $(s, t)$  left to be computed we update  $\mathcal{C}$  calling the subroutine  $UpdateC$  on a matching  $\mu \in \Pi(\tau(s, a), \tau(t, a))$ , for each  $a \in A$ .

Then,  $d$  is updated on all visited pairs with the over-approximation  $\gamma_\lambda^{\mathcal{C}}$  by calling  $BellmanOpt$ .

According to the greedy strategy,  $\mathcal{C}$  is successively improved and  $d$  is consequently updated, until no further improvements are possible. Each improvement is demanded by the existence of a better transportation schedule.

When line 10 is reached,  $d(u, v) = \delta_\lambda^{\mathcal{M}}(u, v)$  for all  $(u, v) \in visited$ , therefore visited is added to exact and removed from toComp. If no more pairs have to be considered, the exact distance on  $Q$  is returned.

The subroutine  $UpdateC$  (Algorithm 5) updates the coupling  $\mathcal{C}$  and recursively populates it on all demanded pairs.

$BellmanOpt(\lambda, \mathcal{C}, d)$  solves the smallest independent subsystem of the Bellman optimality equation system on the MDP induced by  $\mathcal{C}$ , that contains all the visited pairs.

Notice that, the equation system can be further reduced by Gaussian elimination, substituting the variables associated with pairs  $(u, v) \in exact$  with  $d(u, v)$ .

**Compositional Optimizations:** Algorithm 4 can be modified to handle composite MDPs efficiently.

Assume  $\mathcal{M} = op(\mathcal{M}_1, \dots, \mathcal{M}_n)$  and that we have a safe coupling composition operator  $op^*$  that lifts  $op$ .

The compositional heuristic described in Section 13.6 suggests to start from the coupling  $op^*(\mathcal{D}_1, \dots, \mathcal{D}_n)$  obtained by composing the optimal couplings  $\mathcal{D}_i$  for each  $\mathcal{M}_i$ .

This is done running Algorithm 4 in two modalities: master/slave. For each  $\mathcal{M}_i$ , the master shares the data structures  $\mathcal{C}_i$ ,  $d_i$ ,  $visited_i$ ,  $toComp_i$  and  $exact_i$  with the corresponding slave to keep track of the computation of  $\delta_\lambda^{\mathcal{M}_i}$ .

When a new pair  $((s_1, \dots, s_n), (t_1, \dots, t_n))$  is considered, the master runs (possibly in parallel)  $n$  slave threads of Algorithm 4 on the query  $\{(s_i, t_i)\}$ .

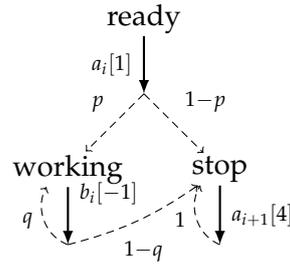


Figure 13.1: Pipeline

Query	Instance	OTF	COTF	# States
All pairs	$E_0 \parallel E_1$	0.654791	0.97248	9
	$E_1 \parallel E_2$	0.702105	0.801121	9
	$E_0 \parallel E_0 \parallel E_1$	48.5982	13.5731	27
	$E_0 \parallel E_1 \parallel E_2$	23.1984	19.9137	27
	$E_0 \parallel E_1 \parallel E_1$	126.335	13.6483	27
	$E_0 \parallel E_0 \parallel E_0$	49.1167	14.1075	27
Single pair	$E_0 \parallel E_0 \parallel E_0 \parallel E_1 \parallel E_1$	16.7027	11.6919	243
	$E_0 \parallel E_1 \parallel E_0 \parallel E_1 \parallel E_1$	20.2666	16.6274	243
	$E_2 \parallel E_1 \parallel E_0 \parallel E_1 \parallel E_1$	22.8357	10.4844	243
	$E_1 \parallel E_2 \parallel E_0 \parallel E_0 \parallel E_2$	11.7968	6.76188	243
	$E_1 \parallel E_2 \parallel E_0 \parallel E_0 \parallel E_2 \parallel E_2$	Time-out	79.902	729

Table 13.1: Comparison between the on-the-fly algorithm (OTF) and its compositional optimization (COTF);  $E_0 = E_0(0.7, 0.2)$ ,  $E_1 = E_1(0.6, 0.2)$ , and  $E_2 = E_2(0.5, 0.3)$ .

At the end of these subcomputations, the couplings  $\mathcal{C}_i$  are optimal, and they are composed to obtain a coupling for  $\mathcal{M}$ .

Note that, the master can reuse the values stored by the slaves in their previous computations.

**Experimental Results:** For Markov chains, in [5] and detailed in Chapter 11 of this monograph, it has already been shown that an on-the-fly strategy yields, on average, significant improvements with respect to the corresponding iterative algorithms.

Here we focus on how the compositional optimization affects the performances. To this end, we consider a simple yet meaningful set of experiments performed on a collection of MDPs, parametric in the probabilities, modeling a pipeline.

The Figure 13.1 specifies an element  $E_i(p, q)$  of the pipeline with actions  $A_i = \{a_i, a_{i+1}, b_i\}$ . Pipelines are modeled as the parallel composition of different processing elements, that are connected in series by means of synchronization on shared actions.

Table 13.1 reports the computation times of the tests<sup>2</sup> we have run both on all-pairs

<sup>2</sup>The tests have been made using a prototype implementation coded in Mathematica<sup>®</sup> (available at

queries and single-pair queries for several pipeline instances; timings are expressed in seconds and, as for the single-pair case, they represent the average of 20 randomly chosen queries.

Table 13.1 shows that the required overhead for maintaining the additional data structure for the subcomponents, affects the performances only on very small systems. In all other cases the compositional optimization yields a significant reduction of the computation time that varies from a factor of 2 up to a factor of 10.

Notably, on single-pair queries the compositional version can manage (relatively) large systems whereas the non-compositional one exceeds a time-bound of 3 minutes. Interestingly, we observe better reductions on all-pairs queries than in single-pairs; this may be due to fact that the exact distances collected during the computation are used to further reduce the size of the equation systems that are successively encountered.

## 13.8 Concluding Remarks

We have proposed a general notion of composition operator on MDPs and identified safeness as a sufficient condition for ensuring non-extensiveness. We showed that the class of safe operators is general enough to cover a wide range of known composition operators. Moreover, we presented an algorithm for computing bisimilarity distances on MDPs, which is able to exploit the compositional structure of the system and applies on MDPs built over any safe operators. This is the first proposal for a compositional algorithm for computing bisimilarity distances; before our contribution, the known tools were based on iterative methods that, by their nature, cannot take advantage of the structure of the systems.

Our work can be extended in several directions. For instance, the notion of safeness can be easily adapted to other contexts where bisimilarity pseudometrics have a fixed point characterization. In the same spirit, one may obtain a sufficient condition that ensures continuity of operators, which is the natural generalization of non-extensiveness.

# Chapter 14

## A Distance for Continuous-Time Markov Chains

### 14.1 Introduction

Continuous-time Markov chains (CTMCs) are one of the most prominent models in performance and dependability analysis. They are exploited in a broad range of applications, and constitute the underlying semantics of many modeling formalisms for real-time probabilistic systems such as Markovian queuing networks, stochastic process algebras, and calculi for systems biology.

An example of CTMC is presented in Figure 14.1(left). Here, state  $s_1$  goes to state  $s_3$  and  $s_4$  with probability  $\frac{1}{3}$  and  $\frac{2}{3}$ , respectively. Each state has an associated *exit-rate* representing the rate of an exponentially distributed random variable that characterizes the residence-time in the state.

For example, the probability to move from  $s_1$  to any other state within time  $t \geq 0$  is given by

$$\int_0^t 3e^{-3x} dx = 1 - e^{-3t}.$$

A state with no outgoing transitions (as  $s_3$  in Figure 14.1) is called an *absorbing state*, and it represents a terminating state of the system.

The concept of *probabilistic bisimulation*, introduced by Larsen and Skou for discrete-time Markov chains (MCs) [32], has been extended to several type of probabilistic systems, including CTMCs. In Figure 14.1(left)  $s_4$  and  $s_5$  are bisimilar. Moreover, although  $s_1$  and  $s_2$  move with different probabilities to state  $s_4$  and  $s_5$ , their probabilities to reach any bisimilarity class is the same; hence, also  $s_1$  and  $s_2$  are bisimilar.

However, as with the Markovian models discussed in the previous Chapters of this monograph, also for CTMCs any behavioral analysis based on a notion of equivalence is too fragile, as it only relates processes with identical behaviors. This issue is illustrated in Figure 14.1(right), where the states  $t_1$  and  $t_2$  (i.e., the counterpart of  $s_1$  and  $s_2$ ,

respectively, after a perturbation of the transition probabilities) are not bisimilar. A similar situation occurs considering perturbations on the exit-rates or on associated labels, if one assumes they are taken from a metric space.

In this Chapter we introduce a bisimilarity pseudometric over CTMCs that extends that of Desharnais et al. over MCs [22], and we consider the problem of computing it both from a theoretical and practical point of view.

We show that the proposed distance can be computed in polynomial time in the size of the CTMC. This is obtained by reducing the problem of computing the distance to that of finding an optimal solution of a linear program that can be solved using the ellipsoid method. Differently from the proposal of [14], our linear program characterization has a number of constraints that is bounded by a polynomial in the size of the CTMC. This, in particular, allows us to avoid the use of the ellipsoid algorithm in favor of the simplex or the interior point methods.

Also in this case, the linear program characterization turns out to be inefficient in practice, even for small CTMCs. Nevertheless, supported by the encouraging results in our previous work [5] summarized in the previous Chapters, we propose to follow an on-the-fly approach for computing the distance. This is inspired by an alternative characterization of the bisimilarity pseudometric based on the notion of *coupling structure* for a CTMC.

Each coupling structure is associated with a *discrepancy function* that represents an over-approximation of the distance. The problem of computing the pseudometric is then reduced to that of searching for an *optimal* coupling structure whose associated discrepancy coincides with the distance.

The exploration of the coupling structures is based on a greedy strategy that, given a coupling structure, moves to a new one by ensuring an actual improvement of the current discrepancy function. This strategy will eventually find an optimal coupling structure. The method is sound independently from the initial starting coupling structure.

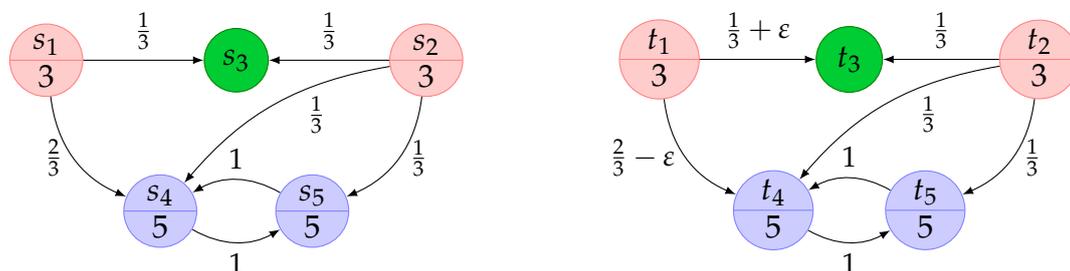


Figure 14.1: A CTMC (left) and an  $\epsilon$ -perturbation of it (right), for some  $\epsilon \in (0, \frac{2}{3})$ . Labels are represented by different colors; states are additionally labelled with their exit rates; and transitions with probability 0 are omitted.

Notably, the moving strategy is based on a *local* update of the current coupling structure. Since the update is local, when the goal is to compute the distance only between certain pairs of states, the construction of the coupling structures can be done *on-the-fly*, delimiting the exploration only to those states that are demanded during the computation.

The efficiency of our algorithm has been evaluated on a significant set of randomly generated CTMCs. The results show that our algorithm performs orders of magnitude better than the corresponding iterative and linear program implementations. Moreover, we provide empirical evidence that our algorithm enjoys good execution running times.

One of the main practical advantages of our approach consists in that one can focus on computing only the distances between states that are of particular interest. This is useful in practice, for instance when large systems are considered and visiting the entire state space is computationally expensive. A similar issue has been considered by Comanici et al., in [19] in the case of Markov decision processes with rewards, who noticed that for computing the approximated pseudometric one does not need to update the current value for all the pairs at each iteration, but it is sufficient only to focus on the pairs where changes are happening rapidly.

In our approach, the termination condition is checked locally, still ensuring that the local optimum corresponds to the global one.

Our methods can also be used in combination with approximation techniques as, for instance, to provide a least over-approximation of the behavioral distance given over-estimates of some particular distances.

## 14.2 CTMCs and Bisimilarity Pseudometrics

We recall the definitions of (finite)  $L$ -labelled continuous-time Markov chains (CTMCs) for a nonempty set of labels  $L$ , and *stochastic bisimilarity* over them. Then, we introduce a *behavioral pseudometric* over CTMCs to be considered as a quantitative generalization of stochastic bisimilarity.

Given a set  $X$ , a discrete probability distribution over it is a finitely supported function  $\mu: X \rightarrow [0, 1]$  such that  $\mu(X) = 1$ , where  $\mu(E) = \sum_{x \in E} \mu(x)$ , for  $E \subseteq X$ . We denote the set of discrete probability distributions over  $X$  by  $\mathcal{D}(X)$ .

**Definition 14.2.1** (Continuous-time Markov chain). *An  $L$ -labelled continuous-time Markov chain is a tuple  $\mathcal{M} = (S, A, \tau, \rho, \ell)$  consisting of*

- a countable nonempty finite set  $S$  of states,
- a set  $A \subseteq S$  of absorbing states,
- a transition probability function  $\tau: S \setminus A \rightarrow \mathcal{D}(S)$ ,

- an exit rate function  $\rho: S \setminus A \rightarrow \mathbb{R}_{>0}$ , and
- a labeling function  $\ell: S \rightarrow L$ .

■

The labels in  $L$  represent properties of interest that hold in a particular state according to the labeling function  $\ell: S \rightarrow L$ .

If  $s \in S$  is the current state of the system and  $E \subseteq S$  is a subset of states,  $\tau(s)(E) \in [0, 1]$  corresponds to the probability that a transition from  $s$  to arbitrary  $s' \in E$  is taken, and  $\rho(s) \in \mathbb{R}_{>0}$  represents the rate of an exponentially distributed random variable that characterizes the residence time in the state  $s$  before any transition is taken. Therefore, the probability to make a transition from state  $s$  to any  $s' \in E$  within time unit  $t \in \mathbb{R}_{\geq 0}$  is given by

$$\tau(s)(E) \cdot \text{Exp}(\rho(s))([0, t]),$$

where  $\text{Exp}(r)(E) = \int_E r e^{-rx} dx$ , for any Borel subset  $E \subseteq \mathbb{R}_{\geq 0}$  and  $r > 0$ .

Absorbing states in  $A \subseteq S$  are used to represent termination or deadlock states. An example of CTMC is shown in Figure 14.1.

For discrete-time Markov chains, the standard notion of behavioral equivalence is probabilistic bisimulation of Larsen and Skou [32]. The following definition extends it to CTMCs.

To ease the notation, for  $A \subseteq S$ , we introduce the relation  $\equiv_A \subseteq S \times S$  defined by

$$s \equiv_A s' \text{ if either } s, s' \in A \text{ or } s, s' \notin A.$$

As before in this monograph, for an equivalence relation  $R$  on a set  $S$  we denote by  $S/R$  the set of  $R$ -closed sets. Hence, each element of  $S/R$  is a union of  $R$ -equivalence classes.

**Definition 14.2.2** (Stochastic Bisimulation). *Let  $\mathcal{M} = (S, A, \tau, \rho, \ell)$  be a CTMC. An equivalence relation  $R \subseteq S \times S$  is a stochastic bisimulation on  $\mathcal{M}$  if whenever  $s R t$ , then*

1.  $s \equiv_A t$ ,  $\ell(s) = \ell(t)$ , and
2. if  $s, t \in A$ , then  $\rho(s) = \rho(t)$  and, for all  $C \in S/R$ ,  $\tau(s)(C) = \tau(t)(C)$ .

*Two states  $s, t \in S$  are bisimilar with respect to  $\mathcal{M}$ , written  $s \sim_{\mathcal{M}} t$ , if they are related by some probabilistic bisimulation on  $\mathcal{M}$ .*

Intuitively, two states are bisimilar if they have the same labels, they are simultaneously absorbing or non-absorbing states and, in the case they are non-absorbing, their residence-time distributions and probability of moving by a single transition to any given class of bisimilar states is always the same.

As an example of two stochastic bisimilar states, consider  $s_1$  and  $s_2$  in the CTMC depicted on the left hand side of in Figure 14.1. A bisimulation relation that relates them is the equivalence relation with equivalence classes given by  $\{s_1, s_2\}$ ,  $\{s_3\}$ , and  $\{s_4, s_5\}$ .

### 14.2.1 Bisimilarity Pseudometrics on CTMCs

In this section, we introduce a family of pseudometrics on CTMCs parametric in a discount factor  $\lambda \in (0, 1)$ .

Following the approach of [46], given a CTMC  $\mathcal{M} = (S, A, \tau, \rho, \ell)$  we define a (1-bounded) pseudometric on  $S$  as the least fixed point of an operator on the set  $[0, 1]^{S \times S}$  of functions from  $S \times S$  to  $[0, 1]$ . This pseudometric is then shown to be adequate with respect to stochastic bisimilarity: we prove that two states are stochastic bisimilar if and only if they have distance zero.

The operator we are going to introduce will use three key ingredients:

- a 1-bounded metric<sup>1</sup>  $d_L: L \times L \rightarrow [0, 1]$  on the set of labels;
- a distance between residence-time distributions;
- a distance between transition distributions.

The first is meant to measure the *static* differences with respect to the labels associated with the states; the last two are meant to capture the differences in the *dynamics* with respect to the continuous and discrete probabilistic choices, respectively.

To this end, we consider two distances over probability distributions. The first one is the, so called, *total variation metric*, defined for arbitrary distributions  $\mu, \nu \in \mathcal{D}(\mathbb{R}_{\geq 0})$  as follows.

$$\|\mu - \nu\|_{\text{TV}} = \sup_E |\mu(E) - \nu(E)|,$$

where the supremum is taken over the Borel measurable sets of  $\mathbb{R}_{\geq 0}$ .

The second one is the *Kantorovich distance*, based on the notion of *coupling* of probability measures that we already introduced in the previous Chapters. Hereafter, we recall this definition for the case of probability distributions over finite sets.

**Definition 14.2.3 (Coupling).** *Let  $S$  be a finite set, and let  $\mu, \nu \in \mathcal{D}(S)$ . A probability distribution  $\omega \in \mathcal{D}(S \times S)$  is called a coupling for  $(\mu, \nu)$  if for arbitrary  $u, v \in S$*

$$\sum_{v \in S} \omega(u, v) = \mu(u) \quad \text{and} \quad \sum_{u \in S} \omega(u, v) = \nu(v).$$

In other words,  $\omega$  is a joint probability distribution with  $\mu$  and  $\nu$  being the left and the right marginal, respectively.

We denote by  $\Omega(\mu, \nu)$ , the set of couplings for  $(\mu, \nu)$ .

---

<sup>1</sup>Since the set  $S$  of states is assumed to be finite, one may assume the set of labels is finite as well. Thus, the metric  $d_L$  on labels can be bounded without loss of generality.

For a finite set  $S$  and a 1-bounded distance  $d: S \times S \rightarrow [0, 1]$  over it, the *Kantorovich distance* is defined, for arbitrary distributions  $\mu, \nu \in \mathcal{D}(S)$  as follows

$$\mathcal{K}_d(\mu, \nu) = \min \left\{ \sum_{u, v \in S} d(u, v) \cdot \omega(u, v) \mid \omega \in \Omega(\mu, \nu) \right\}.$$

Intuitively,  $\mathcal{K}_d$  lifts a (1-bounded) distance over  $S$  to a (1-bounded) distance over its probability distributions. It can be easily verified that  $\mathcal{K}_d$  is a (pseudo)metric whenever  $d$  is a (pseudo)metric.

Now, consider the following functional operator.

**Definition 14.2.4.** Let  $\mathcal{M} = (S, A, \tau, \rho, \ell)$  be a CTMC and  $\lambda \in (0, 1)$  a discount factor. The function  $\Delta_\lambda^{\mathcal{M}}: [0, 1]^{S \times S} \rightarrow [0, 1]^{S \times S}$  is defined, for arbitrary  $d: S \times S \rightarrow [0, 1]$  and  $s, t \in S$ , as follows.

$$\Delta_\lambda^{\mathcal{M}}(d)(s, t) = \begin{cases} 1 & \text{if } s \notin_A t \\ \mathcal{L}(s, t) & \text{if } s, t \in A \\ \max\{\mathcal{L}(s, t), \lambda \cdot \mathcal{T}(d)(s, t)\} & \text{if } s, t \notin A \end{cases}$$

where  $\mathcal{T}: [0, 1]^{S \times S} \rightarrow [0, 1]^{S \times S}$  and  $\mathcal{L}, \mathcal{E}: S \times S \rightarrow [0, 1]$  are defined by

$$\mathcal{T}(d)(s, t) = \mathcal{E}(s, t) + (1 - \mathcal{E}(s, t)) \cdot \mathcal{K}_d(\tau(s), \tau(t)),$$

$$\mathcal{L}(s, t) = d_L(\ell(s), \ell(t)),$$

$$\mathcal{E}(s, t) = \|\text{Exp}(\rho(s)) - \text{Exp}(\rho(t))\|_{\text{TV}}.$$

The functional  $\Delta_\lambda^{\mathcal{M}}$  measures the difference of two states with respect to: their labels (by means of the pseudometric  $\mathcal{L}$ ), their residence-time distributions (by means of the pseudometric  $\mathcal{E}$ ), and their discrete probabilities to move to the next state (by means of the Kantorovich distance).

If two states are not absorbing or non-absorbing in the same time, they are considered incomparable, and their distance is set to 1.

If both states are absorbing, they express no dynamic behaviour, hence they are compared statically, and their distance corresponds to that occurring between their labels.

Finally, if the states are non-absorbing, then they are compared with respect to both their static and dynamic features, namely, taking the maximum among their respective associated distances.

Specifically, the value  $\mathcal{E}(s, t)$  corresponds to the least probability that two transitions are taken independently from the states  $s$  and  $t$  at different moments in time. This value is used by the functional  $\mathcal{T}$  to measure the overall differences that might occur in the dynamics of the two states in combination with the Kantorovich distance between their transition probability distributions.

The set  $[0, 1]^{S \times S}$  is endowed with the partial order  $\sqsubseteq$  defined by

$$d \sqsubseteq d' \text{ iff } d(s, t) \leq d'(s, t),$$

for all  $s, t \in S$  and it forms a complete lattice. The bottom element 0 is the constant 0 function, while the top element is the constant 1 function.

For any subset  $D \subseteq [0, 1]^{S \times S}$ , the least upper bound  $\bigsqcup D$ , and greatest lower bound  $\bigsqcap D$  are, respectively, given by

$$(\bigsqcup D)(s, t) = \sup_{d \in D} d(s, t)$$

and

$$(\bigsqcap D)(s, t) = \inf_{d \in D} d(s, t),$$

for all  $s, t \in S$ .

It is easy to check that, for any  $\mathcal{M}$  and  $\lambda \in (0, 1)$ ,  $\Delta_\lambda^{\mathcal{M}}$  is monotone, i.e.,

$$\text{if } d \sqsubseteq d', \text{ then } \Delta_\lambda^{\mathcal{M}}(d) \sqsubseteq \Delta_\lambda^{\mathcal{M}}(d'),$$

thus, since  $([0, 1]^{S \times S}, \sqsubseteq)$  is a complete lattice, Tarski's fixed point theorem guarantees that  $\Delta_\lambda^{\mathcal{M}}$  admits least and greatest fixed points.

**Definition 14.2.5** (Bisimilarity distance). *Let  $\mathcal{M}$  be a CTMC and  $\lambda \in (0, 1)$ . The  $\lambda$ -discounted bisimilarity pseudometric on  $\mathcal{M}$ , denoted by  $\delta_\lambda^{\mathcal{M}}$ , is the least fixed point of  $\Delta_\lambda^{\mathcal{M}}$ .*

Now we will show that the least fixed point  $\delta_\lambda^{\mathcal{M}}$  is indeed a pseudometric and, moreover, it is adequate with respect to stochastic bisimilarity (Theorem 14.2.9). This justifies the definition above.

To this end, we need some technical lemmas. In particular, we prove that  $\Delta_\lambda^{\mathcal{M}}$  preserves pseudometrics (Lemma 14.2.6) and it is non-expansive (Lemma 14.2.7).

Hereafter, unless mentioned otherwise, we fix a CTMC  $\mathcal{M} = (S, A, \tau, \rho, \ell)$  and a discount factor  $\lambda \in (0, 1)$ . To ease the notation, whenever  $\mathcal{M}$  is clear from the context, instead of  $\Delta_\lambda^{\mathcal{M}}$ ,  $\delta_\lambda^{\mathcal{M}}$  or  $\sim_{\mathcal{M}}$  we will simply use  $\Delta_\lambda$ ,  $\delta_\lambda$  and  $\sim$ , respectively.

**Lemma 14.2.6.** *The operator  $\Delta_\lambda$  preserves pseudometrics.*

*Proof.* Let  $d: S \times S \rightarrow [0, 1]$  be a pseudometric. We want to prove that  $\Delta_\lambda(d)$  is a pseudometric as well.

Recall that  $\mathcal{L}, \mathcal{E}: S \times S \rightarrow [0, 1]$  are pseudometrics. Thus, since the point wise maximum of pseudometrics is a pseudometric, it suffices to prove that  $\mathcal{T}$  preserves pseudometrics.

Recall also that  $\mathcal{K}_d: \mathcal{D}(S) \times \mathcal{D}(S) \rightarrow [0, 1]$  is a pseudometric, since  $d$  is. Thus, reflexivity and symmetry are immediate.

The only nontrivial case is triangular inequality.

Let  $s, t, u \in S$ , we want to prove  $\mathcal{T}(d)(s, t) \leq \mathcal{T}(d)(s, u) + \mathcal{T}(d)(u, t)$ .

First note that, for any  $0 \leq \alpha, \beta \leq 1$  and  $\alpha' \geq \alpha$ , the following holds:

$$\begin{aligned} \alpha + (1 - \alpha)\beta &= \beta - \beta + \alpha + (1 - \alpha)\beta \\ &= \beta - \alpha\beta - (1 - \alpha)\beta + \alpha + (1 - \alpha)\beta && (0 \leq \alpha \leq 1) \\ &= \beta - \alpha\beta + \alpha = \beta + (1 - \beta)\alpha \\ &\leq \beta + (1 - \beta)\alpha' && (\alpha \leq \alpha' \text{ and } 0 \leq \beta \leq 1) \\ &= \alpha' + (1 - \alpha')\beta. \end{aligned}$$

Thus, since  $\mathcal{E}$  is a pseudometric, by triangular inequality and the above we have

$$\begin{aligned} \mathcal{T}(d)(s, t) &= \mathcal{E}(s, t) + (1 - \mathcal{E}(s, t)) \cdot \mathcal{K}_d(\tau(s), \tau(t)) && (\text{def. } \mathcal{T}) \\ &\leq \mathcal{E}(s, u) + \mathcal{E}(u, t) + (1 - (\mathcal{E}(s, u) + \mathcal{E}(u, t))) \cdot \mathcal{K}_d(\tau(s), \tau(t)) && (*) \end{aligned}$$

If we show that the last summand in (\*) is smaller or equal than the sum of  $(1 - \mathcal{E}(s, u)) \cdot \mathcal{K}_d(\tau(s), \tau(u))$  and  $(1 - \mathcal{E}(u, t)) \cdot \mathcal{K}_d(\tau(u), \tau(t))$ , we get the following

$$\begin{aligned} &\leq \mathcal{E}(s, u) + (1 - \mathcal{E}(s, u)) \cdot \mathcal{K}_d(\tau(s), \tau(u)) + \mathcal{E}(u, t) + (1 - \mathcal{E}(u, t)) \cdot \mathcal{K}_d(\tau(u), \tau(t)) \\ &= \mathcal{T}(d)(s, t) + \mathcal{T}(d)(s, t). && (\text{def. } \mathcal{T}) \end{aligned}$$

To this end, consider two cases.

If  $\mathcal{E}(s, u) + \mathcal{E}(u, t) > 1$  then the inequality holds trivially, since  $1 - (\mathcal{E}(s, u) + \mathcal{E}(u, t)) < 0$ , so that the last summand in (\*) is negative.

If  $\mathcal{E}(s, u) + \mathcal{E}(u, t) \leq 1$ , then  $1 - (\mathcal{E}(s, u) + \mathcal{E}(u, t)) \geq 0$ , so we have

$$\begin{aligned} &(1 - (\mathcal{E}(s, u) + \mathcal{E}(u, t))) \cdot \mathcal{K}_d(\tau(s), \tau(t)) \\ &\leq (1 - (\mathcal{E}(s, u) + \mathcal{E}(u, t))) \cdot (\mathcal{K}_d(\tau(s), \tau(u)) + \mathcal{K}_d(\tau(u), \tau(t))) && (\text{triang. } \mathcal{K}_d) \\ &= (1 - (\mathcal{E}(s, u) + \mathcal{E}(u, t))) \cdot \mathcal{K}_d(\tau(s), \tau(u)) + (1 - (\mathcal{E}(s, u) + \mathcal{E}(u, t))) \cdot \mathcal{K}_d(\tau(u), \tau(t)) \\ &\leq (1 - \mathcal{E}(s, u)) \cdot \mathcal{K}_d(\tau(s), \tau(u)) + (1 - \mathcal{E}(u, t)) \cdot \mathcal{K}_d(\tau(u), \tau(t)) \end{aligned}$$

and we are done.  $\square$

The set  $[0, 1]^{S \times S}$  can be turned into a metric space by means of the supremum norm

$$\|d - d'\| = \sup_{s, t \in S} |d(s, t) - d'(s, t)|.$$

Next we show that the  $\lambda$ -discounted functional operator  $\Delta_\lambda$  is  $\lambda$ -non-expansive, that is

$$\|\Delta_\lambda(d') - \Delta_\lambda(d)\| \leq \lambda \cdot \|d' - d\|,$$

for any  $d, d' \in [0, 1]^{S \times S}$ .

**Lemma 14.2.7.** *The operator  $\Delta_\lambda$  is  $\lambda$ -non-expansive*

*Proof.* By [45, Corollary 1], to prove that  $\Delta_\lambda$  is  $\lambda$ -non-expansive it suffices to show that, whenever  $d \sqsubseteq d'$  then, for all  $s, t \in S$ ,

$$\Delta_\lambda(d')(s, t) - \Delta_\lambda(d)(s, t) \leq \lambda \cdot \|d' - d\|.$$

The only nontrivial case is when  $s, t \notin A$  and  $\mathcal{L}(s, t) < \lambda \cdot \mathcal{T}(d)(s, t)$ .

Assume that for some  $\omega \in \Omega(\tau(s), \tau(t))$ ,

$$\mathcal{K}_d(s, t)(\tau(s), \tau(t)) = \sum_{u, v \in S} d(u, v) \cdot \omega(u, v).$$

Then, we have

$$\begin{aligned} \Delta_\lambda(d')(s, t) - \Delta_\lambda(d)(s, t) &\leq \lambda \cdot (\mathcal{T}(d')(s, s') - \mathcal{T}(d)(s, s')) \\ &\leq \lambda \cdot (\mathcal{K}_{d'}(s, t) - \mathcal{K}_d(s, t)) \\ &\leq \lambda \cdot \left( \sum_{u, v \in S} d'(u, v) \cdot \omega(u, v) - \sum_{u, v \in S} d(u, v) \cdot \omega(u, v) \right) \\ &= \lambda \cdot \left( \sum_{u, v \in S} (d'(u, v) - d(u, v)) \cdot \omega(u, v) \right) \\ &\leq \lambda \cdot \left( \sum_{u, v \in S} \|d' - d\| \cdot \omega(u, v) \right) \\ &= \lambda \cdot (\|d' - d\| \cdot \sum_{u, v \in S} \omega(u, v)) \\ &= \lambda \cdot \|d' - d\|. \end{aligned}$$

□

It is standard that  $[0, 1]^{S \times S}$  with the supremum norm forms a complete metric space (i.e, every Cauchy sequence converges). Therefore, since  $\lambda \in (0, 1)$ , a direct consequence of Lemma 14.2.7 and Banach's fixed point theorem is the following.

**Theorem 14.2.8.** *For any  $\lambda \in (0, 1)$ ,  $\delta_\lambda$  is the unique fixed point of  $\Delta_\lambda$ . Moreover, for any  $n \in \mathcal{N}$  and  $d: S \times S \rightarrow [0, 1]$  the following inequality holds.*

$$\|\delta_\lambda - \Delta_\lambda^n(d)\| \leq \frac{\lambda^n}{1 - \lambda} \|\Delta_\lambda(d) - d\|.$$

Now we are ready to state the main theorem of this section.

**Theorem 14.2.9** (Bisimilarity pseudometric). *The least fixed point  $\delta_\lambda$  of  $\Delta_\lambda$  is a pseudometric. Moreover, for any  $s, t \in S$ ,*

$$s \sim t \text{ iff } \delta_\lambda(s, t) = 0.$$

*Proof.* We first prove that  $\delta_\lambda$  is a pseudometric.

By Lemma 14.2.7 and Banach's fixed point theorem,

$$\delta_\lambda = \bigsqcup_{n \in \mathcal{N}} \Delta_\lambda^n(\mathbf{0}).$$

Clearly,  $\mathbf{0}$  is a pseudometric. Thus, by Lemma 14.2.6, a simple induction on  $n$  shows that, for all  $n \in \mathcal{N}$ ,  $\Delta_\lambda^n(\mathbf{0})$  is a pseudometric.

Since the least upper bound with respect to  $\sqsubseteq$  preserves pseudometrics, we have that  $\delta_\lambda$  is a pseudometric too.

Now we are left to prove that, for any  $s, t \in S$ ,

$$s \sim t \text{ iff } \delta_\lambda(s, t) = 0.$$

( $\Rightarrow$ ) We prove that  $R = \{(s, s') \mid \delta_\lambda(s, t) = 0\}$  is a stochastic bisimulation. Clearly,  $R$  is an equivalence.

Assume  $(s, s') \in R$ , then, by definition of  $\Delta_\lambda$ , one of the following holds:

1.  $s, t \in A$  and  $\mathcal{L}(s, t) = 0$ ;
2.  $s, t \notin A$ ,  $\mathcal{L}(s, t) = 0$ , and  $\mathcal{T}(\delta_\lambda)(s, t) = 0$ .

If (1) holds, from  $\mathcal{L}(s, t) = 0$ , we get that  $\ell(s) = \ell(t)$ .

If (2) holds, then we have  $\mathcal{E}(s, t) = 0$  and  $\mathcal{K}_{\delta_\lambda}(\tau(s), \tau(t)) = 0$ . Since  $\mathcal{E}(s, t) = 0$  we get  $\text{Exp}(\rho(s)) = \text{Exp}(\rho(t))$  and hence  $\rho(s) = \rho(t)$ . By [25, Lemma 3.1],  $\mathcal{K}_{\delta_\lambda}(\tau(s), \tau(t)) = 0$  implies that, for all  $C \in S/R$ ,  $\tau(s)(C) = \tau(t)(C)$ . Therefore  $R$  is a bisimulation.

( $\Leftarrow$ ) Let  $R \subseteq S \times S$  be a stochastic bisimulation on  $\mathcal{M}$ , and define  $d_R: S \times S \rightarrow [0, 1]$  by  $d_R(s, t) = 0$  if  $(s, t) \in R$  and  $d_R(s, t) = 1$  if  $(s, t) \notin R$ .

We show that  $\Delta_\lambda(d_R) \sqsubseteq d_R$ .

If  $(s, t) \notin R$ , then  $d_R(s, t) = 1 \geq \Delta_\lambda(d_R)(s, t)$ .

If  $(s, t) \in R$ , then  $\ell(s) = \ell(t)$  and one of the following holds:

1.  $s, t \in A$ ;
2.  $s, t \notin A$ ,  $\rho(s) = \rho(t)$  and,  $\forall C \in S/R$ .  $\tau(s)(C) = \tau(t)(C)$ .

If (1) holds,  $\Delta_\lambda(d_R)(s, t) = \mathcal{L}(s, t) = 0 = d_R(s, t)$ .

If (2) holds, by [25, Lemma 3.1] and the fact that, for all  $C \in S/R$ ,  $\tau(s)(C) = \tau(t)(C)$ , we have  $\mathcal{K}_{d_R}(\tau(s), \tau(t)) = 0$ . Moreover  $\mathcal{E}(s, t) = 0$ . This gives us

$$\Delta_\lambda(d_R)(s, t) = 0 = d_R(s, t).$$

By the generality of the bisimulation relation  $R$  and by Tarski's fixed point theorem, we have that  $s \sim s'$  implies  $\delta_\lambda(s, s') = 0$ .  $\square$

## 14.3 Complexity and Linear Programming representation

In this section, we study the problem of computing the bisimilarity distance by considering two different approaches.

The first approach is an iterative method that approximates  $\delta_\lambda$  from below (resp. above) successively applying the operator  $\Delta_\lambda$  starting from the least (resp. greatest) element in  $[0, 1]^{S \times S}$ .

The second approach is based on a linear program characterization of  $\delta_\lambda$  that is based on the Kantorovich duality [48]. In contrast to an analogous proposal in [14], our linear program has a number of constraints that is polynomially bounded in the size of the CTMC. As a consequence, the bisimilarity distance  $\delta_\lambda$  can be computed in polynomial time in the size of the CTMC.

### 14.3.1 Iterative method

From Theorem 14.2.8 it follows that in order to get  $\varepsilon$ -close to  $\delta_\lambda$ , for any  $\varepsilon > 0$ , it is sufficient to iterate the application of the fixed point operator  $\lceil \log_\lambda \varepsilon \rceil$  times.

**Proposition 14.3.1.** *For any  $\varepsilon > 0$  and  $d: S \times S \rightarrow [0, 1]$ ,*

$$\|\delta_\lambda - \Delta_\lambda^{\lceil \log_\lambda \varepsilon \rceil}(d)\| \leq \varepsilon.$$

*Proof.* From Theorem 14.2.8 we have

$$\|\delta_\lambda - \Delta_\lambda^n(d)\| \leq \frac{\lambda^n}{1-\lambda} \|\Delta_\lambda(d) - d\|$$

and since  $\|\Delta_\lambda(d) - d\| \leq 1$ , we get  $\|\delta_\lambda - \Delta_\lambda^n(d)\| \leq \frac{\lambda^n}{1-\lambda}$ .

For  $n = \log_\lambda(\varepsilon - \varepsilon\lambda)$ , we have  $\varepsilon = \frac{\lambda^n}{1-\lambda}$ . Therefore, using Lemma 14.2.7 and the fact that  $\lceil \log_\lambda \varepsilon \rceil \geq \log_\lambda(\varepsilon - \varepsilon\lambda)$ , we obtain

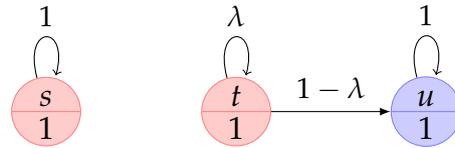
$$\|\delta_\lambda - \Delta_\lambda^{\lceil \log_\lambda \varepsilon \rceil}(d)\| \leq \varepsilon.$$

□

The above result provides us with a simple method for approximating  $\delta_\lambda$ . If the starting point is  $\mathbf{0}$  we obtain an under-approximation, whereas starting from  $\mathbf{1}$  we get an over-approximation. Both the approximations can be taken arbitrary close to the exact value.

However, as shown in the following example, the exact distance value cannot be reached in general. This holds for any discount factor.

**Example 14.3.2** ([14]). *Consider the {red, blue}-labeled CTMC represented in the figure below.*



Let  $d_L: L \times L \rightarrow [0, 1]$  be the discrete metric over  $L$ , defined as  $d_L(l, l') = 0$  if  $l = l'$  and 1 otherwise. One can check that  $\delta_\lambda(s, t) = \frac{\lambda - \lambda^2}{1 - \lambda^2}$  and, for all  $n \in \mathcal{N}$ ,

$$\Delta_\lambda^n(\mathbf{0})(s, t) \leq \frac{\lambda - \lambda^{2n+1}}{1 + \lambda}.$$

Since, for all  $n \in \mathcal{N}$ ,

$$\frac{\lambda - \lambda^{2n+1}}{1 + \lambda} < \frac{\lambda - \lambda^2}{1 - \lambda^2},$$

we have that the fixed point cannot be reached in a finite number of iterations.  $\blacksquare$

In [14] it is shown that the bisimilarity distance of Desharnais et al. [22] can be computed *exactly* by iterating the fixed point operator up to a precision that allows one to use the continued fraction algorithm to yield the exact value of the fixed point. This method can be applied provided that the pseudometric has rational values. In their case, this is ensured assuming that the transition probabilities are rational.

Unfortunately, in our case this cannot be ensured under the same conditions. Indeed, the total variation distance between exponential distributions with rates  $r, r' > 0$  is analytically solved as follows

$$\|Exp(r) - Exp(r')\|_{TV} = \begin{cases} 0 & \text{if } r = r' \\ \left| \left(\frac{r'}{r}\right)^{\frac{r}{r-r'}} - \left(\frac{r'}{r}\right)^{\frac{r'}{r-r'}} \right| & \text{otherwise} \end{cases} \quad (14.3.1)$$

Thus, even restricting to rational exit-rates and probabilities, the distance may have irrational values. As a consequence, we cannot compute, in general, the exact distance values.

### 14.3.2 Linear Program Characterization

Our linear program characterization leverages on two key results. The first one is the uniqueness of the fixed point of  $\Delta_\lambda$  (Theorem 14.2.8). The second one is a *dual* linear program characterization of the Kantorovich distance.

For  $S$  finite,  $d: S \times S \rightarrow [0, 1]$ , and  $\mu, \nu \in \mathcal{D}(S)$ , the value  $\mathcal{K}_d(\mu, \nu)$  coincides with the optimal value of the following linear program

$$\begin{aligned} \mathcal{K}_d(\mu, \nu) = \min_{\omega} \sum_{s,t \in S} d(u, v) \cdot \omega_{u,v} & \quad (14.3.2) \\ \sum_v \omega_{u,v} = \mu(u) & \quad \forall u \in S \\ \sum_u \omega_{u,v} = \nu(v) & \quad \forall v \in S. \end{aligned}$$

By a standard argument in linear optimization, the above can be alternatively represented by the following dual linear program

$$\begin{aligned} \mathcal{K}_d(\mu, \nu) = \max_y \sum_{u \in S} (\mu(u) - \nu(u)) \cdot y_u & \quad (14.3.3) \\ y_u - y_v \leq d(u, v) & \quad \forall u, v \in S. \end{aligned}$$

This alternative characterization is a special case of a more general result commonly known as the *Kantorovich duality* and largely studied in linear optimization theory (see [48]).

Let  $n = |S|$  and  $m = n - |A|$ . Consider the linear program in Figure 14.2, hereafter denoted by  $D_\lambda(\mathcal{M})$ , with variables  $d \in \mathbb{R}^{n^2}$ ,  $y \in \mathbb{R}^{m^2+n}$  and  $k, m \in \mathbb{R}^{m^2}$ .

The objective function of  $D_\lambda(\mathcal{M})$  attains its optimal value when  $k$  and  $m$  are maximized in every component. Therefore, according to (14.3.3), an optimal solution  $(d^*, y^*, k^*, m^*) \in D_\lambda(\mathcal{M})$  satisfies the following equalities

$$\begin{aligned} \forall s, t \notin A. \quad m_{s,t}^* &= \min\{\mathcal{L}(s, t), \lambda(\mathcal{E}(s, t) + (1 - \mathcal{E}(s, t))\mathcal{K}_{d^*}(\tau(s), \tau(t)))\} \\ \forall s, t \notin A. \quad k_{s,t}^* &= \mathcal{K}_{d^*}(\tau(s), \tau(t)) \end{aligned}$$

By the above equalities and by the feasibility of the solution, it follows that  $d^*$  is a fixed point of  $\Delta_\lambda$ . Since the distance is the unique fixed point of  $\Delta_\lambda$ ,  $d^* = \delta_\lambda$ .

**Theorem 14.3.3.** *Let  $\mathcal{M}$  be a CTMC,  $\lambda \in (0, 1)$  be a discount factor, and  $(d^*, y^*, k^*, m^*)$  be an optimal solution of  $D_\lambda(\mathcal{M})$ . Then, for all  $s, t \in S$ ,  $d_{s,t}^* = \delta_\lambda(s, t)$ .*

*Proof.* Let  $(d^*, y^*, k^*, m^*)$  be an optimal solution of  $D_\lambda(\mathcal{M})$ , and consider the following linear program, hereafter denoted by  $D'_\lambda$ .

$$\begin{aligned} \arg \max_{y, k, m} \sum_{s,t \notin A} k_{s,t} + m_{s,t} & \\ m_{s,t} \leq \mathcal{L}(s, t) & \quad \forall s, t \notin A \\ m_{s,t} \leq \lambda(\mathcal{E}(s, t) + (1 - \mathcal{E}(s, t))k_{s,t}) & \quad \forall s, t \notin A \\ k_{s,t} = \sum_{u \in S} (\tau(s)(u) - \tau(t)(u)) \cdot y_u^{s,t} & \quad \forall s, t \notin A \\ y_u^{s,t} - y_v^{s,t} \leq d_{u,v}^* & \quad \forall s, t \notin A, \forall u, v \in S \end{aligned}$$

$\arg \max_{d,y,k,m} \sum_{s,t \notin A} k_{s,t} + m_{s,t}$	
$d_{s,t} = 1$	$\forall s, t \in S. s \neq_A t$
$d_{s,t} = \mathcal{L}(s, t)$	$\forall s, t \in A$
$d_{s,t} = \mathcal{L}(s, t) + \lambda(\mathcal{E}(s, t) + (1 - \mathcal{E}(s, t))k_{s,t}) - m_{s,t}$	$\forall s, t \notin A$
$m_{s,t} \leq \mathcal{L}(s, t)$	$\forall s, t \notin A$
$m_{s,t} \leq \lambda(\mathcal{E}(s, t) + (1 - \mathcal{E}(s, t))k_{s,t})$	$\forall s, t \notin A$
$k_{s,t} = \sum_{u \in S} (\tau(s)(u) - \tau(t)(u)) \cdot y_u^{s,t}$	$\forall s, t \notin A$
$y_u^{s,t} - y_v^{s,t} \leq d_{u,v}$	$\forall s, t \notin A, \forall u, v \in S$

Figure 14.2: Linear program characterization of  $\delta_\lambda$  for  $\lambda \in (0, 1)$ .

Since  $d^*$  is part of a feasible solution of  $D_\lambda(\mathcal{M})$  and the constraints of  $D'_\lambda$  are a subset of those of  $D_\lambda(\mathcal{M})$ , the optimal value for  $D'_\lambda$  is greater or equal than the one for  $D_\lambda(\mathcal{M})$ .

For each  $s, t \notin A$ , when  $k_{s,t}$  is maximal also  $m_{s,t}$  reaches its maximal value. Therefore, an optimal solution for  $D'_\lambda$  will be achieved when, for each  $s, t \notin A$ , the following holds

$$k_{s,t} = \max_y \sum_{u \in S} (\tau(s)(u) - \tau(t)(u)) \cdot y_u \quad (14.3.4)$$

$$y_u - y_v \leq d_{u,v}^* \quad \forall u, v \in S$$

By the Kantorowich duality (14.3.3), the optimal value of (14.3.4) is  $\mathcal{K}_{d^*}(\tau(s), \tau(t))$ . Since  $m_{s,t}$  is a lower bound for  $\{\mathcal{L}(s, t), \lambda(\mathcal{E}(s, t) + (1 - \mathcal{E}(s, t))k_{s,t})\}$ , we obtain an optimal solution for  $D'_\lambda$  by instantiating the variables as follows

$$\forall s, t \notin A. \quad m_{s,t} = \min\{\mathcal{L}(s, t), \lambda(\mathcal{E}(s, t) + (1 - \mathcal{E}(s, t))\mathcal{K}_{d^*}(\tau(s), \tau(t)))\} \quad (14.3.5)$$

$$\forall s, t \notin A. \quad k_{s,t} = \mathcal{K}_{d^*}(\tau(s), \tau(t)) \quad (14.3.6)$$

$$\forall s, t \notin A. \quad y^{s,t} \in \arg \max_y \sum_{u \in S} (\tau(s)(u) - \tau(t)(u)) \cdot y_u \quad (14.3.7)$$

$$y_u - y_v \leq d_{u,v}^* \quad \forall u, v \in S$$

We will show that the above instantiation is part of a feasible solution of  $D_\lambda(\mathcal{M})$ , provided that  $d^*$  is a fixed point for  $\Delta_\lambda$ . For any  $s, t \notin A$  the following holds

$$\begin{aligned} d_{s,t}^* &= \mathcal{L}(s, t) + \lambda(\mathcal{E}(s, t) + (1 - \mathcal{E}(s, t))k_{s,t}) - m_{s,t} && \text{(by def. } D_\lambda(\mathcal{M})) \\ &= \mathcal{L}(s, t) + \lambda(\mathcal{E}(s, t) + (1 - \mathcal{E}(s, t))\mathcal{K}_{d^*}(\tau(s), \tau(t))) - m_{s,t} && \text{(by (14.3.6))} \\ &= \mathcal{L}(s, t) + \mathcal{T}(d^*)(s, t) - m_{s,t} && \text{(by def. } \mathcal{T}) \\ &= \mathcal{L}(s, t) + \mathcal{T}(d^*)(s, t) - \min\{\mathcal{L}(s, t), \mathcal{T}(d^*)(s, t)\} && \text{(by (14.3.5) and def. } \mathcal{T}) \\ &= \max\{\mathcal{L}(s, t), \mathcal{T}(d^*)(s, t)\}, \end{aligned}$$

thus  $d^*$  is a fixed point of  $\Delta_\lambda$ . The thesis follows by Theorem 14.2.8.  $\square$

In [14] it has been shown that the bisimilarity distance of Desharnais et al. can be computed in polynomial time as the solution of a linear program that can be solved by using the ellipsoid method. However, in their proposal the number of constraints may be exponential in the size of the model. This is due to the fact that the Kantorovich distance is resolved listing all the couplings that correspond to vertices of the transportation polytopes involved in the definition of the distance [23].

In contrast, our proposal has a number of constraints and unknowns<sup>2</sup> bounded by  $4|S|^2$  and  $|S|^3 + 3|S|^2$ , respectively. This allows one to use general algorithms for solving LP problems (such as the *simplex* and the *interior point* methods) that, in practice, are more efficient than the ellipsoid method.

Moreover, this allows us to state the following complexity result.

**Theorem 14.3.4.**  $\delta_\lambda$  can be computed in polynomial-time in the size of  $\mathcal{M}$ .

*Proof.* Using Theorem 14.3.3,  $\delta_\lambda$  can be computed within the time it takes to construct and solve  $D_\lambda(\mathcal{M})$ .

$D_\lambda(\mathcal{M})$  has a number of constraints and unknowns that is bounded by a polynomial in the size of  $\mathcal{M}$ , therefore its construction can be performed in polynomial time. For the same reason,  $D_\lambda(\mathcal{M})$  admits a polynomial time separation algorithm: whenever a solution is given, its feasibility is checked by scanning each inequality in  $D_\lambda(\mathcal{M})$ ; otherwise, the first encountered inequality that is not satisfied is returned as a separation hyperplane.

Therefore, the thesis follows by solving  $D_\lambda(\mathcal{M})$  using the ellipsoid method together with the naive separation algorithm described above.  $\square$

## 14.4 Alternative Characterization of the Pseudometric

In the following, we propose an alternative characterization of the bisimilarity distance  $\delta_\lambda$ , based on the notion of *coupling structure*. Our result generalizes the one proposed in [5, 14] for MCs, to the continuous-time settings.

**Definition 14.4.1** (Coupling Structure). Let  $\mathcal{M} = (S, A, \pi, \ell)$  be a CTMC. A coupling structure for  $\mathcal{M}$  is a function

$$\mathcal{C}: (S \setminus A) \times (S \setminus A) \rightarrow \mathcal{D}(S \times S)$$

such that, for all  $s, t \notin A$ ,

$$\mathcal{C}(s, t) \in \Omega(\tau(s), \tau(t)).$$

---

<sup>2</sup>Actually, the variables  $k$  are used only for ease the presentation but they can be removed by substitution.

Intuitively, a coupling structure for  $\mathcal{M}$  can be thought of as a joint transition probability distribution with left and right marginals point wisely equal to  $\tau: S \setminus A \rightarrow \mathcal{D}(S)$ .

The following definition adapts the definition of the operator  $\Delta_\lambda$  (see Definition 14.2.4) with respect to the notion of coupling structure for a CTMC.

**Definition 14.4.2.** Let  $\mathcal{M} = (S, A, \tau, \rho, \ell)$  be a CTMC,  $\mathcal{C}$  a coupling structure for  $\mathcal{M}$ , and  $\lambda \in (0, 1)$  a discount factor. The function  $\Gamma_\lambda^{\mathcal{C}}: [0, 1]^{S \times S} \rightarrow [0, 1]^{S \times S}$  is defined, for arbitrary  $d: S \times S \rightarrow [0, 1]$  and  $s, t \in S$  as follows.

$$\Gamma_\lambda^{\mathcal{C}}(d)(s, t) = \begin{cases} 1 & \text{if } s \notin_A t \\ \mathcal{L}(s, t) & \text{if } s, t \in A \\ \max\{\mathcal{L}(s, t), \lambda \cdot \Theta(d)(s, t)\} & \text{if } s, t \notin A \end{cases}$$

where  $\mathcal{L}, \mathcal{E}: S \times S \rightarrow [0, 1]$  are as in Definition 14.2.4 and  $\Theta: [0, 1]^{S \times S} \rightarrow [0, 1]^{S \times S}$  is given by

$$\Theta(d)(s, t) = \mathcal{E}(s, t) + (1 - \mathcal{E}(s, t)) \cdot \sum_{u, v \in S} d(u, v) \cdot \mathcal{C}(s, t)(u, v).$$

Recall that the Kantorovich distance between two distributions  $\mu$  and  $\nu$  is defined as

$$\mathcal{K}_d(\mu, \nu) = \min_{\omega} \sum_{u, v \in S} d(u, v) \cdot \omega(u, v),$$

where the minimum is taken over all the possible couplings  $\omega \in \Omega(\mu, \nu)$ . Thus, the operator  $\Gamma_\lambda^{\mathcal{C}}$  can intuitively be thought of as a possible instance of  $\Delta_\lambda^{\mathcal{M}}$  with respect to a fixed choice of the couplings given by  $\mathcal{C}$ .

One can easily check that  $\Gamma_\lambda^{\mathcal{C}}$  is monotone, thus, applying Tarski's fixed point theorem, it admits least and greatest fixed points. The least fixed point, in particular, will be denoted by  $\gamma_\lambda^{\mathcal{C}}$  and referred to as the  $\lambda$ -discrepancy of  $\mathcal{C}$ .

**Theorem 14.4.3 (Minimum coupling).** For an arbitrary CTMC  $\mathcal{M}$ ,

$$\delta_\lambda = \min\{\gamma_\lambda^{\mathcal{C}} \mid \mathcal{C} \text{ coupling structure for } \mathcal{M}\}.$$

*Proof.* We first prove that  $\delta_\lambda \sqsubseteq \gamma_\lambda^{\mathcal{C}}$ , for any coupling structure  $\mathcal{C}$  for  $\mathcal{M}$ .

By Tarski's fixed point theorem, it suffices to prove that, for any  $d: S \times S \rightarrow [0, 1]$ ,  $\Delta_\lambda(d) \sqsubseteq \Gamma_\lambda^{\mathcal{C}}(d)$ .

The only nontrivial case is when  $s, t \notin A$ , which follows by definition of  $\mathcal{K}_d$ , by noticing that  $\mathcal{T}(d)(s, t) \leq \Theta(d)(s, t)$  and that the maximum is order preserving.

It remains to prove that the minimum is attained. To this end, define a coupling structure  $\mathcal{C}^*$  as  $\mathcal{C}^*(s, t) = \omega_{s, t}$ , for  $s, t \notin A$ , where  $\omega_{s, t} \in \Omega(\tau(s), \tau(t))$  is such that

$$\mathcal{K}_{\delta_\lambda}(\tau(s), \tau(t)) = \sum_{u, v} \delta_\lambda(u, v) \cdot \omega_{s, t}(u, v).$$

By construction,  $\delta_\lambda = \Gamma^{\mathcal{C}^*}(\delta_\lambda)$ , hence  $\gamma_\lambda^{\mathcal{C}^*} \sqsubseteq \delta_\lambda$ .

Since  $\mathcal{C}^*$  is a coupling structure for  $\mathcal{M}$ , by what we have shown above we also have  $\delta_\lambda \sqsubseteq \gamma_\lambda^{\mathcal{C}^*}$ . Therefore,  $\delta_\lambda = \gamma_\lambda^{\mathcal{C}^*}$ .  $\square$

## 14.5 Greedy Computation of the Bisimilarity Distance

Inspired by the characterization given in Theorem 14.4.3, we propose a procedure to compute the bisimilarity pseudometric that is alternative to those previously described.

The set of coupling structures for  $\mathcal{M}$  can be endowed with the preorder  $\preceq_\lambda$  defined by

$$\mathcal{C} \preceq_\lambda \mathcal{C}' \quad \text{iff} \quad \gamma_\lambda^{\mathcal{C}} \sqsubseteq \gamma_\lambda^{\mathcal{C}'}$$

Theorem 14.4.3 suggests to look at all the coupling structures  $\mathcal{C}$  for  $\mathcal{M}$  in order to find an optimal one, i.e., minimal w.r.t.  $\preceq_\lambda$ .

However, it is clear that the enumeration of all the couplings is unfeasible, therefore it is crucial to provide an efficient search strategy which prevents us to do that. Moreover, we also need an efficient method for computing the  $\lambda$ -discrepancy associated with a coupling structure.

### 14.5.1 Computing the $\lambda$ -Discrepancy

In this subsection we consider the problem of computing the  $\lambda$ -discrepancy associated with a coupling structure.

Using Tarski's fixed point theorem,  $\gamma_\lambda^{\mathcal{C}}$  corresponds to the least pre-fixed point of  $\Gamma^{\mathcal{C}}$ , that is

$$\gamma_\lambda^{\mathcal{C}} = \bigsqcap \{d \in [0, 1]^{S \times S} \mid \Gamma^{\mathcal{C}}(d) \sqsubseteq d\}.$$

This allows us to compute the  $\lambda$ -discrepancy associated with  $\mathcal{C}$  as the optimal solution of the following linear program, denoted by  $Discr_\lambda(\mathcal{C})$ .

$$\begin{aligned} \arg \min_d \quad & \sum_{s,t \in S} d_{s,t} \\ & d_{s,t} \geq 1 && \text{if } s \not\equiv_A t \\ & d_{s,t} \geq \mathcal{L}(s,t) && \text{if } s \equiv_A t \\ & d_{s,t} \geq \lambda(\mathcal{E}(s,t) + (1 - \mathcal{E}(s,t)) \cdot \sum_{u,v \in S} d_{u,v} \cdot \mathcal{C}(s,t)(u,v)) && \text{if } s, t \notin A \end{aligned}$$

$Discr_\lambda(\mathcal{C})$  has a number of inequalities that is bounded by  $2|S|^2$  and  $|S|^2$  unknowns, thus, it can be efficiently solved using the interior point method.

**Remark 14.5.1.** If one is interested in computing the  $\lambda$ -discrepancy for a particular pair of states  $(s, t)$ , the method above can be applied on the least independent set of inequalities containing the variable  $d_{s,t}$ . Moreover, assuming that for some pairs the values associated to  $d$  are known, the set of constraints can be further decreased by substitution. ■

## 14.5.2 Greedy Strategy for Optimal Coupling Structures

Now we propose a *greedy strategy* that moves toward an optimal coupling structure starting from any given one. Then, we provide sufficient and necessary conditions for a coupling structure to ensure that its associated  $\lambda$ -discrepancy coincides with  $\delta_\lambda$ .

Hereafter we fix a CTMC  $\mathcal{M} = (S, A, \tau, \rho, \ell)$  and a coupling structure  $\mathcal{C}$  for it.

The greedy strategy takes a coupling structure and locally updates it at a given pair of states in such a way that it decreases it with respect to  $\preceq_\lambda$ .

For  $s, t \notin A$  and  $\omega \in \Omega(\tau(s), \tau(t))$ , we denote by  $\mathcal{C}[(s, t)/\omega]$  the *update* of  $\mathcal{C}$  at  $(s, t)$  with  $\omega$ , defined as

$$\mathcal{C}[(s, t)/\omega](u, v) = \mathcal{C}(u, v),$$

for all  $(u, v) \neq (s, t)$  and

$$\mathcal{C}[(s, t)/\omega](s, t) = \omega;$$

it is worth noting that, by construction,  $\mathcal{C}[(s, t)/\omega]$  is a coupling structure of  $\mathcal{M}$ .

The next lemma gives a sufficient condition for an update to be effective for the strategy.

**Lemma 14.5.2.** *Let  $s, t \notin A$  and  $\omega \in \Omega(\tau(s), \tau(t))$ . Then, for  $\mathcal{D} = \mathcal{C}[(s, t)/\omega]$  and any  $\lambda \in (0, 1]$ ,*

$$\text{if } \Gamma_\lambda^{\mathcal{D}}(\gamma_\lambda^{\mathcal{C}})(s, t) < \gamma_\lambda^{\mathcal{C}}(s, t) \text{ then } \gamma_\lambda^{\mathcal{D}} \sqsubset \gamma_\lambda^{\mathcal{C}}.$$

*Proof.* It suffices to show that  $\Gamma_\lambda^{\mathcal{D}}(\gamma_\lambda^{\mathcal{C}}) \sqsubset \gamma_\lambda^{\mathcal{C}}$ , i.e., that  $\gamma_\lambda^{\mathcal{C}}$  is a strict post-fixed point of  $\Gamma_\lambda^{\mathcal{D}}$ . Then, the thesis follows by Tarski's fixed point theorem.

Let  $u, v \in S$ . If  $u \neq_A v$ , then

$$\Gamma_\lambda^{\mathcal{D}}(\gamma_\lambda^{\mathcal{C}})(u, v) = 1 = \Gamma_\lambda^{\mathcal{C}}(\gamma_\lambda^{\mathcal{C}})(u, v) = \gamma_\lambda^{\mathcal{C}}(u, v).$$

If  $u, v \in A$ , then

$$\Gamma_\lambda^{\mathcal{D}}(\gamma_\lambda^{\mathcal{C}})(u, v) = \mathcal{L}(u, v) = \Gamma_\lambda^{\mathcal{C}}(\gamma_\lambda^{\mathcal{C}})(u, v) = \gamma_\lambda^{\mathcal{C}}(u, v).$$

If  $u, v \notin A$  and  $(u, v) \neq (s, t)$ , by definition of  $\mathcal{D}$ , we have that  $\bar{\omega}(u, v) = \omega(u, v)$ , hence

$$\Gamma_\lambda^{\mathcal{C}}(\gamma_\lambda^{\mathcal{C}})(u, v) = \Gamma_\lambda^{\mathcal{D}}(\gamma_\lambda^{\mathcal{C}})(u, v).$$

The remaining case, i.e.,  $(u, v) = (s, t)$ , holds by hypothesis.

This proves  $\Gamma_\lambda^{\mathcal{D}}(\gamma_\lambda^{\mathcal{C}}) \sqsubset \gamma_\lambda^{\mathcal{C}}$ . □

Lemma 14.5.2 states that  $\mathcal{C}$  can be improved w.r.t.  $\preceq_\lambda$  by updating it at  $(s, t)$ , if  $s, t \notin A$  and there exists a coupling  $\omega \in \Omega(\tau(s), \tau(t))$  such that the following holds

$$\sum_{u, v \in S} \gamma_\lambda^{\mathcal{C}}(u, v) \cdot \omega(u, v) < \sum_{u, v \in S} \gamma_\lambda^{\mathcal{C}}(u, v) \cdot \mathcal{C}(s, t)(u, v).$$

A coupling that enjoys the above condition is  $\omega \in TP(\gamma_\lambda^c, \tau(s), \tau(t))$  where, for arbitrary  $\mu, \nu \in \mathcal{D}(S)$  and  $c: S \times S \rightarrow [0, 1]$

$$\begin{aligned} TP(c, \mu, \nu) &= \arg \min_{\omega} \sum_{s,t \in S} c(u, v) \cdot \omega_{u,v} \\ &\quad \sum_v \omega_{u,v} = \mu(u) \quad \forall u \in S \\ &\quad \sum_u \omega_{u,v} = \nu(v) \quad \forall v \in S. \end{aligned} \tag{14.5.1}$$

The above mentioned problem is usually referred to as the (homogeneous) *transportation problem* with  $\mu$  and  $\nu$  as the left and the right marginals, respectively, and transportation costs  $c$ . This problem has been extensively studied and comes with (several) efficient polynomial algorithmic solutions [20, 26].

This gives us an efficient solution to update any coupling structure, that, together with Lemma 14.5.2 represents a strategy for moving toward  $\delta_\lambda$  by successive improvements on the coupling structures.

Now we proceed giving a sufficient and necessary condition for termination.

**Lemma 14.5.3.** *Let  $\mathcal{C}$  be a coupling for the CTMC  $\mathcal{M}$ . If  $\gamma_\lambda^c \neq \delta_\lambda$ , then there exist  $s, t \notin A$  and a coupling structure  $\mathcal{D} = \mathcal{C}[(s, t)/\omega]$  for  $\mathcal{M}$  such that*

$$\Gamma_\lambda^{\mathcal{D}}(\gamma_\lambda^c)(s, t) < \gamma_\lambda^c(s, t).$$

*Proof.* We proceed by contraposition.

If for all  $s, t \notin A$  and  $\omega \in \Omega(\tau(s), \tau(t))$ ,

$$\Gamma_\lambda^{\mathcal{D}}(\gamma_\lambda^c)(s, t) \geq \gamma_\lambda^c(s, t),$$

then  $\gamma_\lambda^c = \Delta_\lambda(\gamma_\lambda^c)$ .

Since, by Theorem 14.2.8,  $\Delta_\lambda$  has a unique fixed point,  $\gamma_\lambda^c = \delta_\lambda$ . □

The above result ensures that, unless  $\mathcal{C}$  is optimal w.r.t  $\preceq_\lambda$ , the hypothesis of Lemma 14.5.2 are satisfied, so that, we can further improve  $\mathcal{C}$  as aforesaid.

The next statement proves that this search strategy is correct.

**Theorem 14.5.4.** *Let  $\mathcal{C}$  be a coupling for the CTMC  $\mathcal{M}$ . Then,  $\delta_\lambda = \gamma_\lambda^c$  iff there is no coupling  $\mathcal{D}$  for  $\mathcal{M}$  such that  $\Gamma_\lambda^{\mathcal{D}}(\gamma_\lambda^c) \sqsubset \gamma_\lambda^c$ .*

*Proof.* We prove that  $\delta_\lambda \neq \gamma_\lambda^c$  iff there exists  $\mathcal{D}$  such that  $\Gamma_\lambda^{\mathcal{D}}(\gamma_\lambda^c) \sqsubset \gamma_\lambda^c$ .

( $\Rightarrow$ ) Assume  $\delta_\lambda \neq \gamma_\lambda^c$ .

By Lemma 14.5.3, there exists a pair of states  $s, t \in S$  and a coupling  $\omega \in \Omega(\tau(s), \tau(t))$  such that

$$\lambda \cdot \sum_{u,v \in S} \gamma_\lambda^c(u, v) \cdot \omega(u, v) < \gamma_\lambda^c(s, t).$$

As in the proof of Lemma 14.5.2, we have that  $\mathcal{D} = \mathcal{C}[(s, t)/\omega]$  satisfies  $\Gamma^{\mathcal{D}}(\gamma_\lambda^{\mathcal{C}}) \sqsubset \gamma_\lambda^{\mathcal{C}}$ .

( $\Leftarrow$ ) Let  $\mathcal{D}$  be such that  $\Gamma_\lambda^{\mathcal{D}}(\gamma_\lambda^{\mathcal{C}}) \sqsubset \gamma_\lambda^{\mathcal{C}}$ .

By Tarski's fixed point theorem  $\gamma_\lambda^{\mathcal{D}} \sqsubset \gamma_\lambda^{\mathcal{C}}$ . Applying Theorem 14.4.3,

$$\delta_\lambda \sqsubseteq \gamma_\lambda^{\mathcal{D}} \sqsubset \gamma_\lambda^{\mathcal{C}}.$$

Therefore,  $\delta_\lambda \neq \gamma_\lambda^{\mathcal{C}}$ . □

**Remark 14.5.5.** Note that, in general, there could be an infinite number of couplings structures for a given CTMC.

However, for each fixed  $d \in [0, 1]^{S \times S}$ , the linear function mapping  $\omega$  to  $\sum_{u,v \in S} d(u, v) \cdot \omega(u, v)$  achieves its minimum at some vertex in the transportation polytope  $\Omega(\tau(s), \tau(t))$ .

Since the number of such vertices are finite, using the optimal TP schedule for the update, ensures that the search strategy is always terminating. ■

## 14.6 The On-the-Fly Algorithm

In this section we describe an on-the-fly technique for computing the bisimilarity distance  $\delta_\lambda$  fully exploiting the greedy strategy of Section 14.5.2.

Let  $Q \subseteq S \times S$  and consider the problem of computing  $\delta_\lambda(s, t)$  for all  $(s, t) \in Q$ . Recall that the strategy proposed in Section 14.5.2 consists in a traversal

$$\mathcal{C}_0 \triangleright_\lambda \mathcal{C}_1 \triangleright_\lambda \cdots \triangleright_\lambda \mathcal{C}_n$$

of the set of coupling structures for  $\mathcal{M}$  that starts from an arbitrary coupling structure  $\mathcal{C}_0$  and leads to an optimal one  $\mathcal{C}_n$ . We observe that, for any  $i < n$

1. the improvement of each coupling structure  $\mathcal{C}_i$  is obtained by a *local update* at some pair of states  $u, v \notin A$ , namely  $\mathcal{C}_{i+1} = \mathcal{C}_i[(u, v)/\omega]$  for some  $\omega \in TP(\gamma_\lambda^{\mathcal{C}_i}, \tau(u), \tau(v))$ ;
2. the pair  $(u, v)$  is chosen according to an optimality check that is performed *locally* among the couplings in  $\Omega(\tau(u), \tau(v))$ , i.e.,  $\mathcal{C}_i(u, v) \notin TP(\gamma_\lambda^{\mathcal{C}_i}, \tau(u), \tau(v))$ ;
3. whenever a coupling structure  $\mathcal{C}_i$  is considered, its associated  $\lambda$ -discrepancy  $\gamma_\lambda^{\mathcal{C}_i}$  can be computed by solving the linear program  $Discr_\lambda(\mathcal{C}_i)$  of Section 14.5.1.

Among the observations above, only the last one requires to look at the coupling structure  $\mathcal{C}_i$ . However, as noticed in Remark 14.5.1, the value  $\gamma_\lambda^{\mathcal{C}_i}(s, t)$  can be computed without considering the entire set of constraints of  $Discr_\lambda(\mathcal{C}_i)$ , but only the least independent set of inequalities that contains the variable  $d_{s,t}$ . Moreover, provided that for some pairs of states  $E \subseteq S \times S$  the value of the distance is known, the linear program  $Discr_\lambda(\mathcal{C}_i)$  can be further reduced by substituting the occurrences of the unknown  $d_{u,v}$  by the constant  $\delta_\lambda(u, v)$ , for each  $(u, v) \in E$ . This suggests that we do not need to store the entire coupling structures, but they can be constructed *on-the-fly* during the calculation.

Specifically, the couplings that are demanded to compute  $\gamma_\lambda^{C_i}(s, t)$  are only those  $C_i(u, v)$  such that  $(s, t) \rightsquigarrow_{C_i, E}^* (u, v)$ , where  $\rightsquigarrow_{C_i, E}^*$  is the reflexive and transitive closure of  $\rightsquigarrow_{C_i, E} \subseteq S^2 \times S^2$ , defined by

$$(s', t') \rightsquigarrow_{C_i, E} (u', v') \quad \text{iff} \quad C_i(s', t')(u', v') > 0 \text{ and } (u', v') \notin E.$$

The computation of the bisimilarity pseudometric is implemented by Algorithm 6. It takes as input a finite CTMC  $\mathcal{M} = (S, A, \tau, \rho, \ell)$ , a discount factor  $\lambda \in (0, 1)$ , and a query set  $Q \subseteq S \times S$ . We assume the following global variables to store:

- $\mathcal{C}$ : the current (partial) coupling structure;
- $d$ : the  $\lambda$ -discrepancy associated with  $\mathcal{C}$ ;
- *ToCompute*: the pairs of states for which the distance has to be computed;
- *Exact*: the set pairs of states  $(s, t)$  such that  $d(s, t) = \delta_\lambda(s, t)$ , hence those pairs which do not need to be further improved<sup>3</sup>;
- *Visited*: the set of pairs of states that as been visited so far.

Moreover,  $\mathcal{R}_{s,t}(\text{Exact}, \mathcal{C})$  will denote the set  $\{(u, v) \mid (s, t) \rightsquigarrow_{\mathcal{C}, \text{Exact}}^* (u, v)\}$ .

At the beginning (line 1–2) both the coupling structure  $\mathcal{C}$  and the discrepancy  $d$  are empty, there are no visited states, no exact computed distances, and the pairs to be computed are those in the input query.

While there are still pairs left to be computed (line 3), we pick one (line 4), say  $(s, t)$ . According to the definition of  $\delta_\lambda$ , if  $s \notin_A t$  then  $\delta_\lambda(s, t) = 1$ ; if  $s = t$  then  $\delta_\lambda(s, t) = 0$  and if  $s, t \in A$  then  $\delta_\lambda(s, t) = \mathcal{L}(s, t)$ , so that,  $d(s, t)$  is set accordingly, and  $(s, t)$  is added to *Exact* (lines 5–10).

Otherwise, if  $(s, t)$  was not previously visited, a coupling  $\omega \in \Omega(\tau(s), \tau(t))$  is guessed, and the routine *SetPair* updates the coupling structure  $\mathcal{C}$  at  $(s, t)$  with  $\omega$  (line 14), then the routine *Discrepancy* updates  $d$  with the  $\lambda$ -discrepancy associated with  $\mathcal{C}$  (line 16).

According to the greedy strategy,  $\mathcal{C}$  is successively improved and  $d$  is consequently updated, until no further improvements are possible (lines 17–21).

Each improvement is obtained by replacing a sub-optimal coupling  $C(u, v)$ , for some  $(u, v) \in \mathcal{R}_{s,t}(\text{Exact}, \mathcal{C})$ , by one taken from  $TP(d, \tau(u), \tau(v))$  (line 17). Note that, each improvement actually affects the current value of  $d(s, t)$ , since the update is performed on a pair in  $\mathcal{R}_{s,t}(\text{Exact}, \mathcal{C})$ .

It is worth to note that  $\mathcal{C}$  and *Exact* are constantly updated, hence  $\mathcal{R}_{s,t}(\text{Exact}, \mathcal{C})$  may differ from one iteration to another.

When line 22 is reached, for each  $(u, v) \in \mathcal{R}_{s,t}(\text{Exact}, \mathcal{C})$ , we are guaranteed that  $d(u, v) = \delta_\lambda(s, t)$ , therefore  $\mathcal{R}_{s,t}(\text{Exact}, \mathcal{C})$  is added to *Exact* and, for these pairs,  $d$  will

<sup>3</sup>Actually, the set *Exact* contains those pairs such that  $|d(s, t) - \delta_\lambda(s, t)| \leq \varepsilon$  where  $\varepsilon$  corresponds to the precision of the machine. In our implementation  $\varepsilon = 10^{-9}$ .

**Algorithm 6** On-the-Fly Bisimilarity Pseudometric**Input:** CTMC  $\mathcal{M} = (S, A, \tau, \rho, \ell)$ ; discount factor  $\lambda \in (0, 1)$ ; query  $Q \subseteq S \times S$ .

- 
1.  $\mathcal{C} \leftarrow \text{empty}$ ;  $\text{Dom}_{\mathcal{C}} \leftarrow \emptyset$ ;  $d \leftarrow \text{empty}$ ; —initialize data structures—
  2.  $\text{Visited} \leftarrow \emptyset$ ;  $\text{Exact} \leftarrow \emptyset$ ;  $\text{ToCompute} \leftarrow Q$
  3. **while**  $\text{ToCompute} \neq \emptyset$  **do**
  4.   pick  $(s, t) \in \text{ToCompute}$
  5.   **if**  $s \not\equiv_A t$  **then**
  6.      $d(s, t) \leftarrow 1$ ;  $\text{Exact} \leftarrow \text{Exact} \cup \{(s, t)\}$ ;  $\text{Visited} \leftarrow \text{Visited} \cup \{(s, t)\}$
  7.   **else if**  $s = t$  **then**
  8.      $d(s, t) \leftarrow 0$ ;  $\text{Exact} \leftarrow \text{Exact} \cup \{(s, t)\}$ ;  $\text{Visited} \leftarrow \text{Visited} \cup \{(s, t)\}$
  9.   **else if**  $s, t \in A$  **then**
  10.      $d(s, t) \leftarrow \mathcal{L}(s, t)$ ;  $\text{Exact} \leftarrow \text{Exact} \cup \{(s, t)\}$ ;  $\text{Visited} \leftarrow \text{Visited} \cup \{(s, t)\}$
  11.   **else** —if  $(s, t)$  is nontrivial—
  12.     **if**  $(s, t) \notin \text{Visited}$  **then** —if  $(s, t)$  has not been encountered so far—
  13.       pick  $\omega \in \Omega(\tau(s), \tau(t))$  —guess a coupling—
  14.        $\text{SetPair}(\mathcal{M}, (s, t), \omega)$  —update the current coupling structure—
  15.     **end if**
  16.      $\text{Discrepancy}(\lambda, (s, t))$  —update  $d$  as the  $\lambda$ -discrepancy for  $\mathcal{C}$ —
  17.     **while**  $\exists (u, v) \in \mathcal{R}_{s,t}(\text{Exact}, \mathcal{C})$  such that  $\mathcal{C}(u, v) \notin \text{TP}(d, \tau(u), \tau(v))$  **do**
  18.        $\omega \in \text{TP}(d, \tau(u), \tau(v))$  —pick an optimal coupling for  $s, t$  w.r.t.  $d$ —
  19.        $\text{SetPair}(\mathcal{M}, (u, v), \omega)$  —improve the current coupling structure—
  20.        $\text{Discrepancy}(\lambda, (s, t))$  —update  $d$  as the  $\lambda$ -discrepancy for  $\mathcal{C}$ —
  21.     **end while**
  22.      $\text{Exact} \leftarrow \text{Exact} \cup \mathcal{R}_{\mathcal{C}}(s, t)$  —add new exact distances—
  23.     remove from  $\mathcal{C}$  all the couplings associated with a pair in  $\text{Exact}$
  24.   **end if**
  25.    $\text{ToCompute} \leftarrow \text{ToCompute} \setminus \text{Exact}$  —remove exactly computed pairs—
  26. **end while**
  27. **return**  $d|_Q$  —return the distance for all pairs in  $Q$ —
- 

no longer be updated. At this point (line 23), the couplings associated with the pairs in  $\text{Exact}$  can be removed from  $\mathcal{C}$ .

At line 25, the exact pairs computed so far are removed from  $\text{ToCompute}$ .

Finally, if no more pairs need be considered, the exact distance on  $Q$  is returned (line 27).

Algorithm 6 calls the subroutines  $\text{SetPair}$  and  $\text{Discrepancy}$ . The former is used to construct and update the coupling structure  $\mathcal{C}$ , the latter to update the current over-approximation  $d$  during the computation. Next, we explain how they work.

$\text{SetPair}$  (Algorithm 7) takes as input a CTMC  $\mathcal{M} = (S, A, \tau, \rho, \ell)$ , a pair of states  $s, t \in S$ , and a coupling  $\omega \in \Omega(\tau(s), \tau(t))$ .

**Algorithm 7** *SetPair*( $\mathcal{M}, (s, t), \omega$ )**Input:** CTMC  $\mathcal{M} = (S, A, \tau, \rho, \ell); s, t \in S; \omega \in \Omega(\tau(s), \tau(t))$ 

1.  $\mathcal{C}(s, t) \leftarrow \omega$  —update the coupling at  $(s, t)$  with  $\omega$ —
2.  $Visited \leftarrow Visited \cup \{(s, t)\}$  —set  $(s, t)$  as visited—
3. **for all**  $(u, v) \notin Visited$  such that  $(s, t) \rightsquigarrow_{\mathcal{C}, Exact} (u, v)$  **do** —for all demanded pairs—
4.    $Visited \leftarrow Visited \cup \{(u, v)\}$
5.   **if**  $u = v$  **then**  $d(u, v) \leftarrow 0; Exact \leftarrow Exact \cup \{(u, v)\};$
6.   **if**  $u \not\equiv_A v$  **then**  $d(u, v) \leftarrow 1; Exact \leftarrow Exact \cup \{(u, v)\};$
7.   **if**  $u, v \in A$  **then**  $d(u, v) \leftarrow \mathcal{L}(u, v); Exact \leftarrow Exact \cup \{(u, v)\};$
8.   // propagate the construction
9.   **if**  $(u, v) \notin Exact$  **then**
10.     pick  $\omega' \in \Omega(\tau(u), \tau(v))$  —guess a matching—
11.      $SetPair(\mathcal{M}, (u, v), \omega')$
12.   **end if**
13. **end for**

**Algorithm 8** *Discrepancy*( $\lambda, (s, t)$ )**Input:** discount factor  $\lambda \in (0, 1); s, t \in S \setminus A$ 

1. Let LP be the linear program obtained from  $Discr_\lambda(\mathcal{C})$  by keeping only the inequalities associated with pairs in  $\mathcal{R}_{s,t}(Exact, \mathcal{C})$  and replacing the unknown  $d_{u,v}$  by the constant  $d(u, v)$ , for all  $(u, v) \in Exact$ .
2.  $d^* \leftarrow$  optimal solution of LP
3. **for all**  $(u, v) \in \mathcal{R}_{s,t}(Exact, \mathcal{C})$  **do** —update distances—
4.    $d(u, v) \leftarrow d^*_{u,v}$
5.   **if**  $d(u, v) = 0$  or  $d(u, v) = \mathcal{L}(u, v)$  **then**
6.      $Exact \leftarrow Exact \cup \{(u, v)\}$
7.   **end if**
8. **end for**

In lines 1–2, the coupling structure  $\mathcal{C}$  is set to  $\omega$  at  $(s, t)$ , then  $(s, t)$  is added to *Visited*.

The on-the-fly construction of the coupling structure is recursively propagated to the demanded successor pairs of  $(s, t)$  according to the information accumulated so far.

During this construction, if some states with trivial distances are encountered,  $d$  and *Exact* are updated accordingly (lines 5–7).

*Discrepancy* (Algorithm 8) takes as input a discount factor  $\lambda \in (0, 1)$  and a pair of states  $s, t \notin A$ . It constructs the least independent linear program obtained from  $Discr_\lambda(\mathcal{C})$ , that can compute  $\gamma_\lambda^{\mathcal{C}}(s, t)$  using the information accumulated so far (line 1).

At the lines 3–7 the current  $\lambda$ -discrepancy is updated accordingly; and those pairs  $(u, v) \in \mathcal{R}_{s,t}(Exact, \mathcal{C})$  for which the current  $\lambda$ -discrepancy coincide with the distance are added to *Exact*.

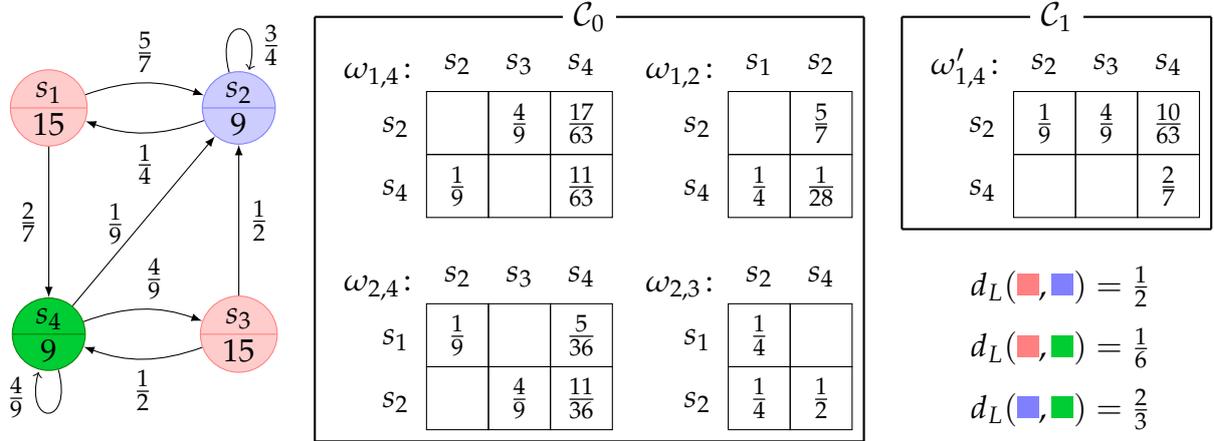


Figure 14.3: Execution trace for the computation of  $\delta_{\frac{1}{2}}(1,4)$  (details in Example 14.6.1).

Next, we present a simple example of Algorithm 6, emphasizing the main features of our method:

1. the on-the-fly construction of the (partial) coupling, and
2. the restriction only to those variables which are demanded for the solution of the system of linear equations.

**Example 14.6.1 (On-the-fly computation).** Consider the CTMC in Figure 14.3, and assume that we want to compute the  $\lambda$ -discounted bisimilarity distance between states  $s_1$  and  $s_4$ , for  $\lambda = \frac{1}{2}$ .

Algorithm 6 starts by guessing an initial coupling structure  $\mathcal{C}_0$ . This is done by considering only the pairs of states which are really needed in the computation.

Starting from the pair  $(s_1, s_4)$  a coupling in  $\omega_{1,4} \in \Omega(\tau(s_1), \tau(s_4))$  is guessed as in Figure 14.3 and assigned to  $\mathcal{C}_0(s_1, s_4)$ .

This demands for the exploration of the pairs  $(s_2, s_3)$ ,  $(s_2, s_4)$ ,  $(s_1, s_2)$  and the guess of three new couplings  $\omega_{2,3} \in \Omega(\tau(s_2), \tau(s_3))$ ,  $\omega_{2,4} \in \Omega(\tau(s_2), \tau(s_4))$ , and  $\omega_{1,2} \in \Omega(\tau(s_1), \tau(s_2))$ , to be associated in  $\mathcal{C}_0$  with their corresponding pairs.

Since no other pairs are demanded, the construction of  $\mathcal{C}_0$  terminates as shown in Figure 14.3.

The  $\lambda$ -discrepancy associated with  $\mathcal{C}_0$  for the pair  $(s_1, s_4)$  is obtained as the solution of the

following reduced linear program

$$\begin{aligned}
 & \arg \min_d (d_{1,4} + d_{2,3} + d_{2,4} + d_{1,2}) \\
 & d_{1,4} \geq \frac{1}{6} \\
 & d_{1,4} \geq \frac{\alpha}{2} + \frac{(1-\alpha)}{2} \cdot \left( \frac{4}{9} \cdot d_{2,3} + \frac{8}{21} \cdot d_{2,4} + \frac{11}{63} \cdot \overbrace{d_{4,4}}^{=0} \right) \\
 & d_{2,3} \geq \frac{1}{2} \\
 & d_{2,3} \geq \frac{\alpha}{2} + \frac{(1-\alpha)}{2} \cdot \left( \frac{1}{4} \cdot d_{1,2} + \frac{1}{4} \cdot \overbrace{d_{2,2}}^{=0} + \frac{1}{2} \cdot d_{2,4} \right) \\
 & d_{2,4} \geq \frac{2}{3} \\
 & d_{2,4} \geq \frac{1}{2} \cdot \left( \frac{1}{9} \cdot d_{1,2} + \frac{5}{36} \cdot d_{1,4} + \frac{4}{9} \cdot d_{2,3} + \frac{11}{36} \cdot d_{2,4} \right) \\
 & d_{1,2} \geq \frac{1}{2} \\
 & d_{1,2} \geq \frac{\alpha}{2} + \frac{(1-\alpha)}{2} \cdot \left( \frac{5}{7} \cdot \overbrace{d_{2,2}}^{=0} + \frac{1}{4} \cdot d_{1,4} + \frac{1}{28} \cdot d_{2,4} \right)
 \end{aligned}$$

where  $\alpha = \|\text{Exp}(15) - \text{Exp}(9)\|_{\text{TV}} = \frac{\sqrt[6]{3/5}}{25}$  (by Equation (14.3.1)).

Note that, the bisimilarity distance for the pairs  $(s_2, s_2)$  and  $(s_4, s_4)$  is always 0, thus  $d_{2,2}$  and  $d_{4,4}$  are substituted accordingly.

The solution of the above linear program is

$$d^{\mathcal{C}_0}(s_1, s_4) = \frac{\alpha}{2} + \frac{5(1-\alpha)}{21},$$

$$d^{\mathcal{C}_0}(s_2, s_3) = \frac{1}{2},$$

$$d^{\mathcal{C}_0}(s_2, s_4) = \frac{2}{3},$$

and

$$d^{\mathcal{C}_0}(s_1, s_2) = \frac{1}{2}.$$

Since, the  $\lambda$ -discrepancy for  $(s_2, s_3)$ ,  $(s_2, s_4)$ , and  $(s_1, s_2)$  equals the distance  $\mathcal{L}$  between their labels, it coincides with the bisimilarity distance, hence it cannot be further decreased. Consequently, the pairs of states are added to the set *Exact* and their associated couplings are removed from  $\mathcal{C}_0$ . Note that, these pairs will no longer be considered in the construction of a coupling structure.

In order to decrease the  $\lambda$ -discrepancy of  $(s_1, s_4)$ , Algorithm 6 constructs a new coupling structure  $\mathcal{C}_1$ .

According to our greedy strategy,  $\mathcal{C}_1$  is obtained from  $\mathcal{C}_0$  updating  $\mathcal{C}_0(s_1, s_4)$  (i.e., the only coupling left) by the coupling  $\omega'_{1,4} \in \Omega(\tau(s_1), \tau(s_4))$  (shown in Figure 14.3) that is obtained as the solution of a transportation problem with marginals  $\tau(s_1)$  and  $\tau(s_4)$ , where the current  $\lambda$ -discrepancy is taken as cost function.

The resulting coupling does not demand for the exploration of new pairs in the CTMC, hence the construction of  $\mathcal{C}_1$  terminates. The reduced linear program associated with  $\mathcal{C}_1$  is given by

$$\begin{aligned} & \arg \min_d d_{1,4} \\ & d_{1,4} \geq \frac{1}{6} \\ & d_{1,4} \geq \frac{\alpha}{2} + \frac{(1-\alpha)}{2} \cdot \left( \frac{1}{9} \cdot \overbrace{d_{2,2}}^{=0} + \frac{4}{9} \cdot \overbrace{d_{2,3}}^{=\frac{1}{2}} + \frac{10}{63} \cdot \overbrace{d_{2,4}}^{=\frac{2}{3}} + \frac{2}{7} \cdot \overbrace{d_{4,4}}^{=0} \right) \end{aligned}$$

whose solution is  $d^{\mathcal{C}_1}(s_1, s_4) = \frac{\alpha}{2} + \frac{31(1-\alpha)}{189}$ .

Solving again a new transportation problem with the improved current  $\lambda$ -discrepancy as cost function, we discover that the coupling structure  $\mathcal{C}_1$  cannot be further improved, hence we stop the computation, returning

$$\delta_\lambda(s_1, s_4) = d^{\mathcal{C}_1}(s_1, s_4) = \frac{\alpha}{2} + \frac{31(1-\alpha)}{189}.$$

■

**Remark 14.6.2.** Algorithm 6 can also be used for computing over-approximated distances. Indeed, assuming over-estimates for some particular distances are already known, they can be taken as inputs and used in our algorithm simply storing them in the variable  $d$  and treated as “exact” values. In this way, our method will return the least over-approximation of the distance agreeing with the given over-estimates. This modification of the algorithm can be used to further decrease the exploration of the CTMC. Moreover, it can be employed in combination with approximated algorithms, having the advantage of an on-the-fly state space exploration. ■

## 14.7 Experimental Results

In this section, we evaluate the performances of the on-the-fly algorithm on a collection of randomly generated<sup>4</sup> CTMCs.

Firstly, we compare the execution times of the on-the-fly algorithm with those of the iterative method proposed in Section 14.3.1. Since the iterative method only allows for

<sup>4</sup> The tests have been performed on a prototype implementation coded in Wolfram Mathematica<sup>®</sup> 9 (available at <http://people.cs.aau.dk/~mardare/tools.html>) running on an Intel Core-i7 3.4 GHz processor with 12GB of RAM.

# States	On-the-Fly (exact)		Iterating (approximated)		Approx. Error
	Time (s)	# TPs	# Iterations	# TPs	
10	0.352	10.500	2.660	266.667	0.0339
12	0.772	19.700	2.850	410.403	0.0388
14	2.496	35.800	3.880	760.480	0.0318
16	4.549	50.607	5.142	1316.570	0.0230
18	13.709	78.611	6.638	2151.021	0.0206
20	22.044	109.146	7.243	2897.560	0.0149
22	50.258	140.727	7.409	3586.010	0.0145
24	67.049	175.481	7.826	4508.310	0.0141
26	112.924	219.255	9.509	6428.150	0.0025
28	247.583	295.533	11.133	8728.530	0.0004
30	284.252	307.698	10.679	9611.320	0.0006
40	296.633	330.824	11.294	18070.600	0.0004
50	807.522	368.500	16.900	42250.000	0.00001

Table 14.1: Comparison between the on-the-fly algorithm and the iterative method.

the computation of the distance for all state pairs at once, the comparison is (in fairness) made with respect to runs of our on-the-fly algorithm with input query being the set of all state pairs. For each input instance, the comparison involves the following steps:

1. we run the on-the-fly algorithm, storing both execution time and the number of solved transportation problems,
2. then, on the same instance, we execute the iterative method until the running time exceeds that of step 1. We report the number of iterations and the number of solved transportation problems.
3. Finally, we calculate the approximation error between the exact solution  $\delta_\lambda$  computed by our method at step 1 and the approximate result  $d$  obtained in step 2 by the iterative method, as  $\|\delta_\lambda - d\|$ .

This has been made on a collection of CTMCs varying from 10 to 50 states. For each  $n = 10, \dots, 30$ , we have considered 40 randomly generated CTMCs per out-degree, varying from 3 to  $n$ ; whereas for  $n = 40$  and 50, the out-degree varies from 3 to 10. Table 14.1 reports the average results of the comparison obtained for a discount factor  $\lambda = \frac{1}{2}$ .

As it can be seen, our use of a greedy strategy in the construction of the couplings leads to a significant improvement in the performances. We are able to compute the exact solution before the iterative method can under-approximate it with an absolute error of  $\approx 0.03$ , which is a non-negligible error for a value within the interval  $[0, 1]$ .

So far, we only examined the case when the on-the-fly algorithm is run on all state pairs at once. Now, we show how the performance of our method is improved even further when the distance is computed only for single pairs of states.

# States	out-deg = 3		$3 \leq \text{out-deg} \leq \# \text{ States}/2$		# States	out-deg = 3	
	Time (s)	# TPs	Time (s)	# TPs		Time (s)	# TPs
30	0.304	0.383	18.113	21.379	60	34.858	12.053
40	2.045	0.954	34.582	22.877	70	48.016	14.166
50	7.832	16.304	50.258	139.427	80	73.419	29.383
					90	75.591	13.116
					100	158.027	20.301

Table 14.2: Average performances of the on-the-fly algorithm on single-pair queries. Execution times and number of performed TPs are reported for CTMCs with different out-degree. For instances with more than 50 states the out-degree is fixed to 3;

Table 14.2 shows the average execution times and number of solved transportation problems for (nontrivial) single-pair queries for randomly generated of CTMCs with number of states varying from 30 to 100. In the first two columns we consider CTMCs with out-degree equal to 3, while the last two columns show the average values for out-degrees varying from 3 to half of the number of states of the CTMCs.

The results show that, when the out-degree of the CTMCs is low, our algorithm performs orders of magnitude better than in the general case.

Notably, our on-the-fly method scales well when the out-degree is small and successive computation of the current  $\lambda$ -discrepancy are performed on a relatively small set of pairs.

As for the linear program characterization of the bisimilarity distance illustrated in Section 14.3.2, tests performed on small CTMCs show that solving  $D_\lambda(\mathcal{M})$  is inefficient in practice, both using the simplex and the interior-point methods<sup>5</sup>. Even for CTMCs with less than 20 states, the computation times are in the order of hours. For this reason, the efficiency of our on-the-fly technique is by no mean comparable to the linear program solution.

## 14.8 Conclusive Remarks

In this Chapter, we proposed a bisimilarity pseudometric for measuring the behavioral similarity between CTMCs, that extends the one on MCs introduced by Desharnais et al. in [22].

Moreover, we gave a novel linear program characterization of the distance that, differently from similar previous proposals, have a number of constraints which is polynomial in the size of the CTMC. This proved that the bisimilarity pseudometric can be computed in polynomial time.

<sup>5</sup>The implementation is done in Wolfram Mathematica<sup>®</sup> 9 and uses the Linear Program solvers available in the standard library.

Finally, we defined an on-the-fly algorithm for computing the bisimilarity distance. We demonstrated that, using on-the-fly techniques the computation time is improved with orders of magnitude with respect to the corresponding iterative and linear program approaches. Our technique allows for the computation on a set of target distances that might be done by only investigating a significantly reduced set of states and for further improvement of speed.

Our algorithm can be practically used to address a large spectrum of problems. For instance, it can be seen as a method to decide whether two states of a given CTMC are probabilistic bisimilar, to identify bisimilarity classes, or to solve lumpability problems. It is sufficiently robust to be used with approximation techniques as, for instance, to provide a least over-approximation of the behavioural distance given over-estimates of some particular distances. It can be integrated with other approximate algorithms, having the advantage of the efficient on-the-fly state space exploration.

Having a practically efficient tool to compute bisimilarity distances opens the perspective of new applications already announced in previous Chapters. One of these is the state space reduction problem for CTMCs. Our technique can be used in this context as an indicator for the sets of neighbor states that can be collapsed due to their similarity. It also provides a tool to estimate the difference between the initial CTMC and the reduced one, hence a tool for the approximation theory of CTMCs.



# Chapter 15

## Total Variation Distances of Semi-Markov Chains

### 15.1 Introduction

The growing interest in quantitative aspects in the real world applications motivated the introduction of quantitative models and formal methods for studying their behaviors. Classically, the behavior of two models is compared by means of an equivalence (e.g., bisimilarity, trace equivalence, etc.). However, when the models are parametric on some numerical values subject to error estimates or obtained from statistical samplings, any notion of equivalence for comparing two systems is too strong a concept, where from the necessity to relax it towards a concept of behavioural distance.

As we have already seen in the previous Chapters, the first proposals of behavioral distances in the literature are based on the Kantorovich metric and, like bisimilarity, are branching-time. In this Chapter instead, we consider a linear-time metric. Our attention on linear-time properties is motivated by the fact that in many applications the system to be modeled cannot be internally accessed, but only tested via observations performed over a set of random executions. For instance, this is mostly common in application domains such as systems biology, modeling/testing, and machine learning.

An other aspect we want to consider are real-time properties. These are important for performance evaluation of cyber-physical systems and dependability analysis.

In this respect, we consider as models the so-called semi-Markov chains (SMCs), which are continuous-time probabilistic transition systems where the residence time on states is governed by generic distributions on the positive real line. SMCs subsume many probabilistic models also studied in the previous Chapters of this monograph, such as the Markov chains (MCs) and the continuous-time Markov Chains (CTMCs).

Regarding the behavioural distance, we study the total variation between probability measures induced by an SMC over infinite timed traces, which corresponds to

the largest possible difference between the probabilities that the measures assign to the same event.

One of most appealing features of the total variation distance between MCs is its relation to probabilistic model checking. Specifically, the total variation is known to be a bound for the maximal difference with respect to the probability of satisfying LTL formulas. However, also the indiscrete distance (that returns distance 0 to identical states, 1 otherwise) is a bound for it, but it is clear that it is of no help. What is important, is to establish *how tight is a given bound*.

In the first part of this Chapter we address this problem in the case of SMCs and probabilistic model checking of linear real-time specifications expressed either as metric temporal logic (MTL) formulas [2, 3], or languages accepted by timed automata (TAs) [1].

The relation between the total variation and model checking is even tighter than expected: we show that the bound is actually an *equality*. This holds both for MTL properties and for TAs languages. These are obtained from a more general result (Theorem 15.4.2), which also entails other nontrivial characterizations of the distance.

Since, SMCs and MTL are extension of MCs and LTL, respectively, the same result holds for the discrete-time case.

This further motivates the study of efficient methods for computing the total variation. Unfortunately, in [17, 34] this problem is proven to be NP-hard in the case of MCs, and to the best of our knowledge, its decidability is still an open problem. Nevertheless, we prove that the problem of approximating the total variation distance with arbitrary precision is computable. This is done providing two sequences that converge from below and above to the total variation distance. This result generalizes that of [17] to the real-time setting. Our approach, however, is different, as it is basing on a duality that characterizes the total variation between two measures as the minimal discrepancy associated with their couplings.

The technical contributions of this Chapter can be summarized as follows.

1. We solved the open problem of measuring how tight is the upper-bound given by the total variation distance with respect to the variational distance ranging over MTL formulas and TA specifications, respectively. This has been made possible due to a more general result —the actual main contribution of the section— that yields many other interesting characterizations of the total variation on SMCs.

2. We provide sufficient conditions to construct sequences that converge, from below and above, to the total variation distance. Differently from [17], the converging conditions are not specific to the probabilistic transition system at hand, but the results hold for probability measures on an arbitrary measurable space. These results are instantiated to characterize two specific sequences that converge to the total variation distance between SMCs.

3. Last but not least, we prove the computability of the converging sequences of the previous point. This yields a computable procedure to approximate the total variation distance with arbitrary precision.

## 15.2 Preliminaries

In this section, we recall some basic notions used in the Chapter and fix the notation. For deeper insides on some of these topics, the reader is also referred to the Chapter Preliminaries of this monograph.

As before in this monograph, the set of functions from  $X$  to  $Y$  is denoted by  $Y^X$  and for  $f \in Y^X$ , let

$$\equiv_f = \{(x, x') \mid f(x) = f(x')\}.$$

Given an equivalence relation  $R \subseteq X \times X$ , the set of  $R$ -equivalence classes is denoted by  $X/R$  and  $[x]_R$  denotes the equivalence class of  $x \in X$ .

Hereafter  $(\mathbb{R}_{\geq 0}, \mathbb{B})$  denotes the measurable space of positive real numbers (including zero) with Borel algebra.

Given a measurable space  $(X, \Sigma)$ , in what follows  $\Delta(X, \Sigma)$  denotes the set of probability measures on  $(X, \Sigma)$  and let  $\mathcal{D}(X) = \Delta(X, 2^X)$  be the set of discrete probability distributions on  $(X, \Sigma)$ .

Given a measurable function  $f: (X, \Sigma) \rightarrow (Y, \Theta)$ , any measure  $\mu$  on  $(X, \Sigma)$  defines a measure  $\mu[f]$  on  $(Y, \Theta)$  by

$$\mu[f](E) = \mu(f^{-1}(E)),$$

for all  $E \in \Theta$ ; it is called the *push forward of  $\mu$  under  $f$* .

Given  $\mu$  and  $\nu$  measures on  $(X, \Sigma)$  and  $(Y, \Theta)$ , respectively, the *product measure*  $\mu \times \nu$  on  $(X, \Sigma) \otimes (Y, \Theta)$  is *uniquely* defined by

$$(\mu \times \nu)(E \times F) = \mu(E) \cdot \nu(F),$$

for all  $(E, F) \in \Sigma \times \Theta$ .

A measure  $\omega$  on  $(X, \Sigma) \otimes (Y, \Theta)$  is a *coupling* for  $(\mu, \nu)$  if for all  $E \in \Sigma$  and  $F \in \Theta$ ,  $\omega(E \times Y) = \mu(E)$  and  $\omega(X \times F) = \nu(F)$ ;  $\mu$  is called the *left* and  $\nu$  the *right marginals* of  $\omega$ . We denote by  $\Omega(\mu, \nu)$  the set of couplings for  $(\mu, \nu)$ .

Given a measurable space  $(X, \Sigma)$ , the set of measures  $\Delta(X, \Sigma)$  is metrized by the *total variation distance*, defined for arbitrary  $\mu, \nu \in \Delta(X, \Sigma)$  by

$$\|\mu - \nu\| = \sup_{E \in \Sigma} |\mu(E) - \nu(E)|.$$

**The space of timed paths.**

A *timed path* over a set  $X$  is an infinite sequence

$$\pi = x_0, t_0, x_1, t_1 \dots,$$

where  $x_i \in X$  and  $t_i \in \mathbb{R}_{\geq 0}$ ;  $t_i$  are called *time delays*.

For any  $i \in \mathcal{N}$ , let

$$\begin{aligned} \pi[i] &= x_i, & \pi\langle i \rangle &= t_i, \\ \pi|_i &= x_0, t_0, \dots, t_{i-1}, x_i & \text{ and } & \pi|_i &= x_i, t_i, x_{i+1}, t_{i+1}, \dots \end{aligned}$$

Let  $\Pi(X)$  denote the set of timed paths on  $X$ .

The *cylinder set* (of rank  $n$ ) for  $X_i \subseteq X$  and  $R_i \subseteq \mathbb{R}_{\geq 0}$ ,  $i = 0..n$  is the set

$$\mathfrak{C}(X_0, R_0, \dots, R_{n-1}, X_n) = \{\pi \in \Pi(X) \mid \pi|_n \in X_0 \times R_0 \times \dots \times R_{n-1} \times X_n\}.$$

For  $\mathcal{F} \subseteq 2^X$  and  $\mathcal{I} \subseteq 2^{\mathbb{R}_{\geq 0}}$ , let

$$\mathfrak{C}_n(\mathcal{F}, \mathcal{I}) = \{\mathfrak{C}(X_0, R_0, \dots, R_{n-1}, X_n) \mid X_i \in \mathcal{F}, R_i \in \mathcal{I}\},$$

for  $n \in \mathcal{N}$  and

$$\mathfrak{C}(\mathcal{F}, \mathcal{I}) = \bigcup_{n \in \mathcal{N}} \mathfrak{C}_n(\mathcal{F}, \mathcal{I}).$$

If  $(X, \Sigma)$  is a measurable space,  $\Pi(X, \Sigma)$  denotes the measurable space of timed paths with  $\sigma$ -algebra generated by  $\mathfrak{C}(\Sigma, \mathbb{B})$ .

If  $\Sigma = \sigma(\mathcal{F})$  and  $\mathbb{B} = \sigma(\mathcal{I})$ , then  $\sigma(\mathfrak{C}(\Sigma, \mathbb{B})) = \sigma(\mathfrak{C}(\mathcal{F}, \mathcal{I}))$ . Moreover, if both  $\mathcal{F}$  and  $\mathcal{I}$  are fields, so is  $\mathfrak{C}(\mathcal{F}, \mathcal{I})$ .

Any function  $f: X \rightarrow Y$  can be stepwise extended to  $f^\omega: \Pi(X) \rightarrow \Pi(Y)$ . Note that if  $f$  is measurable, so is  $f^\omega$ .

Before concluding this section we prove a couple of "folklore" results that will be useful in what follows.

**Proposition 15.2.1.** *Let  $A \subseteq \mathbb{R}$  be a bounded nonempty set and  $\bar{A}$  its closure in the topology of  $\mathbb{R}$ . Then,*

- (i)  $\sup A \in \bar{A}$ ;
- (ii)  $\sup A = \sup \bar{A}$ .

*Proof.* First, notice that since  $A \neq \emptyset$  and is bounded, by Dedekind axiom, the supremum of  $A$  (and  $\bar{A}$ ) in  $\mathbb{R}$  exists. Moreover, recall that, for any  $B \subseteq \mathbb{R}$ ,

$$\bar{B} = ad(B) := \{x \in \mathbb{R} \mid \forall \varepsilon > 0. (x - \varepsilon, x + \varepsilon) \cap B \neq \emptyset\},$$

where  $ad(B)$  denotes the set of points *adherent* to  $B$ .

Let  $\alpha = \sup A$ .

(i) We prove that  $\alpha \in \overline{A}$ .

Let  $\varepsilon > 0$ , then  $\alpha - \varepsilon$  is not an upper bound for  $A$ . This means that there exists  $x \in A$  such that  $\alpha - \varepsilon < x \leq \alpha$  and, in particular, that  $x \in (\alpha - \varepsilon, \alpha + \varepsilon) \cap A$ . Therefore  $\alpha \in \overline{A}$ .

(ii) Let  $\beta = \sup \overline{A}$ .

By  $A \subseteq \overline{A} = \overline{\overline{A}}$  and (i), we have  $\alpha \leq \beta \in \overline{A}$ . We prove that  $\alpha = \beta$ .

Assume by contradiction that  $\alpha \neq \beta$  and let  $\varepsilon := \beta - \alpha$ . Clearly  $\varepsilon > 0$ , so that, by  $\beta \in \overline{A}$ , we have that  $(\beta - \varepsilon, \beta + \varepsilon) \cap A \neq \emptyset$ . This means that there exists  $x \in A$  such that  $\alpha = \beta - \varepsilon < x$ , in contradiction with the hypothesis that  $\alpha = \sup A$ .  $\square$

**Proposition 15.2.2.** *Let  $f: X \rightarrow Y$  be a continuous function and  $A \subseteq X$ . Then,  $\overline{f(A)} = f(\overline{A})$ , where  $\overline{A}$  denotes the closure of  $A$  in the corresponding topology.*

*Proof.* ( $\supseteq$ ) A function  $f: X \rightarrow Y$  is continuous iff for all  $B \subseteq X$ ,  $f(\overline{B}) \subseteq \overline{f(B)}$ . Therefore,  $f(\overline{A}) \subseteq \overline{f(A)}$ . Since  $\overline{f(A)}$  is closed, we have  $f(\overline{A}) \subseteq \overline{f(A)}$ .

( $\subseteq$ ) The result follows by  $A \subseteq \overline{A}$  and monotonicity of  $f(\cdot)$  and  $\overline{(\cdot)}$ .  $\square$

**Proposition 15.2.3.** *Let  $X$  be nonempty,  $f: X \rightarrow \mathbb{R}$  be a bounded continuous real-valued function, and  $D \subseteq X$  be dense in  $X$ . Then  $\sup f(D) = \sup f(X)$ .*

*Proof.* Notice that, since  $X \neq \emptyset$  and  $f$  is bounded, by Dedekind axiom, both  $\sup f(D)$  and  $\sup f(X)$  exist. By Propositions 15.2.1, 15.2.2, and  $\overline{D} = X$ , we have

$$\sup f(D) \stackrel{(\text{Prop.15.2.1})}{=} \overline{\sup f(D)} \stackrel{(\text{Prop.15.2.2})}{=} \overline{\sup f(\overline{D})} = \overline{\sup f(X)} \stackrel{(\text{Prop.15.2.1})}{=} \sup f(X),$$

which proves the thesis.  $\square$

## 15.3 Semi-Markov Chains and Trace Distance

In this section we introduce the labelled *semi-Markov chains* (SMCs), models that subsume most of the space-finite Markovian models including Markov chains (MCs) and continuous-time Markov chains (CTMCs). We define the total variation distance between SMCs, called *trace distance*, which measures the difference between two SMCs w.r.t. their probabilities of generating labelled timed traces.

In what follows we fix a countable set  $\mathbb{A}$  of atomic properties.

**Definition 15.3.1** (Semi-Markov Chains). *A labelled semi-Markov chain is a tuple  $\mathcal{M} = (S, \tau, \rho, \ell)$  consisting of*

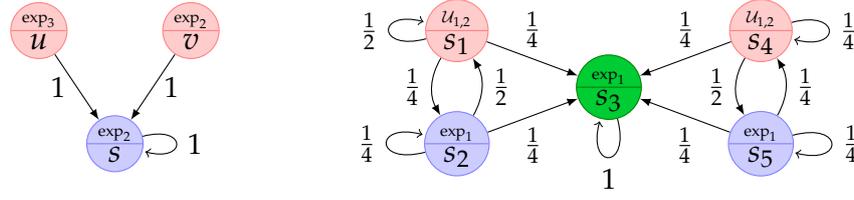


Figure 15.1: Two SMCs. (left) the differences are only in the residence time distributions; (right) the behavioral differences arise only from their transition distributions.

- a finite set  $S$  of states,
- a transition probability function  $\tau: S \rightarrow \mathcal{D}(S)$ ,
- a residence-time probability function  $\rho: S \rightarrow \Delta(\mathbb{R}_{\geq 0})$ ,
- a labelling function  $\ell: S \rightarrow 2^{\mathcal{A}}$ .

In what follows we use  $\mathcal{M} = (S, \tau, \rho, \ell)$  to range over the class of SMCs.

Intuitively, if  $\mathcal{M}$  is in the state  $s$ , it moves to an arbitrary state  $s' \in S$  within time  $t \in \mathbb{R}_{\geq 0}$  with probability  $\rho(s)([0, t]) \cdot \tau(s)(s')$ .

For example, in Fig. 15.1(right) the SMC moves from  $s_1$  to  $s_2$  before time  $t > 0$  with probability  $\frac{1}{4} \cdot U[1, 2]([0, t])$ , where  $U[i, j]$  is the uniform distribution on  $[i, j]$ .

An atomic proposition  $p \in \mathcal{A}$  is said to hold in  $s$  if  $p \in \ell(s)$ .

Notice that MCs are the SMCs s.t. for all  $s \in S$ ,  $\rho(s)$  is the Dirac measure at 0 (transitions happen instantaneously); while CTMCs are the SMCs s.t. for all  $s \in S$ ,  $\rho(s) = \text{Exp}(\lambda)$  —the exponential distribution with rate  $\lambda > 0$ .

An SMC in an initial state is a stochastic processes generating timed paths. They are distributed as in the next definition.

**Definition 15.3.2.** Given an SMC  $\mathcal{M}$  and  $s \in S$  a state in  $\mathcal{M}$ , let  $\mathbb{P}_s$  be the unique probability measure<sup>1</sup> on  $\Pi(S)$  such that for all  $s_i \in S$  and  $R_i \in \mathbb{B}$ ,  $i = 0..n$ ,

$$\mathbb{P}_s(\mathfrak{C}(\{s_0\}, R_0, \dots, R_{n-1}, \{s_n\})) = \mathbb{1}_{\{s\}}(s_0) \cdot \prod_{i=0}^{n-1} P(s_i, R_i, s_{i+1}),$$

where  $\mathbb{1}_A$  is the indicator function of  $A$  and  $P(u, R, v) = \rho(u)(R) \cdot \tau(u)(v)$ .

Since the only things that we observe in a state of an SMC are the atomic properties (labels), timed paths are considered up to label equivalence. This leads to the definition of *trace cylinders*, which are elements in  $\mathfrak{C}(S/\equiv_\ell, \mathbb{B})$ , and to the following equivalence between states.

<sup>1</sup>Existence and uniqueness of  $\mathbb{P}_s$  is guaranteed by the Hahn-Kolmogorov extension theorem and by the fact that, for all  $s \in S$ ,  $\tau(s)$  and  $\rho(s)$  are finite measures.

**Definition 15.3.3** (Trace Equivalence). *For an arbitrary SMC  $\mathcal{M} = (S, \tau, \rho, \ell)$ , the states  $s, s' \in S$  are trace equivalent, written  $s \approx s'$ , if*

$$\text{for all } T \in \mathfrak{C}(S/\equiv_\ell, \mathbb{B}), \mathbb{P}_s(T) = \mathbb{P}_{s'}(T).$$

Hereafter, we use  $\mathcal{T}$  to denote the set  $\mathfrak{C}(S/\equiv_\ell, \mathbb{B})$  of trace cylinders.

If two states of an SMCs are *not* trace equivalent, then their difference is usually measured by the total variation distance between their corresponding probabilities restricted to labelled traces.

**Definition 15.3.4** (Trace Pseudometric). *Given  $\mathcal{M} = (S, \tau, \rho, \ell)$ , the trace pseudometric  $\delta_\lambda: S \times S \rightarrow [0, 1]$  is defined, for arbitrary  $s, s' \in S$ , by*

$$\delta_\lambda(s, s') = \sup_{E \in \sigma(\mathcal{T})} |\mathbb{P}_s(E) - \mathbb{P}_{s'}(E)|.$$

It is not difficult to observe that two states  $s, s' \in S$  are trace equivalent if and only if  $\delta_\lambda(s, s') = 0$ . This demonstrates that the trace equivalence is a *behavioural distance*.

## 15.4 Trace Distance and Probabilistic Model Checking

In this section we investigate the connections between trace distance and model checking SMCs over linear real-time specifications.

We show that the variational distance over measurable sets expressed either as Metric Temporal Logic (MTL) formulas or as languages accepted by Timed Automata (TAs) coincides with the trace distance introduced in the previous section. Both these results are instances of a more general result (Theorem 15.4.2), which also entails other similar nontrivial characterizations of the trace distance.

A measure  $\mu$  on the measurable space  $(X, \Sigma)$  induces the so-called *Fréchet-Nikodym pseudometric* on  $\Sigma$ ,  $d_\mu: \Sigma \times \Sigma \rightarrow \mathbb{R}_{\geq 0}$  defined for arbitrary  $E, F \in \Sigma$ , by

$$d_\mu(E, F) = \mu(E \triangle F),$$

where  $E \triangle F := (E \setminus F) \cup (F \setminus E)$  is the symmetric difference between sets<sup>2</sup>.

Recall that in a (pseudo)metric space a subset  $D$  is dense if its closure  $\overline{D}$  (i.e., the set of all the points arbitrarily close to  $D$ ) coincides with the entire space. In order to prove the aforementioned general result, we need firstly to provide a sufficient condition for a family of measurable sets to be dense with respect to the Fréchet-Nikodym pseudometric for some finite measure.

---

<sup>2</sup>Triangular inequality follows by monotonicity and sub-additivity of  $\mu$  noticing that,  $A \triangle C \subseteq (A \triangle B) \cup (B \triangle C)$ .

**Lemma 15.4.1.** *Let  $(X, \Sigma)$  be a measurable space and  $\mu$  be a finite measure on it. If  $\Sigma$  is generated by a field  $\mathcal{F}$ , then  $\mathcal{F}$  is dense in the pseudometric space  $(\Sigma, d_\mu)$ .*

*Proof.* The closure of  $\mathcal{F}$  under  $d_\mu$  is given by

$$\overline{\mathcal{F}} = \{E \in \Sigma \mid \forall \varepsilon > 0. \exists F \in \mathcal{F}. d_\mu(E, F) < \varepsilon\}.$$

We show that  $\overline{\mathcal{F}} = \Sigma$ .

Clearly,  $\overline{\mathcal{F}} \subseteq \Sigma$ .

The converse inclusion follows from  $\mathcal{F} \subseteq \overline{\mathcal{F}}$  and  $\Sigma = \sigma(\mathcal{F})$ , if we show that  $\overline{\mathcal{F}}$  is a  $\sigma$ -algebra. This is what we prove below.

- *Complement.* Let  $E \in \overline{\mathcal{F}}$ . We want to show that  $E^c \in \overline{\mathcal{F}}$ , where  $E^c := \Sigma \setminus E$  denotes the complement of  $E$  in  $\Sigma$ .

Let  $\varepsilon > 0$ . From  $E \in \overline{\mathcal{F}}$ , there exists  $F \in \mathcal{F}$  such that  $d_\mu(E, F) < \varepsilon$ . Moreover, note that  $E \Delta F = E^c \Delta F^c$ , so

$$d_\mu(E^c, F^c) = \mu(E^c \Delta F^c) = \mu(E \Delta F) = d_\mu(E, F),$$

and  $d_\mu(E^c, F^c) < \varepsilon$ . By hypothesis,  $\mathcal{F}$  is a field, hence  $F^c \in \mathcal{F}$ . Due to the generality of choosing  $\varepsilon > 0$ , we get that  $E^c \in \overline{\mathcal{F}}$ .

- *Countable Union.* Let  $\{E_i \mid i \in \mathcal{N}\} \subseteq \overline{\mathcal{F}}$ . We want to show that  $\bigcup_{i \in \mathcal{N}} E_i \in \overline{\mathcal{F}}$ .

Let  $\varepsilon > 0$ . To prove the thesis it suffices to show that following statements hold:

- there exists  $k \in \mathcal{N}$ , such that  $d_\mu(\bigcup_{i \in \mathcal{N}} E_i, \bigcup_{i=0}^k E_i) < \frac{\varepsilon}{2}$ ;
- for all  $n \in \mathcal{N}$ , there exist  $F_0, \dots, F_n \in \mathcal{F}$ , such that  $d_\mu(\bigcup_{i=0}^n E_i, \bigcup_{i=0}^n F_i) < \frac{\varepsilon}{2}$ .

Indeed, by applying the triangular inequality on (a) and (b), we have that there exist  $k \in \mathcal{N}$  and  $F_0, \dots, F_k \in \mathcal{F}$  such that

$$d_\mu(\bigcup_{i \in \mathcal{N}} E_i, \bigcup_{i=0}^k F_i) \leq d_\mu(\bigcup_{i \in \mathcal{N}} E_i, \bigcup_{i=0}^k E_i) + d_\mu(\bigcup_{i=0}^k E_i, \bigcup_{i=0}^k F_i) < \varepsilon.$$

Since, by hypothesis,  $\mathcal{F}$  is a field, we also have that  $\bigcup_{i=0}^k F_i \in \mathcal{F}$ . Therefore, due to the generality of  $\varepsilon > 0$ , we will obtain that  $\bigcup_{i \in \mathcal{N}} E_i \in \overline{\mathcal{F}}$ .

(a). Since  $(\bigcup_{i=0}^n E_i)_{n \in \mathcal{N}}$  is a countable increasing sequence in  $\Sigma$  converging to  $\bigcup_{i \in \mathcal{N}} E_i$ , by  $\omega$ -continuity from below of  $\mu$ , we have that  $(\mu(\bigcup_{i=0}^n E_i))_{n \in \mathcal{N}}$  converges in  $\mathbb{R}$  to  $\mu(\bigcup_{i \in \mathcal{N}} E_i)$ . This means that there exists an index  $k \in \mathcal{N}$  such that

$$|\mu(\bigcup_{i \in \mathcal{N}} E_i) - \mu(\bigcup_{i=0}^k E_i)| < \frac{\varepsilon}{2}$$

Using the fact that  $\bigcup_{i \in \mathcal{N}} E_i \subseteq \bigcup_{i=0}^k E_i$  and the monotonicity, the additivity and the finiteness of  $\mu$ ,

$$\begin{aligned} d_\mu(\bigcup_{i \in \mathcal{N}} E_i, \bigcup_{i=0}^k E_i) &= \mu(\bigcup_{i \in \mathcal{N}} E_i \Delta \bigcup_{i=0}^k E_i) \\ &= \mu(\bigcup_{i \in \mathcal{N}} E_i \setminus \bigcup_{i=0}^k E_i) \\ &= \mu(\bigcup_{i \in \mathcal{N}} E_i) - \mu(\bigcup_{i=0}^k E_i) < \frac{\varepsilon}{2}. \end{aligned}$$

(b). Let  $n \in \mathcal{N}$ . By  $E_0, \dots, E_n \in \overline{\mathcal{F}}$ , there exists  $F_0, \dots, F_n \in \mathcal{F}$  such that

$$d_\mu(E_i, F_i) < \frac{\varepsilon}{2n}.$$

Moreover, note that

$$\bigcup_{i=0}^n E_i \Delta \bigcup_{i=0}^n F_i \subseteq \bigcup_{i=0}^n (E_i \Delta F_i),$$

so that by monotonicity and sub-additivity of  $\mu$  we have

$$\begin{aligned} d_\mu(\bigcup_{i=0}^n E_i, \bigcup_{i=0}^n F_i) &= \mu(\bigcup_{i=0}^n E_i \Delta \bigcup_{i=0}^n F_i) \\ &\leq \mu(\bigcup_{i=0}^n (E_i \Delta F_i)) \\ &\leq \sum_{i=0}^n \mu(E_i \Delta F_i) < \sum_{i=0}^n \frac{\varepsilon}{2n} = \frac{\varepsilon}{2}. \end{aligned}$$

□

With this result in hands we can state the main theorem of this section.

**Theorem 15.4.2.** *Let  $(X, \Sigma)$  be a measurable space and  $\mu, \nu$  be two finite measures on it. If  $\Sigma$  is generated by a field  $\mathcal{F}$ , then*

$$\|\mu - \nu\| = \sup_{E \in \mathcal{F}} |\mu(E) - \nu(E)|.$$

*Proof.* For  $Y \neq \emptyset$  and  $f: Y \rightarrow \mathbb{R}$  bounded and continuous, if  $D \subseteq Y$  is dense then  $\sup f(D) = \sup f(Y)$ . By Lemma 15.4.1,  $\mathcal{F}$  is dense in  $(\Sigma, d_{\mu+\nu})$ . We show that  $|\mu - \nu|: \Sigma \rightarrow \mathbb{R}$  is bounded and continuous. Let  $E$  and  $F$  be arbitrary measurable sets in  $\Sigma$ , then

$$\begin{aligned} \mu(E) &= \mu(E \setminus F) + \mu(E \cap F) && (\mu \text{ additive}) \\ &\leq \mu((E \setminus F) \cup (F \setminus E)) + \mu(F) && (\mu \text{ monotone}) \\ &= \mu(E \Delta F) + \mu(E) && (\text{by def}) \\ &= \mu(E \Delta F) + \nu(E \Delta F) + \mu(F) && (\nu \text{ positive}) \\ &= d_{\mu+\nu}(E, F) + \mu(F). && (\text{by def}) \end{aligned}$$

This implies that, for all  $E, F \in \Sigma$ ,

$$d_{\mu+\nu}(E, F) \geq |\mu(E) - \mu(F)|,$$

hence  $\mu: \Sigma \rightarrow \mathbb{R}$  is 1-Lipschitz continuous. Analogously, also  $\nu: \Sigma \rightarrow \mathbb{R}$  is 1-Lipschitz continuous. Then, continuity of  $|\mu - \nu|: \Sigma \rightarrow \mathbb{R}$  follows by composition of continuous functions. Moreover,  $|\mu - \nu|$  is bounded because, by hypothesis,  $\mu$  and  $\nu$  are finite. □

### 15.4.1 Model Checking for MTL Formulas

Metric Temporal Logic [2] has been introduced as a formalism for reasoning on sequences of events in a real-time setting. The grammar of formulas is as follows

$$\varphi ::= p \mid \perp \mid \varphi \rightarrow \varphi \mid X^{[t,t']}\varphi \mid \varphi U^{[t,t']}\varphi,$$

where  $p \in \mathbb{A}$  and  $[t, t']$  are positive-real intervals with rational boundaries. MTL extends the linear-temporal logic operators *next* and *until* of LTL with a real time constraint expressed in the form of a closed interval within which the modal observation is realized.

The formal semantics<sup>3</sup> of MTL is given by means of a satisfiability relation defined, for an arbitrary SMC  $\mathcal{M}$  and a timed path  $\pi \in \Pi(S)$ , as follows [38].

$$\begin{array}{ll} \mathcal{M}, \pi \models p & \text{if } p \in \ell(\pi[0]), \\ \mathcal{M}, \pi \models \perp & \text{never,} \\ \mathcal{M}, \pi \models \varphi \rightarrow \psi & \text{if } \mathcal{M}, \pi \models \psi \text{ whenever } \mathcal{M}, \pi \models \varphi, \\ \mathcal{M}, \pi \models X^{[t,t']}\varphi & \text{if } \pi\langle 0 \rangle \in [t, t'], \text{ and } \mathcal{M}, \pi|_1 \models \varphi, \\ \mathcal{M}, \pi \models \varphi U^{[t,t']}\psi & \text{if } \exists i > 0 \text{ such that } \sum_{k=0}^{i-1} \pi\langle k \rangle \in [t, t'], \mathcal{M}, \pi|_i \models \psi, \\ & \text{and } \mathcal{M}, \pi|_j \models \varphi \text{ whenever } 0 \leq j < i. \end{array}$$

Having fixed an SMC  $\mathcal{M}$ , let

$$\llbracket \varphi \rrbracket = \{ \pi \mid \mathcal{M}, \pi \models \varphi \}$$

and

$$\llbracket \mathcal{L} \rrbracket = \{ \llbracket \varphi \rrbracket \mid \varphi \in \mathcal{L} \},$$

for any  $\mathcal{L} \subseteq \text{MTL}$ . Let  $\text{MTL}^-$  be the fragment of MTL without until operator.

**Lemma 15.4.3.** *With the previous notations, the following statements hold.*

1.  $\llbracket \text{MTL} \rrbracket \subseteq \sigma(\mathcal{T})$
2.  $\mathcal{T} \subseteq \sigma(\llbracket \text{MTL}^- \rrbracket)$ .

*Proof.* We prove the two statements separately.

1. By structural induction on the syntax of  $\varphi \in \text{MTL}$  we prove that  $\llbracket \varphi \rrbracket \in \sigma(\mathcal{T})$ .

---

<sup>3</sup>This is known as the *point-based* semantics, since the connectives quantify over a countable set of positions in the path; it differs from the *interval-based* semantics, adopted in [15, 39], which associates a state with each point in the real line, and let the temporal connectives quantify over intervals with uncountable many points.

**Atomic prop.**  $\llbracket a \rrbracket = \{\pi \mid a \in \ell(\pi[0])\} = \bigcup \{\mathfrak{C}([s]_{\equiv_\ell}) \mid s \in \ell^{-1}(\{a\})\}$ . Since  $S$  is finite and  $\mathfrak{C}([s]_{\equiv_\ell}) \in \mathcal{T}$  for all  $s \in S$ , then  $\llbracket a \rrbracket \in \sigma(\mathcal{T})$ .

**False.**  $\llbracket \perp \rrbracket = \emptyset \in \sigma(\mathcal{T})$ .

**Implication.**  $\llbracket \varphi \rightarrow \psi \rrbracket = \llbracket \neg\varphi \vee \psi \rrbracket = \llbracket \varphi \rrbracket^c \cup \llbracket \psi \rrbracket$ . By inductive hypothesis,  $\llbracket \varphi \rrbracket, \llbracket \psi \rrbracket \in \sigma(\mathcal{T})$ , therefore  $\llbracket \varphi \rightarrow \psi \rrbracket \in \sigma(\mathcal{T})$ .

**Next.** Consider  $X^I \varphi$ . The following hold

$$\begin{aligned} \llbracket X^I \varphi \rrbracket &= \{\pi \mid \pi \langle 0 \rangle \in I, \text{ and } \mathcal{M}, \pi|_1 \models \varphi\} && \text{(by def. of } X) \\ &= \{\pi \mid \pi \langle 0 \rangle \in I, \text{ and } \pi|_1 \in \llbracket \varphi \rrbracket\} && \text{(by def. of } \llbracket \cdot \rrbracket) \\ &= (\cdot) \langle 0 \rangle^{-1}(I) \cap (\cdot)|_1^{-1}(\llbracket \varphi \rrbracket) && \text{(by def. of } (\cdot) \langle 0 \rangle \text{ and } (\cdot)|_1) \end{aligned}$$

By inductive hypothesis and the fact that both  $(\cdot) \langle 0 \rangle$  and  $(\cdot)|_1$  are measurable functions, it follows that  $\llbracket X^I \varphi \rrbracket \in \sigma(\mathcal{T})$ .

**Until.** Consider  $\llbracket \varphi U^{[a,b]} \psi \rrbracket$ . For  $k > 0$  we define the set *OnTime@k* as

$$\text{OnTime@}k = \bigcup \left\{ \mathfrak{C}(X) \left[ \begin{array}{l} C_i \in S / \equiv_\ell, t_i^-, t_i^+ \in \mathbb{Q}_+, \text{ for } 0 \leq i \leq k, \\ \sum_{i=0}^{k-1} t_i^- \geq a, \sum_{i=0}^{k-1} t_i^+ \leq b, t_i^- \leq t_i^+, \\ X = C_0, [t_0^-, t_0^+], \dots, [t_{k-1}^-, t_{k-1}^+], C_n \end{array} \right. \right\}$$

Notice that *OnTime@k* is a countable union of cylinders in  $\mathcal{T}$  (the number of unions is bounded by  $|(S \times \mathbb{Q}_+^2)^{k+1}|$ ), hence it is a measurable set in  $\sigma(\mathcal{T})$ .

$$\text{OnTime@}k = \{\pi \mid \forall i < k. \sum_{i=0}^{k-1} \pi \langle i \rangle \in [a, b]\} \quad (15.4.1)$$

The inclusion from left to right trivially holds by definition of *OnTime@k*.

As for the reverse inclusion, let  $\pi$  be a timed path over  $S$ , such that  $\pi \langle i \rangle = t_i$  ( $i = 0..k-1$ ) and  $\sum_{i=0}^{k-1} t_i \in [a, b]$ .

We have to prove that there exist  $t_i^-, t_i^+ \in \mathbb{Q}_+$  such that  $t_i^- \leq t_i \leq t_i^+$ ,  $\sum_{i=0}^{k-1} t_i^- \geq a$ , and  $\sum_{i=0}^{k-1} t_i^+ \leq b$ .

When  $k = 1$  it suffices to take  $t_1^- = a$  and  $t_1^+ = b$ .

Assume  $k > 1$ . Let  $\Delta = 2h/10^h$  for some  $h \in \mathcal{N}$  large enough to satisfy the following two inequalities  $\sum_{i=0}^{k-1} t_i - \Delta > a$  and  $\sum_{i=0}^{k-2} t_i + \Delta < b$ . Let  $t_i^- = t_i = t_i^+$  if  $t_i \in \mathbb{Q}_+$ , otherwise we choose some  $t_i^-, t_i^+ \in \mathbb{Q}_+$  that satisfy

$$t_i^- < t_i < t_i^+, \quad t_i^- > t_i - \Delta/2k \quad \text{and} \quad t_i^+ < t_i + \Delta/2k. \quad (15.4.2)$$

We proceed by showing that the constraints (15.4.2) are sufficient to prove that

$$\sum_{i=0}^{k-1} t_i^- \geq a \quad \text{and} \quad \sum_{i=0}^{k-1} t_i^+ \leq b,$$

then we show how to pick  $t_i^-, t_i^+ \in \mathbb{Q}_+$  in order to satisfy (15.4.2). The following hold

$$\sum_{i=0}^{k-1} t_i - \Delta < \sum_{i=0}^{k-1} (t_i^- + \Delta/2k) - \Delta = \quad (\text{by (15.4.2)})$$

$$= \sum_{i=0}^{k-1} t_i^- - \Delta/2 \leq \sum_{i=0}^{k-1} t_i^-. \quad (\text{by } \Delta \geq 0)$$

By construction,  $\sum_{i=0}^{k-1} t_i - \Delta > a$ , hence  $\sum_{i=0}^{k-1} t_i^- > a$ . Analogously,

$$\sum_{i=0}^{k-2} t_i + \Delta > \sum_{i=0}^{k-1} (t_i^+ - \Delta/2k) - \Delta = \quad (\text{by (15.4.2)})$$

$$= \sum_{i=0}^{k-2} t_i^+ + (k+1)\Delta/2k \geq \sum_{i=0}^{k-2} t_i^+. \quad (\text{by } \Delta \geq 0)$$

By construction,  $\sum_{i=0}^{k-2} t_i + \Delta < b$ , hence  $\sum_{i=0}^{k-2} t_i^+ < b$ .

One can check that the constraints (15.4.2) are easily satisfied if we pick

$$t_i^- = \lfloor t_i \rfloor + \frac{\lfloor 10^h \cdot \{t_i\} \rfloor}{10^h}, \quad t_i^+ = \lfloor t_i \rfloor + \frac{\lfloor 10^h \cdot \{t_i\} \rfloor + 1}{10^h},$$

for some large enough  $h \in \mathcal{N}$ , where  $\{t_i\}$  denotes the fractional part of  $t_i \notin \mathbb{Q}_+$ . This proves (15.4.1).

$$\begin{aligned} & \llbracket \varphi U^{[a,b]} \psi \rrbracket \\ &= \left\{ \pi \left| \begin{array}{l} \exists i > 0. \sum_{k=0}^{i-1} \pi \langle i \rangle \in [a, b], \text{ and } \mathcal{M}, \pi|_i \models \psi, \\ \forall 0 \leq j < i. \mathcal{M}, \pi|_j \models \varphi \end{array} \right. \right\} \quad (\text{by def. } U) \\ &= \left\{ \pi \left| \begin{array}{l} \exists i > 0. \sum_{k=0}^{i-1} \pi \langle i \rangle \in [a, b], \text{ and } \pi|_i \in \llbracket \psi \rrbracket, \\ \forall 0 \leq j < i. \pi|_j \in \llbracket \varphi \rrbracket \end{array} \right. \right\} \quad (\text{by def. } \llbracket \cdot \rrbracket) \\ &= \bigcup_{i>0} \bigcap_{0 \leq j < i} ((\cdot)|_j^{-1}(\llbracket \varphi \rrbracket) \cap (\cdot)|_i^{-1}(\llbracket \psi \rrbracket) \cap \text{OnTime}@i). \quad (\text{by def. } (\cdot)|_k \text{ and (15.4.1)}) \end{aligned}$$

By inductive hypothesis on  $\varphi, \psi$  and measurability of  $(\cdot)|_k$  for arbitrary  $k \in \mathcal{N}$ , it follows that  $\llbracket \varphi U^{[a,b]} \psi \rrbracket \in \sigma(\mathcal{T})$ .

2. We show that  $\sigma(\mathcal{T}) \subseteq \sigma(\llbracket \text{MTL}^- \rrbracket)$ .

Let  $\mathcal{I}$  be the family of closed intervals in  $\mathbb{R}_{\geq 0}$  with rational endpoints. It is standard that  $\sigma(\mathcal{I}) = \mathbb{B}$ , and from it one can easily verify  $\sigma(\mathfrak{C}(S/\equiv_\ell, \mathcal{I})) = \sigma(\mathcal{T})$ .

Therefore, to prove  $\sigma(\mathcal{T}) \subseteq \sigma(\llbracket \text{MTL}^- \rrbracket)$ , it suffices to show  $\mathfrak{C}(S/\equiv_\ell, \mathcal{I}) \subseteq \sigma(\llbracket \text{MTL}^- \rrbracket)$ .

Let define  $Ap: \mathbb{A} \times \mathfrak{C}(S/\equiv_\ell, \mathcal{I}) \rightarrow \text{MTL}^-$  as

$$Ap(a, \mathfrak{C}(C)) = \begin{cases} a & \text{if } C \subseteq \ell^{-1}(a) \\ \neg a & \text{otherwise} \end{cases}$$

$$Ap(a, \mathfrak{C}(C, I, X)) = \begin{cases} a \wedge X^I Ap(a, \mathfrak{C}(X)) & \text{if } C \subseteq \ell^{-1}(a) \\ \neg a \wedge X^I Ap(a, \mathfrak{C}(X)) & \text{otherwise,} \end{cases}$$

Let  $C = \mathfrak{C}(C_0, I_0, \dots, I_{n-1}, C_n) \in \mathfrak{C}(S/\equiv_\ell, \mathcal{I})$ , one can prove by induction on  $n$  that

$$\bigcap_{a \in \mathbb{A}} \llbracket Ap(a, C) \rrbracket = C.$$

Since  $\sigma(\llbracket \text{MTL}^- \rrbracket)$  is closed under countable intersection, we conclude that  $C \in \sigma(\llbracket \text{MTL}^- \rrbracket)$ .  $\square$

Lemma 15.4.3 states that (1) MTL formulas describe events in the  $\sigma$ -algebra generated by the trace cylinders; and (2) the trace cylinders are measurable sets generated by MTL formulas without until operator. Consequently, the probabilistic model checking problem for SMC, which is to determine the probability  $\mathbb{P}_s(\llbracket \varphi \rrbracket)$  given the initial state  $s$  of  $\mathcal{M}$ , is well defined. Moreover, for any  $\mathcal{L} \subseteq \text{MTL}$ ,

$$\delta_{\mathcal{L}}(s, s') = \sup_{\varphi \in \mathcal{L}} |\mathbb{P}_s(\llbracket \varphi \rrbracket) - \mathbb{P}_{s'}(\llbracket \varphi \rrbracket)|$$

is a well-defined pseudometric that distinguishes states w.r.t. their maximal difference in the likelihood of satisfying formulas in  $\mathcal{L}$ .

Obviously, the trace distance  $\delta_\lambda$  is an upper bound of  $\delta_{\mathcal{L}}$ ; however, Theorem 15.4.2 reveals a set of conditions on  $\mathcal{L}$  guaranteeing that the two actually coincide.

**Corollary 15.4.4** (Logical Characterization). *Let  $\mathcal{L}$  be a Boolean-closed fragment of MTL s.t.  $\mathcal{T} \subseteq \sigma(\llbracket \mathcal{L} \rrbracket)$ . Then,  $\delta_{\mathcal{L}} = \delta_\lambda$ . In particular,  $\delta_{\text{MTL}} = \delta_{\text{MTL}^-} = \delta_\lambda$ .*

*Proof.*  $\mathcal{L}$  is closed under all Boolean operators, therefore  $\llbracket \mathcal{L} \rrbracket$  is a field.

Applying Lemma 15.4.3(1), since  $\mathcal{L} \subseteq \text{MTL}$ , we get  $\llbracket \mathcal{L} \rrbracket \subseteq \sigma(\mathcal{T})$ .

Because  $\mathcal{T} \subseteq \sigma(\llbracket \mathcal{L} \rrbracket)$ , it follows that  $\sigma(\llbracket \mathcal{L} \rrbracket) = \sigma(\mathcal{T})$ . The equality  $\delta_{\mathcal{L}} = \delta_\lambda$  now follows by Theorem 15.4.2.

In particular  $\delta_{\text{MTL}} = \delta_{\text{MTL}^-} = \delta_\lambda$  follows from Lemma 15.4.3(2).  $\square$

**Remark 15.4.5.** The supremum in Corollary 15.4.4 is not a maximum. Fig.15.1 shows two examples. The SMC on the right is taken from [17, Example 1]<sup>4</sup>, where it is proven that  $\delta_\lambda(s_1, s_4)$  has a maximizing event that is not an  $\omega$ -regular language. As for the SMC on the left, the maximizing event corresponding to  $\delta_\lambda(u, v)$  should have the form  $X^I \top$  for  $I = [0, \log(3) - \log(2)]$ . However, the previous is not an MTL formula since  $I$  has an irrational endpoint. ■

## 15.4.2 Model Checking for Timed Automata

Timed Automata (TAs) [1] have been introduced to model the behavior of real-time systems over time. Here we consider TAs without location invariants.

Let  $\mathcal{X}$  be a finite set of variables (*clocks*) and  $\mathcal{V}(\mathcal{X})$  the set of *valuations*  $v: \mathcal{X} \rightarrow \mathbb{R}_{\geq 0}$ . As usual, for  $v \in \mathcal{V}(\mathcal{X})$ ,  $t \in \mathbb{R}_{\geq 0}$  and  $X \subseteq \mathcal{X}$ , we denote by  $\mathbf{0}$  the null valuation, by  $v + t$  the  $t$ -delay of  $v$  and by  $v[X := t]$  the update of  $X$  in  $v$ .

A *clock guard*  $g \in \mathcal{G}(\mathcal{X})$  over  $\mathcal{X}$  is a finite set of expressions of the form  $x \bowtie q$ , for  $x \in \mathcal{X}$ ,  $q \in \mathbb{Q}_+$  and  $\bowtie \in \{<, \leq, >, \geq\}$ .

We say that a valuation  $v \in \mathcal{V}(\mathcal{X})$  *satisfies* a clock guard  $g \in \mathcal{G}(\mathcal{X})$ , written  $v \models g$ , if  $v(x) \bowtie n$  holds, for all  $x \bowtie q \in g$ .

Two clock guards  $g, g' \in \mathcal{G}(\mathcal{X})$  are *orthogonal* (or *non-overlapping*), written  $g \perp g'$ , if there is no  $v \in \mathcal{V}(\mathcal{X})$  such that  $v \models g$  and  $v \models g'$ .

**Definition 15.4.6** (Timed Automaton). A timed (Muller) automaton over a set of clocks  $\mathcal{X}$  is a tuple  $\mathbb{A} = (Q, L, q_0, F, \rightarrow)$  consisting of

- a finite set  $Q$  of locations,
- a set  $L$  of input symbols,
- an initial location  $q_0 \in Q$ ,
- a family  $F \subseteq 2^Q$  of final sets of locations,
- a transition relation  $\rightarrow \subseteq Q \times L \times \mathcal{G}(\mathcal{X}) \times 2^{\mathcal{X}} \times Q$ .

$\mathbb{A}$  is deterministic if

$$[(q, a, g, X, q'), (q, a, g', X', q'')] \in \rightarrow \text{ and } g \neq g' \text{ implies } g \perp g';$$

$\mathbb{A}$  is resetting if

$$(q, a, g, X, q') \in \rightarrow \text{ implies } X = \mathcal{X}.$$

<sup>4</sup>The SMC has been adapted to the current setting where the labels are in the state, instead of in the transitions.

Given a TA  $\mathbb{A} = (Q, L, q_0, F, \rightarrow)$ , a *run* of  $\mathbb{A}$  over a timed path  $\pi = a_0, t_0, a_1, t_1, \dots$  is an infinite sequence

$$(q_0, v_0) \xrightarrow{a_0, t_0} (q_1, v_1) \xrightarrow{a_1, t_1} (q_2, v_2) \xrightarrow{a_2, t_2} \dots$$

with  $q_i \in Q$  and  $v_i \in \mathcal{V}(\mathcal{X})$  for all  $i \geq 0$ , satisfying the following requirements:

- (*initialization*)  $v_0 = \mathbf{0}$ ;
- (*consecution*) for all  $i \geq 0$ , exists  $(q_i, a_i, g_i, X_i, q_{i+1}) \in \rightarrow$  such that

$$v_{i+1} = (v_i + t_i)[X_i := 0] \quad \text{and} \quad v_i + t_i \models g_i.$$

A run over  $\pi$  is *accepting* ( $\pi$  is *accepted* by  $\mathbb{A}$ ) if the set of locations visited infinitely often is in  $F$ . Let  $\mathcal{L}(\mathbb{A})$  be the set of timed paths accepted by  $\mathbb{A}$ .

A deterministic timed automata (DTA) is a TA that has at most one accepting run over a given timed path in  $\Pi(L)$ . With respect to TAs, which are only closed under finite union and intersection, DTAs are also closed under complement [1].

To relate TAs and SMCs, consider an SMC  $\mathcal{M} = (S, \tau, \rho, \ell)$  and a TA  $\mathbb{A}$  that uses the labels of  $\mathcal{M}$  as input symbols. Let

$$\llbracket \mathbb{A} \rrbracket = \{ \pi \mid \ell^\omega(\pi) \in \mathcal{L}(\mathbb{A}) \}$$

be the set of timed paths in  $\mathcal{M}$  accepted by  $\mathbb{A}$  and for any set  $\mathcal{F} \in \text{TA}$ , let

$$\llbracket \mathcal{F} \rrbracket = \{ \llbracket \mathbb{A} \rrbracket \mid \mathbb{A} \in \mathcal{F} \}.$$

With this definition we can prove, for TAs a similar lemma with Lemma 15.4.3 that we have previously proven for MTL.

**Lemma 15.4.7.** *With the previous notations, the following statements hold.*

1.  $\llbracket \text{TA} \rrbracket \subseteq \sigma(\mathcal{T})$
2.  $\mathcal{T} \subseteq \sigma(\llbracket \text{DTA} \rrbracket)$ .

*Proof.* We prove the two statements separately.

1. This is proven in [16, Theorem 3.2] and the proof can be identically applied for our case.

2. We show  $\sigma(\mathcal{T}) \subseteq \sigma(\llbracket \text{DTA} \rrbracket)$ .

Let  $\mathcal{I}$  be the family of closed intervals in  $\mathbb{R}_{\geq 0}$  with rational endpoints. It is standard that  $\sigma(\mathcal{I}) = \mathbb{B}$ , and from it one can easily verify  $\sigma(\mathfrak{C}(S/\equiv_\ell, \mathcal{I})) = \sigma(\mathcal{T})$ . Therefore, to show  $\sigma(\mathcal{T}) \subseteq \sigma(\llbracket \text{DTA} \rrbracket)$  it suffices to prove that  $\mathfrak{C}(S/\equiv_\ell, \mathcal{I}) \subseteq \sigma(\llbracket \text{DTA} \rrbracket)$ .

Let  $C = \mathfrak{C}([s_0]_{\equiv_\ell}, I_0, \dots, I_{n-1}, [s_n]_{\equiv_\ell}) \in \mathfrak{C}(S/\equiv_\ell, \mathcal{I})$ .

We define a DTA  $\mathbb{A} = (Q, 2^{\mathcal{X}}, q_0, F, \rightarrow)$  such that  $\llbracket \mathbb{A} \rrbracket = C$ .

Let  $Q = \{q_0, \dots, q_n\}$ ,  $F = \{\{q_n\}\}$  and for a (shared) clock  $x \in \mathcal{X}$  in each guard, let

$$\rightarrow = \{(q_i, \ell(s_i), g_i, \mathcal{X}, q_{i+1}) \mid g_i = a \leq x \leq b \text{ for } I_i = [a, b], 0 \leq i \leq n\} \cup \{(q_n, l, \emptyset, \mathcal{X}, q_n) \mid l \subseteq \mathbb{A}\}.$$

It is easy to see that the only accepted timed paths  $\pi \in \mathcal{L}(\mathbb{A})$  are such that

$$\pi|^n = \ell(s_0), t_0, \dots, t_{n-1}, \ell(s_n)$$

and  $t_i \in I_i$  ( $0 \leq i \leq n-1$ ), because clocks are always resetting. And this concludes our proof.  $\square$

Lemma 15.4.7 states that the model checking problem for an SMC  $\mathcal{M}$  against a TA  $\mathbb{A}$ , which is to determine the probability  $\mathbb{P}_s(\llbracket \mathbb{A} \rrbracket)$  given the initial state  $s$  of  $\mathcal{M}$ , is well defined and for any  $\Phi \subseteq \text{TA}$  we can define the pseudometric

$$\delta_\Phi(s, s') = \sup_{\mathbb{A} \in \Phi} |\mathbb{P}_s(\llbracket \mathbb{A} \rrbracket) - \mathbb{P}_{s'}(\llbracket \mathbb{A} \rrbracket)|$$

that distinguishes states looking at a specific subclass  $\Phi$  of TA specifications. For a generic  $\Phi \subseteq \text{TA}$ , the trace distance is an upper bound of  $\delta_\Phi$ . However, Theorem 15.4.2 provides conditions that guarantee the equality of the two distances.

**Corollary 15.4.8.** *Let  $\Phi \subseteq \text{TA}$  be closed under Boolean operations and such that  $\mathcal{T} \subseteq \sigma(\llbracket \Phi \rrbracket)$ . Then,  $\delta_\Phi = \delta_\lambda$ . In particular,  $\delta_{\text{TA}} = \delta_{\text{DTA}} = \delta_\lambda$ .*

*Proof.*  $\Phi$  is closed under all Boolean operators, therefore  $\llbracket \Phi \rrbracket$  is a field.

By Lemma 15.4.7(1) since  $\Phi \subseteq \text{TA}$ , we get  $\llbracket \Phi \rrbracket \subseteq \sigma(\mathcal{T})$ .

Because  $\mathcal{T} \subseteq \sigma(\llbracket \Phi \rrbracket)$ , it follows that  $\sigma(\llbracket \Phi \rrbracket) = \sigma(\mathcal{T})$ .

The equality  $\delta_\Phi = \delta_\lambda$  follows from Theorem 15.4.2. In particular,  $\delta_{\text{DTA}} = \delta_\lambda$  follows from Lemma 15.4.7(2) and the fact that DTAs are closed under all Boolean operators [1].

The equality  $\delta_{\text{TA}} = \delta_\lambda$ , follows from  $\delta_\lambda \geq \delta_{\text{TA}} \geq \delta_{\text{DTA}}$ .  $\square$

### Single-clock Resetting DTAs.

The decidability of model checking CTMCs against TA specifications is an open problem, even for the subclass of DTAs. Recently, Chen et al. [16] provided a decidable algorithm for the case of *single-clock* DTAs (1-DTAs). In this context, an alternative characterization of the trace distance in terms of 1-DTAs is appealing.

Notice however that Corollary 15.4.8 cannot be applied, since 1-DTAs are not closed under union. However, we show that the *resetting* 1-DTAs (1-RDTA) satisfy the requirements, hence  $\delta_{1\text{-DTA}} = \delta_{1\text{-RDTA}} = \delta_\lambda$ .

**Lemma 15.4.9.** *With the previous notations, the following statements hold.*

1.  $\llbracket 1\text{-RDTA} \rrbracket$  is a field;
2.  $\mathcal{T} \subseteq \sigma(\llbracket 1\text{-RDTA} \rrbracket)$ .

*Proof.* 1. It suffices to prove that 1-RDTAs are closed under union and complement. As for the latter, one only needs to take the complement of the set of final locations (similar to [1]). Closure under union is proven via a product construction similar to [1, p.334], by noticing that the resetting condition on the automata allows one to use a single clock in the product.

2. The proof of Lemma 15.4.7(2) actually uses 1-RDTAs. □

## 15.5 General Convergence Criteria

In this section we provide sufficient conditions to construct sequences that converge, from below and from above, to the total variation distance between a generic pair of probability measures. Eventually, we instantiate these results to the specific case of the trace distance on SMCs.

### Convergence from Below.

To define a converging sequence of under-approximations of the total variation distance, we exploit Theorem 15.4.2.

**Theorem 15.5.1.** *Let  $(X, \Sigma)$  be a measurable space and  $\mu, \nu$  be probability measures on it. Let*

$$\mathcal{F}_0 \subseteq \mathcal{F}_1 \subseteq \mathcal{F}_2 \subseteq \dots$$

*be a sequence s.t.  $\mathcal{F} = \bigcup_{i \in \mathcal{N}} \mathcal{F}_i$  is a field that generates  $\Sigma$  and*

$$l_i = \sup\{|\mu(E) - \nu(E)| \mid E \in \mathcal{F}_i\}.$$

*Then,  $l_i \leq l_{i+1}$  for all  $i \in \mathcal{N}$  and*

$$\sup_{i \in \mathcal{N}} l_i = \|\mu - \nu\|.$$

*Proof.* That  $l_i \leq l_{i+1}$ , follows from  $\mathcal{F}_i \subseteq \mathcal{F}_{i+1}$ .

Because  $\mathcal{F}$  is a field such that  $\sigma(\mathcal{F}) = \Sigma$ ,  $\mu$  and  $\nu$  are finite measures and

$$\sup_{i \in \mathcal{N}} l_i = \sup_{E \in \mathcal{F}} |\mu(E) - \nu(E)|.$$

Further, Theorem 15.4.2 concludes our proof. □

We now instantiate the theorem above to obtain a sequence of under-approximations for the trace distance on SMCs. According to Theorem 15.5.1, to approximate  $\delta_\lambda$  from

below, we need an increasing sequence of families of measurable sets of timed paths whose limit is a field that generates  $\sigma(\mathcal{T})$ .

For  $k \in \mathcal{N}$ , let  $\mathcal{E}_k$  be the set of all finite unions of cylinders in  $\mathfrak{C}_k(S/\equiv_\ell, \mathfrak{R}_k)$ , where

$$\mathfrak{R}_k = \left\{ \left[ \frac{n}{2^k}, \frac{n+1}{2^k} \right) \mid 0 \leq n < k2^k \right\} \cup \{[k, \infty)\}.$$

Note that, these cylinders defined for various  $k$  are pairwise disjoint and, in particular, they form a  $\mathcal{T}$ -measurable partition of  $\Pi(S)$ . The choice is justified by the following result.

**Lemma 15.5.2.** *For all  $k \in \mathcal{N}$ ,  $\mathcal{E}_k \subseteq \mathcal{E}_{k+1}$  and  $\bigcup_{k \in \mathcal{N}} \mathcal{E}_k$  is the field generating  $\sigma(\mathcal{T})$ .*

*Proof.* For  $\mathcal{E}_k \subseteq \mathcal{E}_{k+1}$ , it suffices to prove  $\mathfrak{C}_k(S/\equiv_\ell, \mathfrak{R}_k) \subseteq \mathcal{E}_{k+1}$ . We proceed by induction on  $k \geq 0$ .

The base case is trivial.

Assume  $k > 0$  and let  $C \in \mathfrak{C}_k(S/\equiv_\ell, \mathfrak{R}_k)$ .

Note that, for any  $n \in \mathcal{N}$  such that  $0 \leq n < k2^k$ ,

$$\frac{n}{2^k} = \frac{2n}{2^{k+1}} \quad \text{and} \quad 2n < (k+1)2^{k+1}.$$

From this is immediate to prove that there exists  $\mathcal{F} \subseteq \mathfrak{C}_k(S/\equiv_\ell, \mathfrak{R}_{k+1})$  such that  $C = \bigcup \mathcal{F}$ .

Observe that  $\mathfrak{R}_{k+1}$  is a partition of  $\mathbb{R}_{\geq 0}$ . So, any  $C' = \mathfrak{C}(C_0, R_0, \dots, R_{k-1}, C_k) \in \mathfrak{C}_k(S/\equiv_\ell, \mathfrak{R}_{k+1})$  can be represented as

$$C' = \bigcup \{ \mathfrak{C}(C_0, R_0, \dots, R_{k-1}, C_k, R'', C'') \mid R'' \in \mathfrak{R}_{k+1}, C'' \in S/\equiv_\ell \}.$$

Since  $\mathfrak{R}_{k+1}$  and  $S/\equiv_\ell$  are finite, from the above we get that  $C$  can be represented as a finite union of cylinders in  $\mathfrak{C}_{k+1}(S/\equiv_\ell, \mathfrak{R}_{k+1})$ . Hence,  $C \in \mathcal{E}_{k+1}$ .

Let  $\mathcal{E} = \bigcup_{k \in \mathcal{N}} \mathcal{E}_k$ .

Since each  $\mathfrak{C}_k(S/\equiv_\ell, \mathfrak{R}_k)$  forms a finite partition of  $\Pi(S)$ , it is immediate to prove that  $\mathcal{E}_k$  is a field. Further, because the limit of an increasing sequence of fields is a field, we obtain that  $\mathcal{E}$  is a field.

It remains to show  $\sigma(\mathcal{E}) = \sigma(\mathcal{T})$ .

Clearly  $\mathcal{E} \subseteq \sigma(\mathcal{T})$ , hence  $\sigma(\mathcal{E}) \subseteq \sigma(\mathcal{T})$ .

For the converse inclusion, let  $\mathcal{R} = \bigcup_{k \in \mathcal{N}} \mathfrak{R}_k$  and recall that  $\mathbb{B} = \sigma(\mathcal{CO})$ , where

$$\mathcal{CO} = \{[q, q') \mid q < q' \in \mathbb{Q}_+\} \cup \{[q, \infty) \mid q \in \mathbb{Q}_+\}$$

is the family of left-closed right-open intervals with rational endpoints (or  $\infty$ ).

Let  $q < q' \in \mathbb{Q}_+$ , then the following hold

$$[q, q') = \bigcup \left\{ \left[ \frac{n}{2^{k'}}, \frac{n+1}{2^k} \right) \mid q \leq \frac{n}{2^k} < \frac{n+1}{2^k} \leq q', \text{ for } k \in \mathcal{N}, 0 \leq n < k2^k \right\},$$

$$[q, \infty) = \bigcup \left\{ \left[ \frac{n}{2^{k'}}, \frac{n+1}{2^k} \right) \mid q \leq \frac{n}{2^k}, \text{ for } k \in \mathcal{N}, 0 \leq n < k2^k \right\}.$$

The above suffices to prove  $\mathcal{CO} \subseteq \sigma(\mathcal{R})$ , hence  $\mathbb{B} = \sigma(\mathcal{CO}) \subseteq \sigma(\mathcal{R})$ .

This proves further that  $\sigma(\mathfrak{C}(S/\equiv_{\ell}, \mathcal{R})) \subseteq \sigma(\mathcal{T})$ . Clearly,  $\mathcal{E} \subseteq \sigma(\mathfrak{C}(S/\equiv_{\ell}, \mathcal{R}))$ , therefore  $\sigma(\mathcal{E}) \subseteq \sigma(\mathcal{T})$ .  $\square$

Given an SMC  $\mathcal{M}$ , a sequence of under-approximations of the trace distance  $\delta_\lambda$  is given, for  $k \in \mathcal{N}$ , by  $\delta_\lambda \uparrow_k: S \times S \rightarrow [0, 1]$  defined as follows.

$$\delta_\lambda \uparrow_k(s, s') = \sup \{ |\mathbb{P}_s(E) - \mathbb{P}_{s'}(E)| \mid E \in \mathcal{E}_k \}. \quad (15.5.1)$$

The next result is an immediate consequence of Lemma 15.5.2 and Theorem 15.5.1.

**Corollary 15.5.3.** *For all  $k \in \mathcal{N}$ ,  $\delta_\lambda \uparrow_k \leq \delta_\lambda \uparrow_{k+1}$  and  $\delta_\lambda = \sup_{k \in \mathcal{N}} \delta_\lambda \uparrow_k$ .*

Theorem 15.5.1 suggests alternative constructions of convergent sequences. For example, as lower-approximations of  $\delta_\lambda$  one can use the pseudometrics  $\delta_{\text{MTL}_k^-}$ , where  $\text{MTL}_k^-$  is the set of  $\text{MTL}^-$  formulas of *modal depth* at most  $k \in \mathcal{N}$ .

### Convergence from Above.

The construction of the converging sequence of over-approximations of the total variation is based on a classic duality result asserting that

*the total variation of two measures corresponds to the minimal discrepancy measured among all their possible couplings [33].*

Recall that a coupling  $\omega \in \Omega(\mu, \nu)$  for two probability measures  $\mu, \nu$  on  $(X, \Sigma)$  is a measure in the product space  $(X, \Sigma) \otimes (X, \Sigma)$  whose left and right marginals are  $\mu$  and  $\nu$ , respectively. The *discrepancy* associated to  $\omega$  is the value  $\omega(\not\cong)$ , where

$$\not\cong = \bigcap_{E \in \Sigma} \{(x, y) \mid x \in E \text{ iff } y \in E\}$$

is the *inseparability relation* w.r.t. measurable sets in  $\Sigma$ .

The aforementioned duality is stated in the next theorem [33, Th.5.2].

**Lemma 15.5.4.** Let  $\mu, \nu$  be probability measures on  $(X, \Sigma)$ . Then, provided that  $\cong$  is measurable in  $\Sigma \otimes \Sigma$ ,

$$\|\mu - \nu\| = \min\{\omega(\cong) \mid \omega \in \Omega(\mu, \nu)\}.$$

*Proof.* We prove that  $\|\mu - \nu\|$  is a lower bound for  $\{\omega(\cong) \mid \omega \in \Omega(\mu, \nu)\}$ .

Let  $\omega \in \Omega(\mu, \nu)$  and  $E \in \Sigma$ , then

$$\begin{aligned} \mu(E) &= \omega(E \times X) && (\omega \in \Omega(\mu, \nu)) \\ &\geq \omega((X \times E) \cap \cong) && (\text{def. } \cong) \\ &= 1 - \omega((X \times E)^c \cup \cong) && (\text{complement}) \\ &\geq 1 - \omega((X \times E)^c) - \omega(\cong) && (\text{sub additivity}) \\ &= \omega(X \times E) - \omega(\cong) && (\text{complement}) \\ &= \nu(E) - \omega(\cong). && (\omega \in \Omega(\mu, \nu)) \end{aligned}$$

Thus, by the generality of  $\omega \in \Omega(\mu, \nu)$  and  $E \in \Sigma$ , it immediately follows that

$$\|\mu - \nu\| = \sup_{E \in \Sigma} |\mu(E) - \nu(E)| \leq \min\{\omega(\cong) \mid \omega \in \Omega(\mu, \nu)\}.$$

Now we prove that there exists an optimal coupling  $\omega^* \in \Omega(\mu, \nu)$  such that

$$\omega^*(\cong) = \|\mu - \nu\|.$$

Define  $\psi: X \rightarrow X \times X$  by  $\psi(x) = (x, x)$  (it is measurable because  $\psi^{-1}(E \times E') = E \cap E'$ , for all  $E, E' \in \Sigma$ ). Note that  $\psi^{-1}(\cong) = X$ , since  $\psi(x) = (x, x) \in \cong$ .

If  $\mu = \nu$ , define  $\omega^* = \mu[\psi]$  (to check that this is a coupling and that it is such that  $\omega^*(\cong) = \|\mu - \nu\|$  is trivial).

Let  $\mu \neq \nu$ . Define  $\mu \wedge \nu: \Sigma \rightarrow \mathbb{R}_{\geq 0}$  as follows, for  $E \in \Sigma$

$$(\mu \wedge \nu)(E) = \inf\{\mu(F) + \nu(E \setminus F) \mid F \in \Sigma \text{ and } F \subseteq E\}.$$

The above is a well defined measure (a.k.a. the meet of  $\mu$  and  $\nu$ , see [24, Corr.6 pp.163]). Now define the following derived measures

$$\eta = \mu - (\mu \wedge \nu), \quad \eta' = \nu - (\mu \wedge \nu), \quad \omega^* = \frac{\eta \times \eta'}{1 - \gamma} + (\mu \wedge \nu)[\psi].$$

where  $\gamma = (\mu \wedge \nu)[\psi](\cong)$ .

Note that, since  $\psi^{-1}(\cong) = X$ ,  $(\mu \wedge \nu)[\psi]$  puts all its mass in  $\cong$ . Moreover, since  $\mu \neq \nu$ , we get  $\gamma < 1$ , so  $\omega^*$  is well defined and, in particular,  $\omega^*(\cong) = \gamma$ .

Now we show that  $\omega^* \in \Omega(\mu, \nu)$ . Let  $E \in \Sigma$ , then

$$\begin{aligned}
\omega^*(E \times X) &= \frac{\eta(E) \cdot \eta'(X)}{1 - \gamma} + (\mu \wedge \nu)[\psi](E \times \Pi(S)) && \text{(def. } \omega^*) \\
&= \frac{\eta(E) \cdot (\nu(X) - (\mu \wedge \nu)(X))}{1 - \gamma} + (\mu \wedge \nu)[\psi](E \times X) && \text{(def. } \eta') \\
&= \frac{\eta(E) \cdot (1 - \gamma)}{1 - \gamma} + (\mu \wedge \nu)[\psi](E \times X) && \text{(def. } \mu \wedge \nu) \\
&= \mu(E) - (\mu \wedge \nu)(E) + (\mu \wedge \nu)[\psi](E \times X) && \text{(def. } \eta) \\
&= \mu(E) - (\mu \wedge \nu)(E) + (\mu \wedge \nu)(E) && \text{(def. } (\mu \wedge \nu)[\psi]) \\
&= \mu(E).
\end{aligned}$$

Similarly  $\omega^*(X \times E) = \nu(E)$ . The following shows that  $\omega^*$  is optimal

$$\begin{aligned}
\|\mu - \nu\| &= 1 - (\mu \wedge \nu)(X) && \text{(def. } \mu \wedge \nu \text{ and compl.)} \\
&= 1 - (\mu \wedge \nu)[\psi](\cong) && \text{(def. } \psi) \\
&= 1 - \gamma && \text{(def. } \gamma) \\
&= 1 - \omega^*(\cong) && \text{(def. } \omega^*) \\
&= \omega^*(\not\cong) && \text{(compl.)}
\end{aligned}$$

□

Given the above result, we can state a second general converging criterion to approach the total variation distance from above.

**Theorem 15.5.5.** *Let  $(X, \Sigma)$  be a measurable space s.t.  $\cong \in \Sigma \otimes \Sigma$  and  $\mu, \nu$  be probability measures on it. Let*

$$\Omega_0 \subseteq \Omega_1 \subseteq \Omega_2 \dots$$

*be an increasing sequence s.t.  $\bigcup_{i \in \mathcal{N}} \Omega_i$  is dense in  $\Omega(\mu, \nu)$  w.r.t. the total variation distance and define*

$$u_i = \inf\{\omega(\not\cong) \mid \omega \in \Omega_i\}.$$

*Then,  $u_i \geq u_{i+1}$ , for all  $i \in \mathcal{N}$  and*

$$\inf_{i \in \mathcal{N}} u_i = \|\mu - \nu\|.$$

*Proof.* The inequality  $u_i \geq u_{i+1}$  follows from  $\Omega_i \subseteq \Omega_{i+1}$ .

To prove  $\inf_{i \in \mathcal{N}} u_i = \|\mu - \nu\|$ , recall that for  $Y \neq \emptyset$  and  $f: Y \rightarrow \mathbb{R}$  bounded and continuous, if  $D \subseteq Y$  is dense then  $\inf f(D) = \inf f(Y)$ .

By hypothesis  $\bigcup_{i \in \mathcal{N}} \Omega_i \subseteq \Omega(\mu, \nu)$  is dense; moreover,  $\mu \times \nu \in \Omega(\mu, \nu) \neq \emptyset$ .

We show that  $ev_{\not\cong}: \Omega(\mu, \nu) \rightarrow \mathbb{R}$ , defined by  $ev_{\not\cong}(\omega) = \omega(\not\cong)$  is bounded and continuous. It is bounded since all  $\omega \in \Omega(\mu, \nu)$  are finite measures. It is continuous because

$$\|\omega - \omega'\| \geq |\omega(\not\cong) - \omega'(\not\cong)| = |ev_{\not\cong}(\omega) - ev_{\not\cong}(\omega')|$$

hence, 1-Lipschitz continuous.

Further, applying Lemma 15.5.4, we derive our result.  $\square$

To conclude this section, we define a sequence of sets of couplings that, according to Theorem 15.5.5, characterizes the trace distance  $\delta_\lambda$  on SMCs.

Observe that the inseparability relation w.r.t. the  $\sigma$ -algebra generated by trace cylinders is measurable and it can be characterized as follows.

**Lemma 15.5.6.**

$$\equiv_{\ell\omega} = \bigcap_{E \in \sigma(\mathcal{T})} \{(\pi, \pi') \mid \pi \in E \text{ iff } \pi' \in E\} \in \sigma(\mathcal{T}) \otimes \sigma(\mathcal{T}).$$

*Proof.* We firstly show that

$$\equiv_{\ell\omega} = \bigcap_{E \in \sigma(\mathcal{T})} \{(\pi, \pi') \mid \pi \in E \text{ iff } \pi' \in E\}.$$

( $\subseteq$ ) It suffices to prove inseparability w.r.t trace cylinders.

Let  $\pi \equiv_{\ell\omega} \pi'$  and  $\pi \in C = \mathfrak{C}(C_0, R_0, \dots, R_{n-1}, C_n) \in \mathcal{T}$ , for  $C_i \in S/\equiv_\ell$  and  $R_i \in \mathbb{B}$ ,  $i = 0..n$ . Then, for all  $j \in \mathcal{N}$ ,  $\ell(\pi[j]) = \ell(\pi'[j])$  (hence,  $\pi[j] \equiv_\ell \pi'[j]$ ) and  $\pi\langle j \rangle = \pi'\langle j \rangle$ , so that  $\pi' \in C$ .

( $\supseteq$ ) By contraposition. Let  $\pi \not\equiv_{\ell\omega} \pi'$ , then there exist  $j \in \mathcal{N}$  such that  $\pi[j] \not\equiv_\ell \pi'[j]$  or  $\pi\langle j \rangle \neq \pi'\langle j \rangle$ . Let  $E = (\cdot)|_j^{-1}(\mathfrak{C}([\pi[j]]_{\equiv_\ell}, \{\pi\langle j \rangle\} | \cdot), S)$ , then  $\pi \in E$  but  $\pi' \notin E$ . The inclusion follows since the function  $(\cdot)|_j$  is measurable.

As for the measurability of  $\equiv_{\ell\omega}$ , it suffices to show that its complement  $\not\equiv_{\ell\omega} \in \sigma(\mathcal{T}) \otimes \sigma(\mathcal{T})$ . Define  $DiffS(k)$  and  $DiffT(k)$ , for  $k \geq 0$ , as

$$DiffS(k) := \bigcup_{C \in S/\equiv_\ell} (\cdot)|_k^{-1}(\mathfrak{C}(C)) \times (\cdot)|_k^{-1}(\mathfrak{C}(S \setminus C)),$$

$$DiffT(k) := \bigcup_{t < t' \in \mathbb{Q}_+} (\cdot)|_k^{-1}(\mathfrak{C}(S, (t, t'), S)) \times (\cdot)|_k^{-1}(\mathfrak{C}(S, \mathbb{R}_{\geq 0} \setminus (t, t'), S)).$$

Measurability of  $DiffS(k)$  and  $DiffT(k)$  follows immediately from the measurability of  $(\cdot)|_k$ .

Now we show that  $\not\equiv_{\ell\omega} = \bigcup_{k \in \mathcal{N}} (DiffS(k) \cup DiffT(k))$ .

( $\subseteq$ ) Let  $\pi \not\equiv_{\ell\omega} \pi'$ . Then,  $\pi[j] \not\equiv_\ell \pi'[j]$  or  $\pi\langle j \rangle \neq \pi'\langle j \rangle$ , for some  $j \in \mathcal{N}$ .

Assume  $\pi[j] \not\equiv_\ell \pi'[j]$ , then  $(\pi, \pi') \in (\cdot)|_j^{-1}(\mathfrak{C}(C)) \times (\cdot)|_j^{-1}(\mathfrak{C}(S \setminus C))$ .

Assume  $\pi\langle j \rangle \neq \pi'\langle j \rangle$ . Let  $\varepsilon = |\pi\langle j \rangle - \pi'\langle j \rangle|$ . Since  $\mathbb{Q}_+$  is dense in  $\mathbb{R}_{\geq 0}$ , every nonempty open set has nonempty intersection with  $\mathbb{Q}_+$ , so that there exist  $t \in \mathbb{Q}_+ \cap (\pi\langle j \rangle - \varepsilon, \pi\langle j \rangle)$  and  $t' \in \mathbb{Q}_+ \cap (\pi\langle j \rangle, \pi\langle j \rangle + \varepsilon)$ . Clearly,  $\pi\langle j \rangle \in (t, t')$  and  $\pi'\langle j \rangle \notin (t, t')$ , therefore,

$$(\pi, \pi') \in (\cdot)|_j^{-1}(\mathfrak{C}(S, (t, t'), S)) \times (\cdot)|_j^{-1}(\mathfrak{C}(S, \mathbb{R}_{\geq 0} \setminus (t, t'), S)).$$

( $\supseteq$ ) Let  $(\pi, \pi') \in \text{DiffS}(k) \cup \text{DiffT}(k)$ , for some  $k \in \mathcal{N}$ . Then, since

$$\begin{aligned} \text{DiffS}(k) &= \bigcup_{C \in \mathcal{S}/\equiv_\ell} \{(\pi, \pi') \mid \pi|_k \in \mathfrak{C}(C) \text{ and } \pi'|_k \in \mathfrak{C}(S \setminus C)\} \\ &= \{(\pi, \pi') \mid \pi[k] \not\equiv_\ell \pi'[k]\}, \end{aligned}$$

$$\begin{aligned} \text{DiffT}(k) &= \bigcup_{t < t' \in \mathbb{Q}_+} \{(\pi, \pi') \mid \pi|_k \in \mathfrak{C}(S, (t, t'), S), \pi'|_k \in \mathfrak{C}(S, \mathbb{R}_{\geq 0} \setminus (t, t'), S)\} \\ &= \bigcup_{t < t' \in \mathbb{Q}_+} \{(\pi, \pi') \mid \pi\langle k \rangle \in (t, t') \text{ and } \pi'\langle k \rangle \notin (t, t')\} \\ &\subseteq \{(\pi, \pi') \mid \pi\langle k \rangle \neq \pi'\langle k \rangle\}, \end{aligned}$$

there exists  $k \in \mathcal{N}$  such that  $\pi[k] \not\equiv_\ell \pi'[k]$  or  $\pi\langle k \rangle \neq \pi'\langle k \rangle$ . Thus,  $\pi \not\equiv_{\ell\omega} \pi'$ .  $\square$

Next we introduce the notion of *coupling structure* for an SMC. Let

$$\Pi^k(S) = \{s_0, t_0, \dots, t_{k-1}, s_k \mid s_i \in S, t_i \in \mathbb{R}_{\geq 0}\}$$

be the measurable space with  $\sigma$ -algebra generated by

$$\mathcal{R}_k = \{\{s_0\} \times R_0 \times \dots \times R_{k-1} \times \{s_k\} \mid s_i \in S, R_i \in \mathbb{B}\}.$$

Note that, the prefix function  $(\cdot)|^k: \Pi(S) \rightarrow \Pi^k(S)$  is measurable, hence, the push forward w.r.t. it on  $\mu \in \Delta(\Pi(S))$ , denoted by  $\mu|^k$ , is a measure in  $\Pi^k(S)$ .

**Definition 15.5.7** (Coupling Structure). *A coupling structure of rank  $k \in \mathcal{N}$  for an SMC  $\mathcal{M}$  is a function  $\mathcal{C}: S \times S \rightarrow \Delta(\Pi^k(S) \times \Pi^k(S))$  such that, for all states  $s, s' \in S$ ,*

$$\mathcal{C}(s, s') \in \Omega(\mathbb{P}_s|^k, \mathbb{P}_{s'}|^k).$$

The set of coupling structures of rank  $k$  for  $\mathcal{M}$  is denoted by  $\mathbf{C}_k(\mathcal{M})$ .

A coupling structure of rank  $k$  together with a distinguished initial pair of states, can be intuitively seen as a stochastic process generating pairs of timed paths divided in multi-steps of length  $k$  and distributed according to the following probability.

**Definition 15.5.8.** *For  $k \in \mathcal{N}$ ,  $s, s' \in S$  states in  $\mathcal{M}$  and  $\mathcal{C} \in \mathbf{C}_k(\mathcal{M})$ , let  $\mathbb{P}_{s, s'}^{\mathcal{C}}$  be the unique probability measure<sup>5</sup> on  $\Pi(S) \otimes \Pi(S)$  such that, for all  $n \in \mathcal{N}$  and*

$$E = \{u_0\} \times R_0 \times \dots \times R_{nk-1} \times \{u_{nk}\}, F = \{v_0\} \times H_0 \times \dots \times H_{nk-1} \times \{v_{nk}\} \in \mathcal{R}_{nk},$$

$$\mathbb{P}_{s, s'}^{\mathcal{C}}(\mathfrak{C}(E) \times \mathfrak{C}(F)) = \mathbf{1}_{\{(s, s')\}}(u_0, v_0) \cdot \prod_{h=0}^{n-1} \mathcal{C}(u_{hk}, v_{hk})(E_h \times F_h),$$

where  $\mathfrak{C}(E)$  denotes the cylinder obtained as the pre-image under  $(\cdot)|^{nk}$  of  $E$  and  $E_h = \{u_{hk}\} \times R_{hk} \times \dots \times R_{(h+1)k-1} \times \{u_{(h+1)k}\}$  (similarly for  $F$ ).

<sup>5</sup>The existence and the uniqueness of this measure follow from Hahn-Kolmogorov extension theorem and the fact that any cylinder of rank  $k$  can always be represented as a disjoint union of cylinders of rank  $k' \geq k$  (see e.g., [8, pp.29–32]).

The name “coupling structure” is justified by the following result.

**Lemma 15.5.9.** *Let  $\mathcal{C}$  be a coupling structure for  $\mathcal{M}$ , then  $\mathbb{P}_{s,s'}^{\mathcal{C}} \in \Omega(\mathbb{P}_s, \mathbb{P}_{s'})$ .*

*Proof.* Let  $\mathcal{C} \in \mathbf{C}_k(\mathcal{M})$ . To prove that  $\mathbb{P}_{s,s'}^{\mathcal{C}} \in \Omega(\mathbb{P}_s, \mathbb{P}_{s'})$  it suffices to show that, for all  $n \in \mathcal{N}$  and  $E = \{u_0\} \times R_0 \times \dots \times R_{nk-1} \times \{u_{nk}\} \in \mathcal{R}_{nk}$

$$\mathbb{P}_{s,s'}^{\mathcal{C}}(\mathfrak{C}(E) \times \Pi(S)) \stackrel{(i)}{=} \mathbb{P}_s(\mathfrak{C}(E)), \quad \mathbb{P}_{s,s'}^{\mathcal{C}}(\Pi(S) \times \mathfrak{C}(E)) \stackrel{(ii)}{=} \mathbb{P}_{s'}(\mathfrak{C}(E)).$$

We prove (i) by induction on  $n \geq 0$ .

The base case is trivial.

Let  $n > 0$ . For any  $\vec{v} \in S^{nk+1}$  define  $F^{\vec{v}} = \{v_0\} \times \mathbb{R}_{\geq 0} \times \dots \times \mathbb{R}_{\geq 0} \times \{v_{nk}\}$  and, for  $h < n$ , let  $F_h^{\vec{v}} = \{v_{hk}\} \times \mathbb{R}_{\geq 0} \times \dots \times \mathbb{R}_{\geq 0} \times \{v_{(h+1)k}\}$ . Then the following hold

$$\begin{aligned} & \mathbb{P}_{s,s'}^{\mathcal{C}}(\mathfrak{C}(E) \times \Pi(S)) = \\ &= \sum_{\vec{v} \in S^{nk+1}} \mathbb{P}_{s,s'}^{\mathcal{C}}(\mathfrak{C}(E) \times \mathfrak{C}(F^{\vec{v}})) && \text{(additivity)} \\ &= \sum_{\vec{v} \in S^{nk+1}} \mathbb{1}_{\{(s,s')\}}(u_0, v_0) \cdot \prod_{h=0}^{n-1} \mathcal{C}(u_{hk}, v_{hk})(E_h \times F_h^{\vec{v}}) && \text{(def. } \mathbb{P}_{s,s'}^{\mathcal{C}}) \\ &= \sum_{\vec{v} \in S^{(n-1)k+1}} \mathbb{P}_{s,s'}^{\mathcal{C}}(\mathfrak{C}(E') \times \mathfrak{C}(F^{\vec{v}})) \cdot \mathcal{C}(s_{(n-1)k}, v_{(n-1)k})(E_{(n-1)} \times \Pi^k(S)) && \text{(def. } \mathbb{P}_{s,s'}^{\mathcal{C}}) \\ &= \sum_{\vec{v} \in S^{(n-1)k+1}} \mathbb{P}_{s,s'}^{\mathcal{C}}(\mathfrak{C}(E') \times \mathfrak{C}(F^{\vec{v}})) \cdot \mathbb{P}_{s_{(n-1)k}}(E_{(n-1)}) && (\mathcal{C} \in \mathbf{C}_k(\mathcal{M})) \\ &= \mathbb{P}_{s,s'}^{\mathcal{C}}(\mathfrak{C}(E') \times \Pi(S)) \cdot \mathbb{P}_{s_{(n-1)k}}(E_{(n-1)}) && \text{(additivity)} \\ &= \mathbb{P}_s(\mathfrak{C}(E')) \cdot \mathbb{P}_{s_{(n-1)k}}(E_{(n-1)}) && \text{(inductive hp.)} \\ &= \mathbb{P}_s(\mathfrak{C}(E)) && \text{(def. } \mathbb{P}_{s,s'}^{\mathcal{C}}) \end{aligned}$$

where  $E' = \{u_0\} \times R_0 \times \dots \times R_{(n-1)k-1} \times \{u_{(n-1)k}\}$ .

(ii) follows similarly.  $\square$

We are finally ready to describe a decreasing sequence that converges to the trace distance on SMCs.

Given  $\mathcal{M}$ , let  $\delta_{\lambda \downarrow k}: S \times S \rightarrow [0, 1]$  for  $k \in \mathcal{N}$ , be

$$\delta_{\lambda \downarrow k}(s, s') = \min\{\mathbb{P}_{s,s'}^{\mathcal{C}}(\neq_{\ell^\omega}) \mid \mathcal{C} \in \mathbf{C}_{2k}(\mathcal{M})\}. \quad (15.5.2)$$

According to Theorem 15.5.1, the following suffices to prove the convergence.

**Lemma 15.5.10.** *Let  $s, s' \in S$  be a pair of states of an SMC  $\mathcal{M}$ . Then,*

- (i) *for all  $k \in \mathcal{N}$ ,  $\{\mathbb{P}_{s,s'}^{\mathcal{C}} \mid \mathcal{C} \in \mathbf{C}_k(\mathcal{M})\} \subseteq \{\mathbb{P}_{s,s'}^{\mathcal{C}} \mid \mathcal{C} \in \mathbf{C}_{2k}(\mathcal{M})\}$ ;*
- (ii)  *$\bigcup_{k \in \mathcal{N}} \{\mathbb{P}_{s,s'}^{\mathcal{C}} \mid \mathcal{C} \in \mathbf{C}_{2k}(\mathcal{M})\}$  is dense in  $\Omega(\mathbb{P}_s, \mathbb{P}_{s'})$  w.r.t. the total variation.*

*Proof.* We prove the two items separately.

(i) Let  $k > 0$  and  $\mathcal{C} \in \mathbf{C}_k(\mathcal{M})$ .

Define, for all  $s, s' \in S$ ,  $\mathcal{D}(s, s')$  as the unique measure on  $\Pi^{2k}(S) \otimes \Pi^{2k}(S)$  s.t., for all

$$E = \{u_0\} \times R_0 \times \dots \times R_{2k-1} \times \{u_{2k}\}$$

and

$$F = \{v_0\} \times H_0 \times \dots \times H_{2k-1} \times \{v_{2k}\}$$

in  $\mathcal{R}_{2k}$

$$\mathcal{D}(s, s')(E \times F) = \mathcal{C}(s, s')(E' \times F') \cdot \mathcal{C}(u_k, v_k)(E'' \times F''),$$

where  $E' = \{u_0\} \times R_0 \times \dots \times R_{k-1} \times \{u_k\}$  and  $E'' = \{u_k\} \times R_k \times \dots \times R_{2k-1} \times \{u_{2k}\}$  (similarly for  $F$ ).

To show  $\mathcal{D} \in \mathbf{C}_{2k}(\mathcal{M})$  we need to prove that for all  $s, s' \in S$ ,  $\mathcal{D}(s, s') \in \Omega(\mathbb{P}_s|^{2k}, \mathbb{P}_{s'}|^{2k})$ . To this end it is sufficient that, for all measurable sets

$$E = \{u_0\} \times R_0 \times \dots \times R_{2k-1} \times \{u_{2k}\} \in \mathcal{R}_{2k},$$

the following hold

$$\mathcal{D}(s, s')(E \times \Pi^{2k}(S)) \stackrel{(*)}{=} \mathbb{P}_s|^{2k}(E), \quad \mathcal{D}(s, s')(\Pi^{2k}(S) \times E) \stackrel{(**)}{=} \mathbb{P}_{s'}|^{2k}(E).$$

We prove only (\*). For any  $\vec{v} \in S^{2k+1}$  define  $F^{\vec{v}} = \{v_0\} \times \mathbb{R}_{\geq 0} \times \dots \times \mathbb{R}_{\geq 0} \times \{v_{2k}\}$  and, for  $h = 0..1$ , let  $F_h^{\vec{v}} = \{v_{hk}\} \times \mathbb{R}_{\geq 0} \times \dots \times \mathbb{R}_{\geq 0} \times \{v_{(h+1)k}\}$ . Then we have

$$\begin{aligned} \mathcal{D}(s, s')(E \times \Pi^{2k}(S)) &= \\ &= \sum_{\vec{v} \in S^{2k+1}} \mathcal{D}(s, s')(E \times F^{\vec{v}}) && \text{(additivity)} \\ &= \sum_{\vec{v} \in S^{2k+1}} \mathcal{C}(s, s')(E' \times F_0^{\vec{v}}) \cdot \mathcal{C}(u_k, v_k)(E'' \times F_1^{\vec{v}}) && \text{(def. } \mathcal{D}) \\ &= \sum_{\vec{v} \in S^{k+1}} \mathcal{C}(s, s')(E' \times F_0^{\vec{v}}) \cdot \mathcal{C}(u_k, v_k)(E'' \times \Pi^k(S)) && \text{(additivity)} \\ &= \sum_{\vec{v} \in S^{k+1}} \mathcal{C}(s, s')(E' \times F_0^{\vec{v}}) \cdot \mathbb{P}_{u_k}|^k(E'') && (\mathcal{C} \in \mathbf{C}_k(\mathcal{M})) \\ &= \mathcal{C}(s, s')(E' \times \Pi^k(S)) \cdot \mathbb{P}_{u_k}|^k(E'') && \text{(additivity)} \\ &= \mathbb{P}_s|^k(E') \cdot \mathbb{P}_{u_k}|^k(E'') && (\mathcal{C} \in \mathbf{C}_k(\mathcal{M})) \\ &= \mathbb{P}_s(\mathfrak{C}(E')) \cdot \mathbb{P}_{u_k}(\mathfrak{C}(E'')) && \text{(preimage)} \\ &= \mathbb{P}_s(\mathfrak{C}(E)) && \text{(def. } \mathbb{P}_s) \\ &= \mathbb{P}_s|^{2k}(E). && \text{(preimage)} \end{aligned}$$

We show that, for arbitrary  $s, s' \in S$ ,  $\mathbb{P}_{s, s'}^{\mathcal{C}} = \mathbb{P}_{s, s'}^{\mathcal{D}}$ . To this end it suffices to check the following for all  $n \in \mathcal{N}$  and

$$\begin{aligned} E &= \{u_0\} \times R_0 \times \dots \times R_{2nk-1} \times \{u_{2nk}\}, \\ F &= \{v_0\} \times H_0 \times \dots \times H_{2nk-1} \times \{v_{2nk}\} \end{aligned}$$

in  $\mathcal{R}_{2nk}$ :

$$\mathbb{P}_{s, s'}^{\mathcal{C}}(\mathfrak{C}(E) \times \mathfrak{C}(F)) = \mathbb{P}_{s, s'}^{\mathcal{D}}(\mathfrak{C}(E) \times \mathfrak{C}(F))$$

We proceed by induction on  $n \geq 0$ . The base case is trivial.

Assume  $n > 0$  and, for  $i \in \{k, 2k\}$ , define  $E_h^i = \{u_{hi}\} \times R_{hi} \times \dots \times R_{(h+1)i-1} \times \{u_{(h+1)i}\}$  (similarly for  $F$ ). Then the following holds:

$$\begin{aligned} \mathbb{P}_{s, s'}^{\mathcal{C}}(\mathfrak{C}(E) \times \mathfrak{C}(F)) &= \mathbb{1}_{\{(s, s')\}}(u_0, v_0) \cdot \prod_{h=0}^{2n-1} \mathcal{C}(u_{hk}, v_{hk})(E_h^k \times F_h^k) && \text{(def. } \mathbb{P}^{\mathcal{C}}) \\ &= \mathbb{1}_{\{(s, s')\}}(u_0, v_0) \cdot \prod_{h=0}^{n-1} \mathcal{D}(u_{2hk}, v_{2hk})(E_h^{2k} \times F_h^{2k}) && \text{(def. } \mathcal{D}) \\ &= \mathbb{P}_{s, s'}^{\mathcal{D}}(\mathfrak{C}(E) \times \mathfrak{C}(F)). && \text{(def. } \mathbb{P}^{\mathcal{D}}) \end{aligned}$$

From the above it immediately follows that  $\mathbf{C}_k(\mathcal{M}) \subseteq \mathbf{C}_{2k}(\mathcal{M})$ .

(ii) We prove the following more general result from which we will obtain (ii).

Let  $(X, \Sigma)$  be a measurable space such that  $\mathcal{F}$  is a field that generates  $\Sigma$  and let  $D \subseteq \Delta(X)$  be such that, for all  $\mu \in \Delta(X)$  and  $F \in \mathcal{F}$ , there exists  $\nu \in D$  such that  $\nu(F) = \mu(F)$ . Then  $D$  is dense in  $\Delta(X)$  w.r.t. the total variation distance.

Let  $E \in \Sigma$  be an arbitrary measurable set and  $d_E: \Delta(X) \times \Delta(X) \rightarrow \mathbb{R}_{\geq 0}$  be the pseudometric defined, for  $\mu, \nu \in \Delta(X)$ , by

$$d_E(\mu, \nu) = |\mu(E) - \nu(E)|.$$

Since  $\|\mu - \nu\| = \sup_{E \in \Sigma} d_E(\mu, \nu)$ , to prove that  $D$  is dense w.r.t. the total variation distance it suffices to show that  $D$  is dense w.r.t.  $d_E$ , for any  $E \in \Sigma$ .

Let  $E \in \Sigma$  and  $\varepsilon > 0$ . For any  $\mu \in \Delta(X)$  we have to provide  $\nu \in D$  such that  $d_E(\mu, \nu) < \varepsilon$ .

Define the measure  $\tilde{\mu}$  as the least upper bound of  $D \cup \{\mu\}$  w.r.t. the point-wise partial order between measures:

$$\nu \sqsubseteq \nu' \text{ iff } \nu(A) \leq \nu'(A), \text{ for all } A \in \Sigma.$$

The existence of  $\tilde{\mu}$  is guaranteed by [24, Corr.6 pp.163] – note that  $\tilde{\mu}$  is not necessarily finite.

Applying Lemma 15.4.1,  $\mathcal{F} \subseteq \Sigma$  is dense in  $(\Sigma, d_{\tilde{\mu}})$ , where  $d_{\tilde{\mu}}$  is the Fréchet-Nikodym pseudometric<sup>6</sup>, hence there exists  $F \in \mathcal{F}$  such that  $d_{\tilde{\mu}}(E, F) < \frac{\varepsilon}{2}$ .

<sup>6</sup>Notice that Lemma 15.4.1 does not assume the measure to be finite, hence it can be applied to  $\tilde{\mu}$ .

By hypothesis, there exists  $\nu \in D$ , such that  $\nu(F) = \mu(F)$ . Let  $\omega \in \{\mu, \nu\}$  then

$$\begin{aligned}
\omega(E) &= \omega(E \setminus F) + \omega(E \cap F) && (\omega \text{ additive}) \\
&\leq \omega((E \setminus F) \cup (F \setminus E)) + \omega(F) && (\omega \text{ monotone}) \\
&= \omega(E \triangle F) + \omega(E) && (\text{by def}) \\
&\leq \tilde{\mu}(E \triangle F) + \omega(F) && (\omega \sqsubseteq \tilde{\mu}) \\
&= d_{\tilde{\mu}}(E, F) + \omega(F). && (\text{by def})
\end{aligned}$$

This implies  $|\omega(E) - \omega(F)| \leq d_{\tilde{\mu}}(E, F)$ , and in particular that  $|\mu(E) - \mu(F)| < \frac{\varepsilon}{2}$  and  $|\nu(E) - \nu(F)| < \frac{\varepsilon}{2}$ . Then, the density of  $D$  follows by

$$\begin{aligned}
d_E(\mu, \nu) &= |\mu(E) - \nu(E)| && (\text{def. } d_E) \\
&\leq |\mu(E) - \mu(F)| + |\mu(F) - \nu(E)| && (\text{triangular ineq.}) \\
&= |\mu(E) - \mu(F)| + |\nu(F) - \nu(E)| && (\nu(F) = \mu(F)) \\
&< \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon.
\end{aligned}$$

Let  $s, s' \in S$ ,  $\Omega = \bigcup_{i \in \mathcal{N}} \{\mathbb{P}_{s, s'}^{\mathcal{C}} \mid \mathcal{C} \in \mathbf{C}_{2^i}(\mathcal{M})\}$ .

Given the general result above, to prove (ii) it is sufficient to provide a field  $\mathcal{F}$  that generates the  $\sigma$ -algebra of  $\Pi(S) \otimes \Pi(S)$  and to show that, for every  $\mu \in \Omega(\mathbb{P}_s, \mathbb{P}_{s'})$  and  $F \in \mathcal{F}$ , there exists  $\omega \in \Omega$  such that  $\omega(F) = \mu(F)$ .

Define  $\mathcal{F} = \bigcup_{k \in \mathcal{N}} \mathcal{F}_k$ , where  $\mathcal{F}_k$  denotes the collection of all finite union of measurable sets of the form  $\mathcal{C}(E) \times \mathcal{C}(F)$ , for some  $E, F \in \mathcal{R}_k$ . It holds that  $\mathcal{F}_k \subseteq \mathcal{F}_{k+1}$  and  $\mathcal{F}_k$  is a field, for all  $k \in \mathcal{N}$ . Therefore  $\mathcal{F}$  is a field that generates the  $\sigma$ -algebra of  $\Pi(S) \otimes \Pi(S)$ .

Let  $\mu \in \Omega(\mathbb{P}_s, \mathbb{P}_{s'})$ ,  $k \in \mathcal{N}$  and  $\mathcal{D} \in \mathbf{C}_k(\mathcal{M})$ . We define  $\omega_k = \mathbb{P}_{s, s'}^{\mathcal{C}_k}$ , for  $\mathcal{C}_k: S \times S \rightarrow \Delta(\Pi^k(S) \times \Pi^k(S))$  defined by

$$\mathcal{C}_k(u, v) = \begin{cases} \mu[(\cdot)^k \times (\cdot)^k] & \text{if } (u, v) = (s, s') \\ \mathcal{D}(u, v) & \text{otherwise} \end{cases}$$

where  $\mu[(\cdot)^k \times (\cdot)^k]$  denotes the push forward of  $\mu$  along  $(\pi, \pi') \mapsto (\pi^k, \pi'^k)$ .

Note that, since  $\mathbf{C}_k(\mathcal{M})$  is nonempty,  $\mathcal{C}_k$  is well defined.

We show  $\mathcal{C}_k \in \mathbf{C}_k(\mathcal{M})$ . We just need to prove  $\mu[(\cdot)^k \times (\cdot)^k] \in \Omega(\mathbb{P}_s^k, \mathbb{P}_{s'}^k)$  that corresponds to check

$$\mu[(\cdot)^k \times (\cdot)^k](E \times \Pi^k(S)) = \mathbb{P}_s^k(E)$$

and

$$\mu[(\cdot)^k \times (\cdot)^k](\Pi^k(S) \times E) = \mathbb{P}_{s'}^k(E),$$

for arbitrary  $E \in \mathcal{R}_k$  (we check one equality, the other follows similarly):

$$\begin{aligned} \mu[(\cdot)^k \times (\cdot)^k](E \times \Pi^k(S)) &= \mu(\mathfrak{C}(E) \times \Pi(S)) && \text{(preimage)} \\ &= \mathbb{P}_s(\mathfrak{C}(E)) && (\mu \in \Omega(\mathbb{P}_s, \mathbb{P}_{s'})) \\ &= \mathbb{P}_s|^k(E). && \text{(preimage)} \end{aligned}$$

Next we prove that for all  $A \in \mathcal{F}_k$ ,  $\omega_k(A) = \mu(A)$ . Note that since  $\mathcal{F}_k \subseteq \mathcal{F}_{k+1}$ , this suffices to show that  $\omega_k(B) = \mu(B)$  holds for all  $B \in \mathcal{F}_j$  such that  $j \leq k$ .

Let  $A = \bigcup_{i=0}^n \mathfrak{C}(E_i) \times \mathfrak{C}(F_i) \in \mathcal{F}_k$ , for some  $n \in \mathcal{N}$  and  $E_i, F_i \in \mathcal{R}_k$  ( $i = 0..n$ ). Without loss of generality we can assume that the  $\mathfrak{C}(E_i) \times \mathfrak{C}(F_i)$ 's forming  $A$  are pairwise disjoint (indeed,  $\mathcal{F}_k$  is a field, hence we can simply replace any two "overlapping" sets by taking the intersection and their symmetric difference).

$$\begin{aligned} \omega_k(A) &= \mathbb{P}_{s,s'}^{\mathcal{C}_k}(A) && \text{(def. } \omega_k) \\ &= \sum_{i=0}^n \mathbb{P}_{s,s'}^{\mathcal{C}_k}(\mathfrak{C}(E_i) \times \mathfrak{C}(F_i)) && \text{(additivity)} \\ &= \sum_{i=0}^n \mathcal{C}_k(s, s')(E_i \times F_i) && \text{(def. } \mathbb{P}_{s,s'}^{\mathcal{C}_k}) \\ &= \sum_{i=0}^n \mu(\mathfrak{C}(E_i) \times \mathfrak{C}(F_i)) && \text{(def. } \mathcal{C}_k) \\ &= \mu(A). && \text{(additivity)} \end{aligned}$$

To conclude the proof, observe that, given  $\mu \in \Omega(\mathbb{P}_s, \mathbb{P}_{s'})$  and  $F \in \mathcal{F}$ , there exists  $i \in \mathcal{N}$  such that  $F \in \mathcal{F}_i$ , and that for  $\omega_{2^i}$  defined as above (w.r.t.  $\mu$ ) is such that  $\omega_{2^i}(F) = \mu(F)$  and  $\omega_{2^i} \in \Omega$ .

□

The following corollary derives from Lemma 15.5.10 and Theorem 15.5.5.

**Corollary 15.5.11.** *For all  $k \in \mathcal{N}$ ,  $\delta_{\lambda \downarrow k} \geq \delta_{\lambda \downarrow k+1}$  and  $\delta_{\lambda} = \inf_{k \in \mathcal{N}} \delta_{\lambda \downarrow k}$ .*

## 15.6 An Approximation Algorithm

This section exploits the aforementioned results to propose a decidable procedure for approximating the trace distance  $\delta_{\lambda}$  on SMCs with arbitrary precision.

Let  $\varepsilon > 0$  and consider the sequences  $\{\delta_{\lambda \uparrow k} \mid k \in \mathcal{N}\}$  and  $\{\delta_{\lambda \downarrow k} \mid k \in \mathcal{N}\}$  from Section 15.5. The procedure proceeds step-wise (by increasing  $k \geq 0$ ) and computing the difference  $\delta_{\lambda \downarrow k} - \delta_{\lambda \uparrow k}$  until is smaller then  $\varepsilon$ . Termination and correctness is ensured by the convergence of the sequences from above and below to the total variation.

**Theorem 15.6.1.** *Let  $\mathcal{M}$  be a SMC. There exists an algorithm that, given  $\varepsilon > 0$ , computes a function  $d: S \times S \rightarrow [0, 1]$  such that  $|d + \delta_{\lambda}| < \varepsilon$ .*

We prove this theorem under two reasonable assumptions regarding SMCs:

**A1.** For all  $s, s' \in S$  and  $q, q' \in \mathbb{Q}_+$ ,  $\rho(s)([q, q'])$  is computable;

**A2.** For all  $s, s' \in S$ ,  $\|\rho(s) - \rho(s')\|$  is computable.

**Lemma 15.6.2.** *Assuming A1,  $\delta_\lambda \uparrow_k$  is computable for all  $k \in \mathcal{N}$ .*

*Proof.* For each  $k \in \mathcal{N}$ , the set  $\mathcal{E}_k$  is finite. Moreover, for each  $s \in S$  and  $E \in \mathcal{E}_k$ ,  $\mathbb{P}_s(E)$  is computable thanks to its additivity and the hypothesis A1.  $\square$

The computability of the sequence  $\{\delta_\lambda \downarrow_k\}_{k \in \mathcal{N}}$  is less trivial. Equation (15.5.2) suggests to look for a coupling structure  $\mathcal{C} \in \mathbf{C}_{2k}(\mathcal{M})$  that minimizes the discrepancy  $\mathbb{P}_{s, s'}^{\mathcal{C}}(\neq_{\ell\omega})$ . This is done by following a searching strategy similar to the one in [5] and detailed in Chapter 11. This is structured as follows:

1. we provide an alternative characterization of the discrepancy associated with a coupling structure (Section 15.6.1);
2. we describe how to construct an optimal coupling structure and show that its associated discrepancy is computable (Section 15.6.2).

### 15.6.1 Fixed Point Characterization of the Discrepancy

We characterize the discrepancy associated with a coupling structure  $\mathcal{C}$  by means of the least fixed point of a suitable operator parametric in  $\mathcal{C}$ .

To define the fixed point operator it is convenient to split a coupling structure into two “projections”: one on discrete state transitions (regardless of time delays); and one on residence times (given that a sequence of transitions has occurred). In order to provide these, consider

$$\mathbf{S}^k: S \rightarrow \mathcal{D}(S^{k+1}) \quad \text{and} \quad \mathbf{T}^k: S^k \rightarrow \Delta(\mathbb{R}_{\geq 0}^k)$$

defined by

$$\mathbf{S}^k(s)(u_0..u_k) = \mathbf{1}_s(u_0) \cdot \prod_{i=0}^{k-1} \tau(u_i)(u_{i+1}), \quad \mathbf{T}^k(v_1..v_k) = \rho(v_1) \times \cdots \times \rho(v_k).$$

**Lemma 15.6.3.** *The set  $\mathbf{C}_k(\mathcal{M})$  is in bijection with the set of pairs of functions  $\tau_{\mathcal{C}}: S \times S \rightarrow \mathcal{D}(S^{k+1} \times S^{k+1})$  and  $\rho_{\mathcal{C}}: S^k \times S^k \rightarrow \Delta(\mathbb{R}_{\geq 0}^k \times \mathbb{R}_{\geq 0}^k)$  such that*

$$\tau_{\mathcal{C}}(u, v) \in \Omega(\mathbf{S}^k(u), \mathbf{S}^k(v)) \quad \text{and} \quad \rho_{\mathcal{C}}(u_1..u_k, v_1..v_k) \in \Omega(\mathbf{T}^k(u_1..u_k), \mathbf{T}^k(v_1..v_k)).$$

*Proof.* Consider the functions  $p_1$  and  $p_2$  defined as follows

$$\begin{aligned} p_1: \Pi^k(S) &\rightarrow S^{k+1} & p_2: \Pi^k(S) &\rightarrow \mathbb{R}_{\geq 0}^k \\ p_1(s_0, t_0, \dots, t_{k-1}, s_k) &= (s_0, \dots, s_k) & p_2(s_0, t_0, \dots, t_{k-1}, s_k) &= (t_0, \dots, t_{k-1}). \end{aligned}$$

These are easily seen to be measurable. For  $\mathcal{C} \in \mathbb{C}_k(\mathcal{M})$  and  $(v, \eta)$  satisfying the conditions of the statement, the bijection is given by  $\mathcal{C} \mapsto (\tau_{\mathcal{C}}, \rho_{\mathcal{C}})$  and  $(v, \eta) \mapsto \mathcal{D}$ , where

$$\begin{aligned} \tau_{\mathcal{C}}(s, s') &= \mathcal{C}(s, s')[p_1 \times p_1], \quad \rho_{\mathcal{C}}(\vec{s}, \vec{s}') = \mathcal{C}(s_0, s'_0)(\cdot | (p_1 \times p_1)^{-1}(\vec{s}, \vec{s}'))[p_2 \times p_2] \\ \mathcal{D}(s, s')(\vec{E}_{0,k} \times \vec{E}'_{0,k}) &= v(s_0, s'_0)(\vec{s}, \vec{s}') \cdot \eta(|\vec{s}|^{k-1}, |\vec{s}'|^{k-1})(\vec{R}, \vec{R}'). \end{aligned}$$

where  $\vec{E}_{0,k}$  and  $\vec{E}'_{0,k}$  are as in Definition 15.5.8.  $\square$

Hereafter we identify the coupling structure  $\mathcal{C}$  with its bijective image  $(\tau_{\mathcal{C}}, \rho_{\mathcal{C}})$ .

Intuitively,  $\tau_{\mathcal{C}}(u, v)(u_0..u_k, v_0..v_k)$  is the probability that two copies of  $\mathcal{M}$ , scheduled according to  $\mathcal{C}$ , have respectively generated the sequences of states  $u_0..u_k$  and  $v_0..v_k$  starting from  $u$  and  $v$ ; while  $\rho(u_0..u_{k-1}, v_0..v_{k-1})(R \times R')$  is the probability that, having observed  $u_0..u_{k-1}$  and  $v_0..v_{k-1}$ , the generated sequence of time delays are in  $R, R' \subseteq \mathbb{R}_{\geq 0}^k$ , respectively.

For a coupling structure  $\mathcal{C} = (\tau_{\mathcal{C}}, \rho_{\mathcal{C}}) \in \mathbb{C}_k(\mathcal{M})$ , we define the self-map  $\Gamma^{\mathcal{C}}$  over  $[0, 1]$ -valued functions on  $S^{k+1} \times S^{k+1}$  as follows<sup>7</sup>.

$$\Gamma^{\mathcal{C}}(d)(u_0..u_k, v_0..v_k) = \begin{cases} 0 & \text{if } \alpha = 0 \\ 1 & \text{if } \alpha \neq 0, \exists i. u_i \neq_{\ell} v_i \\ \beta + (1 - \beta) \cdot \int d \, d\tau_{\mathcal{C}}(u_k, v_k) & \text{otherwise} \end{cases}$$

where  $\beta = \rho_{\mathcal{C}}(u_0..u_{k-1}, v_0..v_{k-1})(\neq)$  and  $\alpha = \tau_{\mathcal{C}}(u_0, v_0)(u_0..u_k, v_0..v_k)$ .

The operator  $\Gamma^{\mathcal{C}}$  is monotonic w.r.t. the point-wise order on  $[0, 1]$ -valued functions. Hence, applying Tarski's fixed point theorem,  $\Gamma^{\mathcal{C}}$  has a least fixed point, which we denote by  $\gamma_{\lambda}^{\mathcal{C}}$ .

The next result shows that  $\gamma_{\lambda}^{\mathcal{C}}$  is closely related to the discrepancy associated with the coupling structure  $\mathcal{C}$ , and this will eventually be used to compute it.

**Lemma 15.6.4.** *For any coupling structure  $\mathcal{C}$ ,*

$$\mathbb{P}_{s, s'}^{\mathcal{C}}(\neq_{\ell} \omega) = \int \gamma_{\lambda}^{\mathcal{C}} \, d\tau_{\mathcal{C}}(s, s').$$

<sup>7</sup>Since, for all  $u, v \in S$ ,  $\tau_{\mathcal{C}}(u, v)$  is a discrete measure on a finite space, the Lebesgue integral  $\int d \, d\tau_{\mathcal{C}}(u, v)$  in the definition of  $\Gamma^{\mathcal{C}}$  is  $\sum_{x, y \in S^{k+1}} d(x, y) \cdot \tau_{\mathcal{C}}(u, v)(x, y)$ .

*Proof.* Let  $k \in \mathcal{N}$  and  $\mathcal{C} = (\tau_{\mathcal{C}}, \rho_{\mathcal{C}}) \in \mathbf{C}_k(\mathcal{M})$  be a coupling structure for  $\mathcal{M} = (S, \tau, \rho, \ell)$ . Define  $g: S^{k+1} \times S^{k+1} \rightarrow [0, 1]$ , for  $x, y \in S^{k+1}$ , as

$$g(x, y) = \begin{cases} 0 & \text{if } \tau_{\mathcal{C}}(x_0, y_0)(x, y) = 0 \\ \mathbb{P}_{x_0, y_0}^{\mathcal{C}}(\not\equiv_{\ell\omega} | \{(\pi_1, \pi_2)[0..k] = (x, y)\}) & \text{otherwise} \end{cases} \quad (15.6.1)$$

where

- $\mathbb{P}(A|B)$  denotes the *conditional probability* of  $A$  given  $B$  w.r.t.  $\mathbb{P}$  defined, as usual, by  $\mathbb{P}(A|B) = \mathbb{P}(A \cap B)/\mathbb{P}(B)$ , when  $\mathbb{P}(B) > 0$ ;
- $\{(\pi_1, \pi_2)[0..k] = (x, y)\}$  stands for the event  $(\cdot, \cdot)[0..k]^{-1}(\{(x, y)\})$ , where the function  $(\cdot, \cdot)[0..k]$  is defined by  $(\pi_1, \pi_2) \mapsto (\pi_1[0].. \pi_1[k], \pi_2[0].. \pi_2[k])$  – it is not difficult to check that it is measurable.

Note that  $g$  is well defined, since

$$\mathbb{P}_{x_0, y_0}^{\mathcal{C}}(\{(\pi_1, \pi_2)[0..k] = (x, y)\}) = \tau_{\mathcal{C}}(x_0, y_0)(x, y).$$

To prove  $\mathbb{P}_{s, s'}^{\mathcal{C}}(\not\equiv_{\ell\omega}) = \int \gamma_{\lambda}^{\mathcal{C}} d\tau_{\mathcal{C}}(s, s')$  it suffices to show that  $g = \gamma_{\lambda}^{\mathcal{C}}$ . Indeed,

$$\begin{aligned} \mathbb{P}_{s, s'}^{\mathcal{C}}(\not\equiv_{\ell\omega}) &= \int \mathbb{P}_{s, s'}^{\mathcal{C}}(\not\equiv_{\ell\omega} | \{(\pi_1, \pi_2)[0..k] = (\cdot, \cdot)\}) d\mathbb{P}_{s, s'}^{\mathcal{C}}[(\cdot, \cdot)[0..k]] && \text{(cond. pr.)} \\ &= \int \mathbb{P}_{s, s'}^{\mathcal{C}}(\not\equiv_{\ell\omega} | \{(\pi_1, \pi_2)[0..k] = (\cdot, \cdot)\}) d\tau_{\mathcal{C}}(s, s') && \text{(def. } \mathbb{P}^{\mathcal{C}}) \\ &= \int g d\tau_{\mathcal{C}}(s, s') = \int \gamma_{\lambda}^{\mathcal{C}} d\tau_{\mathcal{C}}(s, s'). && \text{(by (15.6.1) and } g = \gamma_{\lambda}^{\mathcal{C}}) \end{aligned}$$

Firstly we prove that  $g$  is a fixed point of  $\Gamma^{\mathcal{C}}$ . We proceed by cases

Case  $\tau_{\mathcal{C}}(x_0, y_0)(x, y) = 0$ . By definition of  $\Gamma^{\mathcal{C}}$  and (15.6.1),  $\Gamma^{\mathcal{C}}(g)(x, y) = 0 = g(x, y)$ .

Case  $\tau_{\mathcal{C}}(x_0, y_0)(x, y) > 0$  and  $\exists i \leq k. x_i \not\equiv_{\ell} y_i$ . The following hold

$$\begin{aligned} g(x, y) &= \mathbb{P}_{x_0, y_0}^{\mathcal{C}}(\not\equiv_{\ell\omega} | \{(\pi_1, \pi_2)[0..k] = (x, y)\}) && \text{(by (15.6.1))} \\ &= \frac{\mathbb{P}_{x_0, y_0}^{\mathcal{C}}(\not\equiv_{\ell\omega} \cap \{(\pi_1, \pi_2)[0..k] = (x, y)\})}{\mathbb{P}_{x_0, y_0}^{\mathcal{C}}(\{(\pi_1, \pi_2)[0..k] = (x, y)\})} && \text{(cond. pr.)} \\ &= \frac{\mathbb{P}_{x_0, y_0}^{\mathcal{C}}(\{(\pi_1, \pi_2)[0..k] = (x, y)\})}{\mathbb{P}_{x_0, y_0}^{\mathcal{C}}(\{(\pi_1, \pi_2)[0..k] = (x, y)\})} = 1 = \Gamma^{\mathcal{C}}(g)(x, y), \end{aligned}$$

where the last equalities follow by  $\{(\pi_1, \pi_2)[0..k] = (x, y)\} \subseteq \not\equiv_{\ell\omega}$  (because by hypothesis  $\exists i. x_i \not\equiv_{\ell} y_i$ ) and definition of  $\Gamma^{\mathcal{C}}$ .

Case  $\tau_C(x_0, y_0)(x, y) > 0$  and  $\forall i \leq k. x_i \equiv_\ell y_i$ . Let

$$A = \{(\pi_1, \pi_2)[\cdot]0..k = (x, y)\} \quad \text{and} \quad B = \{(\pi_1, \pi_2)\langle \cdot \rangle 0..k - 1 \in \neq\},$$

i.e., the event is  $(\cdot, \cdot)\langle 0..k - 1 \rangle^{-1}(\neq)$ , where the function  $(\cdot, \cdot)\langle 0..k - 1 \rangle$  is defined by  $(\pi_1, \pi_2) \mapsto (\pi_1\langle 0 \rangle.. \pi_1\langle k - 1 \rangle, \pi_2\langle 0 \rangle.. \pi_2\langle k - 1 \rangle)$  and it is easy to see that it is measurable.

Let  $\beta = \rho_C(x_0..x_{k-1}, y_0..y_{k-1})(\neq)$ . We show that the following hold

- (i)  $\mathbb{P}_{x_0, y_0}^C(\neq_{\ell\omega} \cap B|A) = \beta;$
- (ii)  $\mathbb{P}_{x_0, y_0}^C(\neq_{\ell\omega} \cap B^c|A) = (1 - \beta) \cdot \int g \, d\tau_C(x_k, y_k).$

Note that once we have shown (i–ii),  $g(x, y) = \Gamma^C(g)(x, y)$  follows immediately because

$$\begin{aligned} g(x, y) &= \mathbb{P}_{x_0, y_0}^C(\neq_{\ell\omega} | A) && \text{(by (15.6.1))} \\ &= \mathbb{P}_{x_0, y_0}^C(\neq_{\ell\omega} \cap B|A) + \mathbb{P}_{x_0, y_0}^C(\neq_{\ell\omega} \cap B^c|A) && \text{(by additivity)} \\ &= \beta + (1 - \beta) \cdot \int g \, d\tau_C(x_k, y_k) && \text{(by (i) and (ii))} \\ &= \Gamma^C(g)(x, y). && \text{(by def. } \Gamma^C) \end{aligned}$$

We show (i):

$$\begin{aligned} \mathbb{P}_{x_0, y_0}^C(\neq_{\ell\omega} \cap B|A) &= \mathbb{P}_{x_0, y_0}^C(B|A) && \text{(by } B \subseteq \neq_{\ell\omega}) \\ &= \rho_C(x_0..x_{k-1}, y_0..y_{k-1})(\neq) && \text{(by def. } \mathbb{P}^C) \\ &= \beta && \text{(by def. } \beta) \end{aligned}$$

We show (ii):

$$\begin{aligned} \mathbb{P}_{x_0, y_0}^C(\neq_{\ell\omega} \cap B^c|A) &= \\ &= \frac{\mathbb{P}_{x_0, y_0}^C(\neq_{\ell\omega} \cap B^c \cap A)}{\mathbb{P}_{x_0, y_0}^C(A)} && \text{(by cond. pr.)} \\ &= \frac{\tau_C(x_0, y_0)(x, y) \cdot \rho_C(x_0..x_{k-1}, y_0..y_{k-1})(=) \cdot \mathbb{P}_{x_k, y_k}^C(\neq_{\ell\omega})}{\tau_C(x_0, y_0)(x, y)} && \text{(by def. } \mathbb{P}^C) \\ &= (1 - \beta) \cdot \mathbb{P}_{x_k, y_k}^C(\neq_{\ell\omega}) && \text{(by def. } \beta \text{ and compl.)} \\ &= (1 - \beta) \cdot \int \mathbb{P}_{x_k, y_k}^C(\neq_{\ell\omega} | \{(\pi_1, \pi_2)[0..k] = (\cdot, \cdot)\}) \, d\mathbb{P}_{x_k, y_k}^C[(\cdot, \cdot)[0..k]] && \text{(cond. pr.)} \\ &= (1 - \beta) \cdot \int \mathbb{P}_{x_k, y_k}^C(\neq_{\ell\omega} | \{(\pi_1, \pi_2)[0..k] = (\cdot, \cdot)\}) \, d\tau_C(x_k, y_k) && \text{(def. } \mathbb{P}^C) \\ &= (1 - \beta) \cdot \int g \, d\tau_C(x_k, y_k). && \text{(by (15.6.1))} \end{aligned}$$

Now we prove, by contradiction, that  $g$  is actually the least fixed point of  $\Gamma^C$ , i.e.,  $\gamma_\lambda^C = g$ .

Assume that  $\gamma_\lambda^C < g$  and let

$$m = \max_{x,y \in S^{k+1}} \{g(x,y) - \gamma_\lambda^C(x,y)\}, \quad x M y \iff g(x,y) - \gamma_\lambda^C(x,y) = m.$$

We show that  $m = 0$ , that is  $\gamma_\lambda^C = g$ . Assume  $x M y$ , we distinguish 3 cases

1. If  $\tau_C(x_0, y_0)(x, y) = 0$ , then by definition of  $\Gamma^C$  and the fact that  $g$  and  $\gamma_\lambda^C$  are fixed points of it, we have that  $m = g(x, y) - \gamma_\lambda^C(x, y) = 0 - 0 = 0$ .
2. If  $\tau_C(x_0, y_0)(x, y) > 0$  and  $x_i \not\equiv_\ell y_i$  for some  $0 \leq i \leq k$ . Analogously, we have that  $m = g(x, y) - \gamma_\lambda^C(x, y) = 1 - 1 = 0$ .
3. If  $\tau_C(x_0, y_0)(x, y) > 0$  and  $x_i \equiv_\ell y_i$  for all  $0 \leq i \leq k$ . Let  $\beta = \rho_C(x, y) (\neq)$ , then the following equalities hold

$$\begin{aligned} m &= g(x, y) - \gamma_\lambda^C(x, y) && \text{(by } x M y \text{)} \\ &= \Gamma^C(g)(x, y) - \Gamma^C(\gamma_\lambda^C)(x, y) && \text{(} g \text{ and } \gamma_\lambda^C \text{ fixed points)} \\ &= (1 - \beta) \cdot \int (g - \gamma_\lambda^C) d\tau_C(x_k, y_k) && \text{(by def. } \Gamma^C \text{)} \\ &= (1 - \beta) \cdot \sum_{u,v \in S^{k+1}} (g(u, v) - \gamma_\lambda^C(u, v)) \cdot \tau_C(x_k, y_k)(u, v). \end{aligned} \quad (15.6.2)$$

By hypothesis on  $m$  and  $\tau_C$  we have respectively that  $g(u, v) - \gamma_\lambda^C(u, v) \leq m$  for all  $u, v \in S^{k+1}$  and  $\sum_{u,v \in S^{k+1}} \tau_C(x_k, y_k)(u, v) = 1$ , therefore it holds that

$$(1 - \beta) \cdot \sum_{u,v \in S^{k+1}} (g(u, v) - \gamma_\lambda^C(u, v)) \cdot \tau_C(x_k, y_k)(u, v) \leq (1 - \beta)m. \quad (15.6.3)$$

We distinguish two cases:

- if  $\beta > 0$ , then  $1 - \beta < 1$ . By (15.6.2) and (15.6.3) we have that  $m \leq (1 - \beta)m$ . By the assumption on  $\beta$  this holds only for  $m = 0$ ;
- if  $\beta = 0$ , by (15.6.2) and (15.6.3) we have that  $g(u, v) - \gamma_\lambda^C(u, v) = m$  whenever  $\tau_C(x_k, y_k)(u, v) > 0$ . Thus  $\tau_C(x_k, y_k)$  has support contained in  $M$ . By the generality of  $x$  and  $y$  one can prove that

$$g(x, y) \stackrel{(15.6.1)}{=} \mathbb{P}_{x_0, y_0}^C (\not\equiv_{\ell\omega} | \{ \pi_1[0..k] = x, \pi_2[0..k] = y \}) = 0.$$

Therefore  $\gamma_\lambda^C(x, y) \not\leq g(x, y) = 0$ , hence  $m = 0$ .

This proves that  $\gamma_\lambda^C = g$ . and concludes our proof.

□

## 15.6.2 Construction of an Optimal Coupling Structure

In this subsection we construct an optimal coupling structure by iterating successive updates of a given coupling structure. We provide necessary and sufficient conditions for a coupling structure  $\mathcal{C}$  to ensure that  $\delta_{\lambda \downarrow k}$  is obtained from  $\gamma_{\lambda}^{\mathcal{C}}$ .

To this end, we first introduce the notion of update for a coupling structure.

**Definition 15.6.5** (Update). *Let  $\mathcal{C} = (\tau_{\mathcal{C}}, \rho_{\mathcal{C}}) \in \mathbf{C}_k(\mathcal{M})$ . For  $\mu \in \Omega(\mathbf{S}^k(u), \mathbf{S}^k(v))$  and  $\nu \in \Omega(\mathbb{T}^k(u_1..u_k), \mathbb{T}^k(v_1..v_k))$ , define*

- *transition update:  $\mathcal{C}[(u, v) / \mu] = (\tau_{\mathcal{C}}[(u, v) \mapsto \mu], \rho_{\mathcal{C}})$ ;*
- *delay update:  $\mathcal{C}\langle(u_1..u_k, v_1..v_k) / \nu\rangle = (\tau_{\mathcal{C}}, \rho_{\mathcal{C}}[(u_1..u_k, v_1..v_k) \mapsto \nu])$ .*

where, for a function  $f: X \rightarrow Y$ ,  $f[x \mapsto y]$  denotes the update of  $f$  at  $x$  with  $y$ .

Our update strategy relies on the following result.

**Lemma 15.6.6** (Update criteria). *Let  $\mathcal{C} = (\tau_{\mathcal{C}}, \rho_{\mathcal{C}}) \in \mathbf{C}_k(\mathcal{M})$  be a coupling structure and  $u_0..u_k, v_0..v_k \in S$  be such that  $\tau_{\mathcal{C}}(u_0..u_k, v_0..v_k) > 0$  and, for all  $i \leq k$ ,  $u_i \equiv_{\ell} v_i$ . Then, for  $\mu \in \Omega(\mathbf{S}^k(u_k), \mathbf{S}^k(v_k))$ ,  $\nu \in \Omega(\mathbb{T}^k(u_0..u_{k-1}), \mathbb{T}^k(v_0..v_{k-1}))$  and*

$$\mathcal{D} = \mathcal{C}[(u_k, v_k) / \mu] \langle(u_0..u_{k-1}, v_0..v_{k-1}) / \nu\rangle,$$

the inequality  $\gamma_{\lambda}^{\mathcal{D}} < \gamma_{\lambda}^{\mathcal{C}}$  holds whenever one of the following conditions is fulfilled.

- (i)  $\nu(\neq) < \rho_{\mathcal{C}}(u_0..u_{k-1}, v_0..v_{k-1})(\neq)$  and  $\int \gamma_{\lambda}^{\mathcal{C}} d\mu \leq \int \gamma_{\lambda}^{\mathcal{C}} d\tau_{\mathcal{C}}(u_k, v_k)$ , or
- (ii)  $\nu(\neq) \leq \rho_{\mathcal{C}}(u_0..u_{k-1}, v_0..v_{k-1})(\neq)$  and  $\int \gamma_{\lambda}^{\mathcal{C}} d\mu < \int \gamma_{\lambda}^{\mathcal{C}} d\tau_{\mathcal{C}}(u_k, v_k)$ .

*Proof.* Let  $\mathcal{C} = (\tau_{\mathcal{C}}, \rho_{\mathcal{C}}) \in \mathbf{C}_k(\mathcal{M})$  be a coupling structure and  $u_0..u_k, v_0..v_k \in S$  be such that  $\tau_{\mathcal{C}}(u_0..u_k, v_0..v_k) > 0$  and for all  $i \leq k$ ,  $u_i \equiv_{\ell} v_i$ .

Consider  $\mu \in \Omega(\mathbf{S}^k(u_k), \mathbf{S}^k(v_k))$ ,  $\nu \in \Omega(\mathbb{T}^k(u_0..u_{k-1}), \mathbb{T}^k(v_0..v_{k-1}))$  and let  $\mathcal{D} = \mathcal{C}[(u_k, v_k) / \mu] \langle(u_0..u_{k-1}, v_0..v_{k-1}) / \nu\rangle$  be an update of  $\mathcal{C}$ .

We will prove that if (i) or (ii) holds, then  $\gamma_{\lambda}^{\mathcal{C}}$  is a proper prefixed point of  $\Gamma^{\mathcal{D}}$ , that is,  $\Gamma^{\mathcal{D}}(\gamma_{\lambda}^{\mathcal{C}}) < \gamma_{\lambda}^{\mathcal{C}}$ . Then, the thesis follows by Tarski's fixed point theorem.

To this end, let fix  $\alpha, \alpha'$  and  $\beta, \beta'$  as

$$\begin{aligned} \alpha &= \int \gamma_{\lambda}^{\mathcal{C}} d\mu & \alpha' &= \int \gamma_{\lambda}^{\mathcal{C}} d\tau_{\mathcal{C}}(u_k, v_k), \\ \beta &= \nu(\neq) & \beta' &= \rho_{\mathcal{C}}(u_0..u_{k-1}, v_0..v_{k-1})(\neq). \end{aligned}$$

Then, the following inequalities hold

$$\begin{aligned}
\Gamma^{\mathcal{D}}(\gamma_{\lambda}^{\mathcal{C}})(u_0..u_k, v_0..v_k) &= \\
&= \beta + (1 - \beta)\alpha && \text{(def. } \Gamma^{\mathcal{D}}) \\
&\leq \beta + (1 - \beta)\alpha' && (\alpha \leq \alpha') \\
&= \alpha' - \alpha' + \beta + (1 - \beta)\alpha' \\
&= \alpha' - \beta\alpha' - (1 - \beta)\alpha' + \beta + (1 - \beta)\alpha' && (0 \leq \beta \leq 1) \\
&= \alpha' - \beta\alpha' + \beta = \alpha' + (1 - \alpha')\beta \\
&\leq \alpha' + (1 - \alpha')\beta' && (\beta \leq \beta') \\
&= \beta' + (1 - \beta')\alpha' && \text{(same as for } \beta + (1 - \beta)\alpha' = \alpha' + (1 - \alpha')\beta) \\
&= \Gamma^{\mathcal{C}}(\gamma_{\lambda}^{\mathcal{C}})(u_0..u_k, v_0..v_k) && \text{(def. } \Gamma^{\mathcal{C}}) \\
&= \gamma_{\lambda}^{\mathcal{C}}. && \text{(def. } \gamma_{\lambda}^{\mathcal{C}})
\end{aligned}$$

In particular, for (i)  $\beta < \beta'$  or for (ii)  $\alpha < \alpha'$ , the above inequality is strict.

By construction of  $\mathcal{D}$  and definition of  $\Gamma$ , it is immediate to prove that, for arbitrary  $x, y \in S^{k+1}$ ,  $\Gamma^{\mathcal{D}}(\gamma_{\lambda}^{\mathcal{C}})(x, y) \leq \gamma_{\lambda}^{\mathcal{C}}(x, y)$ .

This proves that if (i) or (ii) holds, then  $\gamma_{\lambda}^{\mathcal{D}} < \gamma_{\lambda}^{\mathcal{C}}$ .  $\square$

The condition (i) in Lemma 15.6.6 ensures that any  $\mathcal{C} = (\tau_{\mathcal{C}}, \rho_{\mathcal{C}}) \in \mathbf{C}_k(\mathcal{M})$  is improved by replacing  $\rho_{\mathcal{C}}$  with the function  $\rho^*: S^k \times S^k \rightarrow \Delta(\mathbb{R}_{\geq 0}^k \times \mathbb{R}_{\geq 0}^k)$  defined as

$$\begin{aligned}
\rho^*(u_0..u_{k-1}, v_0..v_{k-1}) &= \min\{v(\neq) \mid v \in \Omega(\mathbb{T}^k(u_0..u_{k-1}), \mathbb{T}^k(v_0..v_{k-1}))\} \\
&= \|\mathbb{T}^k(u_0..u_{k-1}) - \mathbb{T}^k(v_0..v_{k-1})\| && \text{(Lemma 15.5.4)} \\
&= 1 - \prod_{i=0}^{k-1} (1 - \|\rho(u_i) - \rho(v_i)\|) = \beta^*,
\end{aligned}$$

where the last equality follows from the definition of  $\mathbb{T}^k(u_0..u_{k-1})$  and  $\mathbb{T}^k(v_0..v_{k-1})$  as product measures.

Notice that, assuming A2, the above is computable. By replacing  $\beta$  in the definition of  $\Gamma^{\mathcal{C}}$  with  $\beta^*$ ,  $\gamma_{\lambda}^{\mathcal{C}}$  can be computed as the least solution of the linear equation system induced by the definition of  $\Gamma^{\mathcal{C}}$ .

The condition (ii) of Lemma 15.6.6 suggests to improve  $\mathcal{C}$  with  $\mathcal{C}[(u_k, v_k)/\mu^*]$  where

$$\begin{aligned}
\mu^* &= \arg \min\{\int \gamma_{\lambda}^{\mathcal{C}} d\mu \mid \mu \in \Omega(S^k(u_k), S^k(v_k))\} \\
&= \arg \min\{\sum_{x,y \in S^{k+1}} \gamma_{\lambda}^{\mathcal{C}}(x, y) \cdot \mu(x, y) \mid \mu \in \Omega(S^k(u_k), S^k(v_k))\}.
\end{aligned}$$

The above is a linear program (a.k.a. *transportation problem*), hence computable.

The sufficient conditions for termination is provided by the following lemma.

**Lemma 15.6.7.** *Let  $\mathcal{C} = (\tau_{\mathcal{C}}, \rho^*) \in \mathbf{C}_{2k}(\mathcal{M})$  be such that  $\delta_{\lambda \downarrow k}(u, v) \neq \int \gamma_{\lambda}^{\mathcal{C}} d\tau_{\mathcal{C}}(u, v)$  for some  $u, v \in S$ . Then, there exist  $u', v' \in S$  and  $\mu \in \Omega(S^{2k}(u'), S^{2k}(v'))$  such that*

$$\int \gamma_{\lambda}^{\mathcal{C}} d\mu < \int \gamma_{\lambda}^{\mathcal{C}} d\tau_{\mathcal{C}}(u', v').$$

*Proof.* Assume that  $\delta_{\lambda \downarrow k}(u, v) \neq \int \gamma_{\lambda}^{\mathcal{C}} d\tau_{\mathcal{C}}(u, v)$  for some  $u, v \in S$  and that for all  $u', v' \in S$  and all  $\mu \in \Omega(\mathbb{S}^{2^k}(u'), \mathbb{S}^{2^k}(v'))$ , we have

$$\int \gamma_{\lambda}^{\mathcal{C}} d\mu \geq \int \gamma_{\lambda}^{\mathcal{C}} d\tau_{\mathcal{C}}(u', v').$$

The hypothesis and Lemma 15.6.6, guarantee that

$$\int \gamma_{\lambda}^{\mathcal{C}} d\tau_{\mathcal{C}}(u, v) = \min\{\int \gamma_{\lambda}^{\mathcal{D}} d\tau_{\mathcal{D}}(u, v) \mid \mathcal{D} \in \mathbf{C}_{2^k}(\mathcal{M})\}.$$

But at the same time

$$\begin{aligned} \delta_{\lambda \downarrow k}(u, v) &= \min\{\mathbb{P}_{u, v}^{\mathcal{D}}(\neq_{\ell^{\omega}}) \mid \mathcal{D} \in \mathbf{C}_{2^k}(\mathcal{M})\} && \text{(by (15.5.2))} \\ &= \min\{\int \gamma_{\lambda}^{\mathcal{D}} d\tau_{\mathcal{D}}(u, v) \mid \mathcal{D} \in \mathbf{C}_{2^k}(\mathcal{M})\}. && \text{(by Lemma 15.6.4)} \end{aligned}$$

This contradicts the hypothesis that  $\int \gamma_{\lambda}^{\mathcal{C}} d\tau_{\mathcal{C}}(u, v) \neq \delta_{\lambda \downarrow k}(u, v)$ .  $\square$

Intuitively, the above ensures that, unless  $\mathcal{C}$  is an optimal coupling structure, (ii) in Lemma 15.6.6 is satisfied, so that, we can further improve  $\mathcal{C}$  as aforesaid.

**Proposition 15.6.8.** *Assuming A2,  $\delta_{\lambda \downarrow k}$  is computable for all  $k \in \mathcal{N}$ .*

*Proof.* The aforementioned strategy ensures that the updated couplings are chosen from the vertices of the polytopes  $\Omega(\mathbb{S}^k(u), \mathbb{S}^k(v))$ , for  $u, v \in S$ . Since these polytopes have finitely many vertexes, the procedure eventually terminates. By Lemma 15.6.7, the last coupling describes  $\delta_{\lambda \downarrow k}$ .  $\square$

## 15.7 Conclusive Remarks

In this Chapter we have showed that the trace distance is the appropriate behavioral distance to reason about linear real-time properties. This has been done by closing the gap between the total variation distance and the variational distances with respect to the events represented either as MTL formulas or languages recognized by TAs.

Actually, we showed that the same result is obtained by considering smaller fragments of MTL, that are strictly less expressive than the entire logic. Analogously, we showed that for the case of TAs, it suffices to consider the subclass of single-clock always resetting DTAs.

We have also studied the problem of approximating the trace distance within any absolute error. We showed that the problem is computable by approximating the total variation distance both from above and below by means of the computable sequences  $\{\delta_{\lambda \downarrow k}\}_k$  and  $\{\delta_{\lambda \uparrow k}\}_k$ . This both extends the result of [17] to the real-time setting and gives an alternative way to approximate the total variation distance on MCs.

For future, we intend to further explore the potentiality of the presented results by studying how fast the sequences convergence to the total variation distance. Moreover, we would like to see if similar results can be used to link different behavioral distances, such as the Kantorovich-based bisimilarity distance and the total variation (for which the former is know to be an upper bound of the latter) opening for the possibility to “bridge the gap” between trace and branching-based behavioral distances.

From a computational perspective, also motivated by our previous work [5] on MCs presented in Chapter 11, we would like to implement an on-the-fly algorithm for computing tight over-approximations of the trace distance.



# Bibliography

- [1] R. Alur and D. Dill. Automata for Modeling real-time Systems. In M. S. Paterson, editor, *Automata, Languages and Programming*, volume 443 of *Lecture Notes in Computer Science*, pages 322–335. Springer Berlin Heidelberg, 1990.
- [2] R. Alur and T. A. Henzinger. Real-Time Logics: Complexity and Expressiveness. *Information and Computation*, 104(1):35–77, 1993.
- [3] R. Alur and T. A. Henzinger. A Really Temporal Logic. *Journal of the ACM*, 41(1):181–204, 1994.
- [4] G. Bacci, G. Bacci, K. G. Larsen, and R. Mardare. Computing Behavioral Distances, Compositionally. In *MFCS2013*, volume 8087 of *Lecture Notes in Computer Science*, page 74-85 Springer Berlin Heidelberg, 2013.
- [5] G. Bacci, G. Bacci, K. G. Larsen, and R. Mardare. On-the-Fly Exact Computation of Bisimilarity Distances. In *TACAS*, volume 7795 of *Lecture Notes in Computer Science*, pages 1–15, 2013.
- [6] G. Bacci, G. Bacci, K. G. Larsen, and R. Mardare. On the Total Variation Distance of Semi-Markov Chains. Technical report, Aalborg University, DK, 2014. [http://people.cs.aau.dk/~grbacci/Papers/smc\\_dist.pdf](http://people.cs.aau.dk/~grbacci/Papers/smc_dist.pdf).
- [7] Christel Baier and Joost Pieter Katoen. *Principles of Model Checking*. MIT Press, 2008.
- [8] P. Billingsley. *Probability and Measure*. Wiley, New York, 3rd edition, 1995.
- [9] Graham Brightwell and Jan van den Heuvel, and Leen Stougie. A Linear Bound On The Diameter Of The Transportation Polytope. *Combinatorica*, vol. 26.2:133–139, 2006
- [10] Xiaojuan Cai and Yonggen Gu. Measuring Anonymity. *ISPEC '09*:183–194, 2009.
- [11] P. S. Castro and D. Precup. Using bisimulation for policy transfer in MDPs. In *Proceedings of the 9th International Conference on Autonomous Agents and Multiagent Systems*, volume 1 of *AAMAS '10*, pages 1399–1400, Richland, SC, 2010. International Foundation for Autonomous Agents and Multiagent Systems.

- [12] P. S. Castro and D. Precup. Automatic Construction of Temporally Extended Actions for MDPs Using Bisimulation Metrics. In S. Sanner and M. Hutter, editors, *Recent Advances in Reinforcement Learning*, volume 7188 of *Lecture Notes in Computer Science*, pages 140–152. Springer Berlin Heidelberg, 2012.
- [13] Krishnendu Chatterjee and Luca de Alfaro and Rupak Majumdar and Vishwanath Raman. Algorithms for Game Metrics. *Logical Methods in Computer Science*, 6.3, 2010.
- [14] T. Chen, F. van Breugel, and J. Worrell. On the Complexity of Computing Probabilistic Bisimilarity. In *FoSSaCS*, volume 7213 of *Lecture Notes in Computer Science*, pages 437–451. Springer, 2012.
- [15] T. Chen, M. Diciolla, M. Z. Kwiatkowska, and A. Mereacre. Time-Bounded Verification of CTMCs against Real-Time Specifications. In *FORMATS*, volume 6919 of *Lecture Notes in Computer Science*, pages 26–42, 2011.
- [16] T. Chen, T. Han, J.-P. Katoen, and A. Mereacre. Model checking of continuous-time markov chains against timed automata specifications. *Logical Methods in Computer Science*, 7(1), 2011.
- [17] T. Chen and S. Kiefer. On the Total Variation Distance of Labelled Markov Chains. In *Proc. of CSL-LICS '14*, CSL-LICS '14, pages 33:1–33:10, New York, NY, USA, 2014. ACM.
- [18] Gheorghe Comanici and Doina Precup. Basis function discovery using spectral clustering and bisimulation metrics. *AAMAS '11*, vol.3: 1079–1080, 2011.
- [19] Gheorghe Comanici and Prakash Panangaden and Doina Precup. On-the-Fly Algorithms for Bisimulation Metrics. *International Conference on Quantitative Evaluation of Systems*, Vol. 0:94–103, 2012.
- [20] George B. Dantzig Application of the Simplex method to a transportation problem. *Activity analysis of production and allocation*, 359–373, J. Wiley, New York, 1951.
- [21] L. de Alfaro, R. Majumdar, V. Raman, and M. Stoelinga. Game Relations and Metrics. In *LICS*, pages 99–108, July 2007.
- [22] J. Desharnais, V. Gupta, R. Jagadeesan, and P. Panangaden. Metrics for labelled Markov processes. *Theoretical Computer Science*, 318(3):323–354, 2004.
- [23] O. Demuth. Poznámka k dopravnímu [Czech, with German summary; A remark on the transportation problem] *Časopis pro Pěstování Matematiky*, no.86:103–110, 1961.
- [24] N. Dunford and J. T. Schwartz. *Linear Operators, Part 1, General Theory*. Wiley Classic Library. John Wiley, New York, 1988.

- [25] Norm Ferns and Prakash Panangaden and Doina Precup. Metrics for finite Markov Decision Processes. In proc. of the 20th conference on Uncertainty in Artificial Intelligence, 162–169, 2004.
- [26] Lester Randolph Ford and Delbert Ray Fulkerson. Solving the Transportation Problem. *Management Science*, vol.3.1: 24–32, 1956.
- [27] A. Giacalone, C. Jou, and S. A. Smolka. Algebraic reasoning for probabilistic concurrent systems. In *Proc. IFIP TC2 Working Conference on Programming Concepts and Methods*, pages 443–458. North-Holland, 1990.
- [28] R. Givan, T. Dean, and M. Greig. Equivalence notions and model minimization in Markov decision processes. *Artificial Intelligence*, 147(1-2):163–223, 2003.
- [29] David Griffeath. A maximal coupling for Markov chains. *Probability Theory and Related Fields*, vol.31.2:95–106, 1975.
- [30] Martin Grötschel and László Lovász and Alexander Schrijver. *Geometric Algorithms and Combinatorial Optimization*. Algorithms and Combinatorics, vol.2, 1988 publisher = Springer,
- [31] Victor Klee and Christoph Witzgall. *Mathematics of the Decision Sciences, Part I*. Ed. G. B. Dantzig and A. F. Veniott, American Mathematical Society, 1986.
- [32] Kim Guldstrand Larsen and Arne Skou. Bisimulation through Probabilistic Testing. *Information and Computation*, no. 1, vol. 94:1–28, 1991.
- [33] T. Lindvall. *Lectures on the Coupling Method*. Wiley Series in Probability and Mathematical Statistics. John Wiley, New York, 1992.
- [34] R. B. Lyngsø and C. N. Pedersen. The consensus string problem and the complexity of comparing hidden Markov models. *Journal of Computer and System Sciences*, 65(3):545–569, 2002. Special Issue on Computational Biology 2002.
- [35] Michael Mitzenmacher and Eli Upfal. *Probability and Computing - randomized algorithms and probabilistic analysis*. Cambridge University Press, 2005.
- [36] Prakash Panangaden *Labelled Markov Processes*. Imperial College Press, 2009.
- [37] M. L. Puterman. *Markov Decision Processes: Discrete Stochastic Dynamic Programming*. John Wiley & Sons, Inc., New York, NY, USA, 1st edition, 1994.
- [38] J. Ouaknine and J. Worrell. On the decidability and complexity of Metric Temporal Logic over finite words. *Logical Methods in Computer Science*, 3(8), 2007.
- [39] A. Sharma and J.-P. Katoen. Weighted Lumpability on Markov Chains. In *Ershov Memorial Conference*, volume 7162 of *Lecture Notes in Computer Science*, pages 322–339, 2011.

- [40] Alexander Schrijver. *Theory of linear and integer programming*. John Wiley & Sons, Inc., 1986.
- [41] David Thorsley and Eric Klavins. Approximating stochastic biochemical processes with Wasserstein pseudometrics. *IET Systems Biology*, 4.3:193-211, 2010.
- [42] F. van Breugel, B. Sharma, and J. Worrell. Approximating a Behavioural Pseudometric without Discount for Probabilistic Systems. *Logical Methods in Computer Science*, 4(2), 2008.
- [43] Franck van Breugel and James Worrell. An Algorithm for Quantitative Verification of Probabilistic Transition Systems. In *Proc. CONCUR 2001*:336–350, 2001.
- [44] F. van Breugel and J. Worrell. Approximating and computing behavioural distances in probabilistic transition systems. *Theoretical Computer Science*, 360(3):373–385, 2006.
- [45] Franck van Breugel. On behavioural pseudometrics and closure ordinals. *Information Processing Letters*, 112(19):715–718, 2012.
- [46] Franck van Breugel, Claudio Hermida, Michael Makkai, and James Worrell. Recursively defined metric spaces without contraction. *Theoretical Computer Science*, 380(1-2):143–163, 2007.
- [47] Franck van Breugel and James Worrell. Towards quantitative verification of probabilistic transition systems. In *ICALP*, volume 2076 of *LNCS*, pages 421–432, 2001.
- [48] Cédric Villani. *Topics in Optimal Transportation*. Number 58 in Graduate Study in Mathematics. American Mathematical Society, Providence, RI, 2003.





# Chapter 16

## Dansk résumé

Komplekse systemer, der kombinerer kunstige (software-baserede) komponenter og naturlige komponenter udgør de nye udfordringer indenfor ingeniør- og teknologividen- skab. Sådanne systemer forefindes indenfor adskillige områder, såsom luftfart, bilindus- tri, kemiske processer, civil infrastruktur, energi, sundhedsvesenet, produktionssyste- mer, transport, og husholdningsapparater. Ved analyse af disse systemer, repræsenterer vi dem ofte som stokastiske processer for at modellere manglende viden, usikkerhed, eller tilfældigheder. I denne monografi vil vi udvikle et logisk grundlag, der vil hjælpe os til at forstå hvorledes vi kan validere modeller i forhold til de systemer der modelleres. Målet er at forstå hvornår egenskaber, der observeres på modeller kan konkluderes at være korrekte for det modellerede system. For at opnå dette mål, vil vi udvikle en approksima- tionsteori for stokastiske systemer, der vil formalisere denne korrespondance mellem modeller og systemer, samt understøtte modellering og simulering af komplekse syste- mer. Den foreslåede forskning kombinerer viden fra logik, matematik, datalogi og in- genieurvidenskab og de opnåede resultater forventes at ville have en betydelig effekt på alle disse områder.