Deposited via The University of Sheffield.

White Rose Research Online URL for this paper:
https://eprints.whiterose.ac.uk/id/eprint/237266/

Version: Published Version

Proceedings Paper:

Constructing Witnesses for Lower Bounds on Behavioural Distances

Ruben Turkenburg ✉
Radboud University, Nijmegen, The Netherlands

Harsh Beohar ✉
University of Sheffield, UK

Franck van Breugel ✉
York University, Toronto, Canada

Clemens Kupke ✉
University of Strathclyde, Glasgow, UK

Jurriaan Rot ✉
Radboud University, Nijmegen, The Netherlands

Abstract

Behavioural distances provide a robust alternative to notions of equivalence such as bisimilarity in the context of probabilistic transition systems. They can be defined as least fixed points, whose universal property allows us to exhibit upper bounds on the distance between states, showing them to be *at most* some distance apart.

In this paper, we instead consider the problem of bounding distances from below, showing states to be *at least* some distance apart. Contrary to upper bounds, it is possible to reason about lower bounds inductively. We exploit this by giving an inductive derivation system for lower bounds on an existing definition of behavioural distance for labelled Markov chains. This is inspired by recent work on *apartness* as an inductive counterpart to bisimilarity. Proofs in our system will be shown to closely match the behavioural distance by soundness and (approximate) completeness results.

We further provide a constructive correspondence between our derivation system and formulas in a modal logic with quantitative semantics. This logic was used in recent work of Rady and van Breugel to construct evidence for lower bounds on behavioural distances. Our constructions provide smaller witnessing formulas in many examples.

1 Introduction

Bisimilarity is an important notion of equivalence in the study of state-transition systems. It is *qualitative* in the sense that states are either considered equivalent, or not; there are no degrees of equivalence. When studying systems involving probabilistic transitions,

such qualitative definitions are usually considered too strict; states may be *inequivalent* or *distinguishable* under bisimilarity despite their behaviour being difficult to distinguish by an observer (this problem was first noted by Giacalone, Jou, and Smolka [26]).

To better capture the (in)equivalence of states, quantitative notions of *behavioural distances/metrics* may be used [15, 54, 57]. These assign to each pair of states a number (e.g., in the interval $[0,1]$) representing how close (or how far apart) their behaviours are. Determining these distances has been studied algorithmically, with procedures developed for approximating the distance [56], and the exact computation of distances [11, 2, 4, 49]. The definition of behavioural distances as least or greatest fixed points (depending on the chosen ordering), also gives them a universal property yielding a (co)inductive proof principle [28, 5, 7]. The corresponding proofs are of bounds on the distances showing states to be equivalent to some degree. This is analogous to the qualitative proof technique of exhibiting some bisimulation containing a pair of states, thereby showing them to be bisimilar.

Apartness.   Orthogonally, interest in (qualitative) apartness of states has been growing as an inductive counterpart to bisimilarity [25]. Rather than defining when states behave the same, apartness defines when there is some observable difference between them. A reason for interest in apartness is its inductive and potentially constructive nature. Indeed, this was the original motivation, going back to the school of Brouwer [29]. In the setting of state-based systems, this means giving some (finite) evidence or witness for a difference in behaviours. Think of, for example, a word which is accepted by one state of a finite automaton but not another, or a difference in probability of making a certain observation in probabilistic systems such as labelled Markov chains or Markov decision processes.

As in the case of bisimilarity, we would like these notions to be as robust as possible, making behavioural distances a clear area of interest. In the quantitative setting, the dual of the existing coinductive proof methods allows us to obtain lower bounds on distances between states, i.e., we can show states to be at least a certain distance apart. This type of bound has recently been explored in [6]. They define a measure of how much a candidate for the least fixed point can be decreased. If no such decrease is possible, we have a lower bound.

In this work, we take an alternative approach, based essentially on Kleene's chain construction of least fixed points [33]. For the case of behavioural distances, this starts from an order-preserving functional, say $\Gamma \colon \mathsf{PMet}_X \to \mathsf{PMet}_X$, on the space of pseudometric spaces on a set X. To approach the least fixed point $\mu\Gamma$ from below, we can start from the constant zero distance $\bot$ and iteratively apply Γ giving the chain $\bot \leq \Gamma(\bot) \leq \Gamma^2(\bot) \leq \ldots$. As is noted in [6], fully applying Γ iteratively in this way is not a desirable means of obtaining bounds. Instead, we will develop an inductive derivation system allowing the construction of lower bounds for chosen pairs of states.

However, simply translating the definition of the behavioural distance to proof rules does not work; to obtain a usable derivation system we must show that proof steps need only consider direct successors of the involved states. This means that finite derivations can be constructed also for systems with infinite state spaces. We further reduce the work required for proofs, and thus also the size of the derivations, by showing that recursive proofs are only required for a subset of these successors.

The judgments proved in our system are of the form $x \mathrel{\#_\varepsilon} y$ for x, y states in an LMC and ε a rational in the interval $[0,1]$. To ensure a correct relation to the behavioural distance, we show two properties. First, soundness: that if we can prove $x \mathrel{\#_\varepsilon} y$, then ε is a lower bound on the behavioural distance between x and y. Second, a form of completeness which we call *approximate* completeness. Usual completeness with respect to the behavioural

distance would state that any distance between states can be proved in the derivation system. However, in the spirit of apartness, we consider *finite* evidence, which can only witness finite approximations of distances in general. We thus show that lower bounds can be derived with arbitrarily small error with respect to the true distance.

Logics. Evidence of differences in behaviour can also be given in the form of formulas in some modal logic. This is closely related to Hennessy-Milner type theorems, which show for a given logic that bisimilarity and logical equivalence coincide. Such theorems have been shown for probabilistic bisimilarity and a logic with a modality parameterised by rational probabilities [13, 14, 23]. Later, a quantitative analogue was shown by relating a real-valued logic introduced in [15] to behavioural distances [55]. The most interesting part in the qualitative case is the inclusion of logical equivalence in bisimilarity, also called expressiveness, which can dually be shown by giving, for each pair of non-bisimilar states, a formula which distinguishes them. Quantitatively, this means giving a formula for which the difference in interpretations matches the behavioural distance as closely as possible.

This correspondence of behavioural distances and modal formulas has been investigated for *labelled Markov chains* (LMCs) in [44]. In their terminology, a construction is given of formulas "explaining" the distance between states. Due to the chosen logic, and the possibility of infinite behaviours in LMCs induced by loops, this can not be done exactly. Instead, it is shown that for any finite approximation $\Gamma^i(\bot)(x, y)$ of the distance of states as in the above chain, a formula can be constructed such that $|[\![\varphi]\!](x) - [\![\varphi]\!](y)|$ (the difference in interpretations on states x and y) is equal to the approximation.

We finish by relating our new derivation system to the work of [44] by showing that for any derivation, a formula in the modal logic of *op. cit.* can be constructed which witnesses the same bound. These witnessing formulas are an improvement as they can be given for infinite state systems, and they will be smaller in many examples. Further, for any formula a proof tree witnessing the same lower bound can be given whose depth will be equal to the modal depth of the formula. For more fine-grained notions of size counting the total number of steps in a derivation and number of operators in a formula, the derivations will be larger in general, as they are dependent on the system and thus contain more information. This is exactly what facilitates the aforementioned improvements; we see the steps which lead us to conclude a difference in behaviours which are otherwise somewhat hidden in the semantics of the logic. Proofs also focus on pairs of states, so that at each step we see which states of a system are being used to exhibit a lower bound.

Contributions. Summarising the contributions of the paper:
- We define an inductive derivation system for lower bounds on behavioural distances in labelled Markov chains (Section 3)
- We show the soundness and approximate completeness of the system with respect to the behavioural distance (Sections 3.1 and 3.2)
- We show a constructive correspondence between proofs in our system and formulas of a modal logic (Sections 4 and 5)

We illustrate all of the above with examples, including a system with an infinite state space which was not covered by this modal logic (Example 25).

Related Work. The line of work focussing on proofs of apartness for state-based systems was (re)started by Geuvers and Jacobs [25], with further work on the relation to distinguishing formulas in [24]. A proof system for an apartness notion dual to coalgebraic behavioural equivalence has been given in [51]. The current paper builds directly on these works.

The coinductive proof principle in the context of behavioural distances has been explored coalgebraically in, e.g., [46, 8, 5, 7]. In the greatest fixed point characterisation, the ordering is reversed compared to the definition we use in the rest of this work. Coinduction thus leads to upper bounds on distances under our definition. An approach focussed on bounding greatest fixed points from above (but which dually bounds least fixed points from below as we will do) has more recently been given in [6], as discussed above. There is however no construction given of formulas demonstrating proved bounds.

The construction of distinguishing formulas has also been studied in the qualitative setting in, e.g., [41, 42, 37, 60]. We discuss these works further in Section 6.

A more general account of bounding distances from above is the area of quantitative equational theories [40], which has been applied to give a calculus for upper bounds on distances in Markov chains [3] and regular expressions [45].

Notation. We will write, $\mathcal{D}_{\mathbb{Q}}(X)$ for the set of finitely-supported rational distributions on the set X. These are maps $\mu\colon X \to [0,1]\cap\mathbb{Q}$ such that $\mathrm{supp}(\mu) := \{x \in X \mid \mu(x) \neq 0\}$ is finite and $\sum_{x\in X}\mu(x) = 1$. We may also write such distributions as formal sums: $\sum_{x\in X}\mu(x)\,|x\rangle$, with the $|x\rangle$ ("ket") notation taken from [30] and used simply to separate states and their associated probabilities. It can also be thought of as indicating a sum of Dirac distributions. From now on, we will write $[0,1]_{\mathbb{Q}}$ for $[0,1]\cap\mathbb{Q}$.

We denote by PMet_X the set of pseudometric spaces on a set X, i.e., pairs (X,d) with $d\colon X \times X \to [0,1]$ a pseudometric. We order the unit interval with the usual ordering of the reals, and pseudometrics inherit this ordering pointwise, so that $d_1 \leq d_2$ iff $\forall x,y \in X.\, d_1(x,y) \leq d_2(x,y)$. The smallest element $\bot$ is thereby the constant zero distance. Further, for two pseudometric spaces $(X,d_X) \in \mathsf{PMet}_X$ and $(Y,d_Y) \in \mathsf{PMet}_Y$ a map $f\colon (X,d_X) \to (Y,d_Y)$ will be assumed to be *non-expansive* (also called *1-Lipschitz*, *short*, etc.), i.e., $\forall x,y \in X.\, d_Y(f(x),f(y)) \leq d_X(x,y)$. The Euclidean distance is denoted $d_e\colon [0,1] \times [0,1] \to [0,1]$.

In the interest of space, we write $\mu \cdot h$ for $\sum_{x\in X}\mu(x) \cdot h(x)$ where $\mu\colon X \to [0,1]_{\mathbb{Q}}$ is a distribution and $h\colon X \to \mathbb{R}$ is an arbitrary function. This is motivated by viewing the distributions and functions as vectors indexed by their common domain, so that the operation is the vector dot product. In the sequel, we will often restrict h to maps into $[0,1]_{\mathbb{Q}}$.

2   Behavioural Distances on LMCs

We start with the definition of the type of system which we study in the remainder of the paper: labelled Markov chains.

▶ **Definition 1.** *A labelled Markov chain (LMC) consists of the following data:*
- *a set of states X;*
- *a (non-empty) set of labels L;*
- *a (finitely branching) probabilistic transition function $\tau\colon X \to \mathcal{D}_{\mathbb{Q}}(X)$; and*
- *a labelling function $l\colon X \to L$*

▶ **Example 2.** Let $X = \{x, x_1, x_2\}$ and $L = \{a, b\}$. We represent the LMC (X, L, τ, l) with $\tau(x) = \frac{1}{2}\,|x_1\rangle + \frac{1}{2}\,|x_2\rangle,\, \tau(x_1) = 1\,|x_1\rangle,\, \tau(x_2) = 1\,|x_2\rangle$ and $l(x) = l(x_1) = a,\, l(x_2) = b$ as:
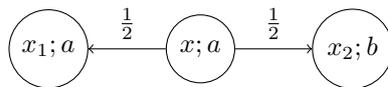
We use the notation $x;a$ for a state $x \in X$ such that $l(x) = a$. Further, any state with no outgoing edges is assumed to have a self-loop with probability 1.

We now recall a definition of the behavioural distance (henceforth written bd) of states in an LMC as the least fixed point of a functional based on non-expansive maps. This distinguishes two cases: states having different labels should be maximally far apart, so they have distance 1; the distance of states with the same label is then defined recursively, and can be seen as an optimisation problem. Intuition for this problem is most often given in terms of its dual based on couplings under the Kantorovich-Rubinstein duality. The distance between distributions can in that setting be seen as the minimal cost of transporting one unit of mass from one distribution to the other, with the cost of transporting m probability mass between states at distance d being $m \cdot d$. In a more general form (as discussed in [58]) the distance below can be seen as the maximisation of profit, where we think of buying from one distribution and selling to the other, with the maps h representing buying/selling costs.

On small examples, these intuitions can often be used to determine the distance by examination; on larger systems a solver for linear programs is usually necessary. For this, the definition below can be transformed into a rational linear program, as we use in the proof of Lemma 13. For further discussions of these distances, see for example [54, 44]

▶ **Definition 3.** *For X a set, and* PMet_X *the set of pseudometric spaces on X, we define* $\Gamma \colon \mathsf{PMet}_X \to \mathsf{PMet}_X$ *by*

$$\Gamma(d)(x,y) = \begin{cases} 1, & \text{if } l(x) \neq l(y), \\ \sup_{h \colon (X,d) \to ([0,1],d_e)} \tau(x) \cdot h - \tau(y) \cdot h, & o.w. \end{cases}$$

Then we define $\mathsf{bd} := \mathrm{lfp}(\Gamma)$.

Note that the least fixed point exists, because PMet_X is a complete lattice, and Γ preserves the pointwise order on PMet_X.

▶ **Example 4.** Consider the following LMC:

Note that $\mathsf{bd}(x_1,y_1) = \mathsf{bd}(x_2,y_2) = 0$ and $\mathsf{bd}(x_1,y_2) = \mathsf{bd}(x_2,y_1) = 1$, which are the values given by $\Gamma(\bot)$. The value $\mathsf{bd}(x,y)$ is then $\Gamma^2(\bot)(x,y)$ for which it can be shown that the supremum is achieved by the map $h_0(z) = \textbf{if } z \in \{x_1,y_1\} \textbf{ then } 1 \textbf{ else } 0$ so that:

$$\begin{aligned} \Gamma^2(\bot)(x,y) &= \sup_{h \colon (X,\Gamma(\bot)) \to ([0,1],d_e)} \tau(x) \cdot h - \tau(y) \cdot h \\ &= \tau(x) \cdot h_0 - \tau(y) \cdot h_0 \\ &= \left(\frac{1}{2} \cdot 1 + \frac{1}{2} \cdot 0 \right) - \left(\frac{2}{5} \cdot 1 + \frac{3}{5} \cdot 0 \right) = \frac{1}{10} \end{aligned}$$

In terms of transportation costs, this $\frac{1}{10}$ can be seen as the cost of transporting the $\frac{1}{10}$ of mass from x_1 to y_2 which can not be transported to y_1 due to its lack of capacity. Intuition based on profits is harder to give in this case, due to the asymmetry of our definition. It is easier to obtain the h in this example by seeing it as grouping the states into "equivalence" classes, spaced as far apart as possible while respecting the distance on X (in this case $\Gamma(\bot)$).

It will be important for the correspondence results of later sections that the behavioural distance can be obtained as a countable supremum, namely the supremum over all finite applications of Γ to the constant zero distance. A similar result for LMCs with non-determinism is shown already in [17, Sec. 3]. It can also be proved using the Kleene fixpoint theorem, or ω-(co)continuity of Γ as shown in [53].

▶ **Proposition 5.** *For any LMC (X, L, τ, l) and $x, y \in X$, we have*

$$\mathsf{bd}(x, y) = \sup_{i < \omega} \Gamma^i(\bot)(x, y)$$

3   Proof System

In this section, we define our derivation system for lower bounds on behavioural distances between states of an LMC. The conclusion of the rules are of the form $x \#_\varepsilon y$ which, as our soundness result will show, implies that $\mathsf{bd}(x, y) \geq \varepsilon$, i.e., the behavioural distance between x and y is at least ε. The definition of bd suggests two rules, one for each case. The label case straightforwardly yields the rule

$$\frac{l(x) \neq l(y)}{x \#_1 y} \text{ (lab)}$$

In the supremum case, $\mathsf{bd}(x, y)$ can be bounded from below by $\tau(x) \cdot h - \tau(y) \cdot h$ for any non-expansive map h by definition. However, it is not immediately clear that this can be done inductively, as we can not assume bd to be known, and thus can not use it to choose a non-expansive h. Fortunately, as long as the system is sound with respect to the behavioural distance, it suffices to have a map h for which a kind of *pairwise non-expansiveness* holds: for any x', y' we have $|h(x') - h(y')| \leq \varepsilon$ for some ε such that we have proved $x' \#_\varepsilon y'$. Soundness then implies that $|h(x') - h(y')| \leq \varepsilon \leq \mathsf{bd}(x', y')$ for all x', y', which is exactly non-expansiveness of h with respect to bd.

Now, in proofs, we could allow arbitrary recursive proofs and require the choice of a pairwise non-expansive map to correctly apply the rule. Alternatively, we can choose to allow arbitrary maps. We are then required to prove that for all $x', y' \in X$, $|h(x') - h(y')|$ is a lower bound on the behavioural distance. We can see the corresponding proof obligations $x' \#_{|h(x')-h(y')|} y'$ as those generated by a chosen map h. The first option fits with a forward reasoning approach to constructing a proof; we prove some bounds and try to find a fitting h. The second is a backward approach; if we wish to show a bound $x \#_\varepsilon y$, we must supply an h and recursively prove its validity.

We choose the latter approach, primarily because it makes the proof obligations clearer, and we will be able to see when a choice of map is not valid. Using the earlier form, a chosen h may be invalid because we have not proved strong enough bounds, or because it is simply not non-expansive with respect to bd. Such a rule can be written as follows:

$$\frac{h \colon X \to [0, 1] \qquad \forall x', y' \in X. \, x' \#_{|h(x')-h(y')|} y' \qquad \tau(x) \cdot h - \tau(y) \cdot h \geq \varepsilon}{x \#_\varepsilon y}$$

However, in this form, the rule does not give a usable derivation system. Namely, in case an LMC has an infinite state space, there will be infinitely many recursive proof obligations. We will ensure that this is always finite, and even reduce the work required to construct proofs beyond this. Further, the map h is so far *real*-valued. This poses a problem both for our approximate completeness result, in which we need to be able to compute these maps, and the construction from proofs to modal formulas whose interpretation will be rational-valued. Summarising, to remedy these issues, we adapt the above rule in the following ways:

■ we show that h need only be defined on a finite subset of the state space thereby generating only finitely many proof obligations;

■ we restrict the codomain of h to rationals;

■ we reduce the number of recursive proof obligations further by not requiring proofs for those bounds which follow from reflexivity and symmetry.

As we show in Theorem 14, the *(lab)* and *(exp)* rules together already yield a system which is (approximately) complete. For the construction of a proof corresponding to a modal formula (Theorem 18), we also require an additional *weakening* rule, essentially to handle formulas using the shift operator ($\varphi \ominus q$, see Definition 15). To make our proof system and its presentation more pleasant, we include two more (derivable) rules: a *zero* rule; and a *symmetry* rule. The three additional rules are inspired by those from quantitative equational theories [40]. Together, this brings us to the following system:

▶ **Definition 6.** *Let* (X, L, τ, l) *be an LMC,* $x, y \in X$, *and* $\varepsilon \in [0, 1]_\mathbb{Q}$. *Further, we define*
$\mathcal{S}_{x,y} := \{s \in X \mid \tau(x)(s) \neq \tau(y)(s)\}$ *and* $\mu \cdot_{\mathcal{S}_{x,y}} h := \sum_{s \in \mathcal{S}_{x,y}} \mu(s) \cdot h(s)$.
  Then, we define the following derivation rules:

$$\frac{}{x \#_0 y} \text{ (zero)} \qquad \frac{y \#_\varepsilon x}{x \#_\varepsilon y} \text{ (symm)} \qquad \frac{x \#_{\varepsilon'} y \quad \varepsilon \leq \varepsilon'}{x \#_\varepsilon y} \text{ (weak)} \qquad \frac{l(x) \neq l(y)}{x \#_1 y} \text{ (lab)}$$

$$\frac{h \colon \mathcal{S}_{x,y} \to [0,1]_\mathbb{Q}}{\frac{\forall x', y' \in \mathcal{S}_{x,y}.\, h(x') > h(y') \implies x' \#_{h(x')-h(y')} y' \qquad \tau(x) \cdot_{\mathcal{S}_{x,y}} h - \tau(y) \cdot_{\mathcal{S}_{x,y}} h \geq \varepsilon}{x \#_\varepsilon y}} \text{ (exp)}$$

We may drop the subscripts x, y *and* $\mathcal{S}$ *whenever clear from the context. We write* $\mathfrak{T}_X$ *for the smallest set which contains all instances of the* (zero) *and* (lab) *rules for* $x, y \in X$, *and is closed under applications of all instances of* (symm), (weak), *and* (exp) *for any* $x, y \in X$ *and* $\varepsilon \in [0, 1]_\mathbb{Q}$.

Note that we write conditions as premises in the rules rather than as separate side-conditions for convenience. As suggested by the definition of $\mathfrak{T}_X$, this makes both the *(zero)* and *(lab)* rules axioms. Further, strictly speaking, *(exp)* is a family of rules, indexed by the maps h. The set $\mathfrak{T}_X$ can be thought of as the set of all proof trees which can be built from the given rules. As is usual, we will write $\vdash x \#_\varepsilon y$ to mean that the given judgment is provable, i.e., there is a proof tree in $\mathfrak{T}_X$ with the given judgment at the root.

▶ **Remark 7.** Note that the restriction to $\mathcal{S}$ in *(exp)* means proofs, which are finite depth by definition, will also be finite breadth even when the LMC under consideration has an infinite state space. This is because $\{s \in X \mid \tau(x)(s) \neq \tau(y)(s)\}$ is a subset of $\text{supp}(\tau(x)) \cup \text{supp}(\tau(y))$, which is finite by assumption. We will see later that this allows us to provide witnesses as both finite proof trees and finite modal formulas. This improves on the earlier work of [44] which restricts to finite state spaces. We illustrate this improvement in Example 25, once we have shown soundness and completeness of the proof system, and its correspondence with modal formulas.

▶ **Example 8.** We continue with the LMC from Example 4 and show how we can prove the distance between x and y shown there as a lower bound. Using the *(lab)* rule, the bounds $u \#_1 v$ can be proved for $u, v \in \{x_1, y_1\}$ and $v \in \{x_2, y_2\}$. This allows us to define $h_0 \colon \mathcal{S} \to [0,1]_\mathbb{Q}$ as before by $h_0(z) = \textbf{if } z \in \{x_1, y_1\} \textbf{ then } 1 \textbf{ else } 0$ for which $\tau(x) \cdot h_0 - \tau(y) \cdot h_0 = \frac{1}{10}$ so that we can prove

$$\frac{\dfrac{}{x_1 \#_1 x_2} \quad \dfrac{}{x_1 \#_1 y_2} \quad \dfrac{}{y_1 \#_1 x_2} \quad \dfrac{}{y_1 \#_1 y_2}}{x \#_{\frac{1}{10}} y}$$

▶ **Example 9.** Our second example serves to illustrate a limitation of our proof system, namely that the behavioural distance of states will not always be exactly provable in our system. It would only be provable if we allowed infinite depth proof trees. The LMC we consider is the same as the one in [44, Thm. 17], which shows that there is an LMC containing states for which it is not possible to give a single formula "explaining" their distance.
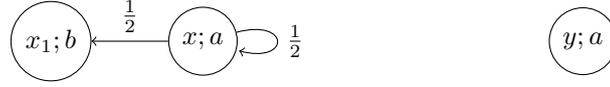
As is discussed in *op. cit.*, the distance $\mathsf{bd}(x, y) = 1$ is reached only in the limit, not by any $\Gamma^i(\bot)$ and thus not by any single tree. Proving the bound given by $\Gamma^i(\bot)$ (for $i > 0$) can be done using $i - 1$ applications of the *(exp)* rule together with two applications each of the *(lab)* and *(zero)* rules. For example, once we have proved $x_1 \#_1 u$ for $u \in \{x, y\}$, we can take the map $h_0 \colon x, y \mapsto 0, x_1 \mapsto 1$ for which $\tau(x) \cdot h_0 - \tau(y) \cdot h_0 = \frac{1}{2}$ and prove:

$$\frac{\overline{\rule{0pt}{1em} x_1 \#_1 x} \quad \overline{\rule{0pt}{1em} x_1 \#_1 y}}{x \#_{\frac{1}{2}} y}$$

This step (plus an application of *(symm)*) allows the next application of *(exp)* with a non-expansive h_0 mapping x to $\frac{1}{2}$, yielding a bound of $\frac{3}{4}$. Continuing to increase the value of $h_0(x)$ in this way, we approach $\mathsf{bd}(x, y)$ from below.

3.1   Soundness

We now move on to showing soundness of the system, i.e., that if we can prove $x \#_\varepsilon y$, then $\mathsf{bd}(x, y) \geq \varepsilon$. The *(zero)* rule is sound as our pseudometrics are valued in $[0, 1]$ and thus 0 is always a sound lower bound. Similarly, the behavioural distance is symmetric, so that the order of states does not change a lower bound's validity. Our weakening rule is sound by transitivity of $\leq$. Soundness of the label rule follows from the definition of Γ. This is similar for the expectation rule, as we discussed at the beginning of this section. We make the intuition given there precise in the coming results.

Due to our restriction of the domain of the map in the *(exp)* rule, we will require the following lemma in the soundness proof:

▶ **Lemma 10.** *For (X, L, τ, l) an LMC, $d \colon X \times X \to [0, 1]$ a pseudometric and $x, y \in X$:*

$$\sup_{h \colon (X, d) \to ([0,1], d_e)} \tau(x) \cdot h - \tau(y) \cdot h = \sup_{h \colon (\mathcal{S}, d|_{\mathcal{S}}) \to ([0,1], d_e)} \tau(x) \cdot_{\mathcal{S}} h - \tau(y) \cdot_{\mathcal{S}} h$$

where $d|_{\mathcal{S}} = d \circ (\iota_{\mathcal{S}} \times \iota_{\mathcal{S}})$ with $\iota_{\mathcal{S}} \colon \mathcal{S} \hookrightarrow X$ the inclusion map.

▶ Remark 11. We can further restrict the space of possible maps h in the above, by seeing the supremum as the optimal solution of a linear program as follows. We encode functions $h \colon \mathcal{S} \to [0, 1]$ as (finite) vectors $\vec{h} \in [0, 1]^{|\mathcal{S}|}$, writing $\vec{h}_x$ for the value of $\vec{h}$ at the position indexed by $x \in \mathcal{S}$. Then, each inequality $|h(x) - h(y)| \leq d(x, y)$ can be expressed by $\vec{a} \cdot \vec{h} \leq d(x, y)$ and $\vec{a}' \cdot \vec{h} \leq d(x, y)$ with $\vec{a}_x = 1, \vec{a}_y = -1, \vec{a}'_x = -1, \vec{a}'_y = 1$ (and all other entries zero). We can enforce $0 \leq \vec{h}_x \leq 1$ similarly. We are thus interested in the problem of maximising $\tau(x) \cdot h - \tau(y) \cdot h$ (a linear expression) subject to the constraints expressed by $\mathbf{A} \cdot \vec{h} \leq \vec{b}$ for an integer matrix $\mathbf{A}$ and vector $\vec{b}$.

The feasible region of this problem is a polytope, which we call $H^d_{x,y}$. It is convex and closed due to the shape of the constraints, and bounded due to the restriction to the unit interval. It is furthermore non-empty as any constant map valued in $[0, 1]$ lies within it. It is known (see, e.g., [50]) that for such a feasible region, the optimal value is achieved in the vertices of the polytope, which we denote by $V(H^d_{x,y})$. Then we can write

$$\sup_{h \colon (X, d) \to ([0,1], d_e)} \tau(x) \cdot h - \tau(y) \cdot h = \max_{h \in V(H^d_{x,y})} \tau(x) \cdot_{\mathcal{S}} h - \tau(y) \cdot_{\mathcal{S}} h$$

The supremum becomes a maximum because finitely many linear inequalities define a polytope with finitely many vertices.

Obtaining a map h in which the supremum is achieved can thus be done using an algorithm such as simplex. We will apply this recursively for $d = \Gamma^i(\bot)$ (the finite approximants of bd), to construct proof trees in our proof of approximate completeness in the next section.

We are now able to show (by structural induction) that any proof in our system yields a lower bound on bd.

▶ **Theorem 12** (Soundness). *For any LMC (X, L, τ, l), any proof tree built from the rules of Definition 6, any $\varepsilon \in [0, 1]_\mathbb{Q}$, and any $x, y \in X$, if the proof tree has $x \#_\varepsilon y$ at the root, then $\mathsf{bd}(x, y) \geq \varepsilon$.*

3.2   Approximate Completeness

The rest of this section is dedicated to proving the approximate completeness of the system. This will show that we can prove lower bounds arbitrarily close to the "true" value given by bd. Our proof relies on the fact that we can get arbitrarily close to bd with its finite approximants $\Gamma^i(\bot)$, and the following lemma, that shows how we can construct proofs exhibiting these finite approximants as lower bounds on the behavioural distance.

▶ **Lemma 13.** *For any $i \in \mathbb{N}$ and $x, y \in X$, we have $\vdash x \#_{\Gamma^i(\bot)(x,y)} y$.*

Proof. By induction on i, where we strengthen the induction by showing that $\Gamma^i(\bot)(x, y)$ is always rational. For the base case, we have $\Gamma^0(\bot)(x, y) = \bot(x, y) = 0$ which is rational, and we can prove $x \#_0 y$ using the *(zero)* rule.

Now let $i \in \mathbb{N}$, and suppose for any $x, y \in X$, that we can prove $x \#_{\Gamma^i(\bot)(x,y)} y$ and that $\Gamma^i(\bot)(x, y)$ is rational. We have

$$\Gamma^{i+1}(\bot)(x, y) = \Gamma(\Gamma^i(\bot))(x, y)$$

$$= \begin{cases} 1, & \text{if } l(x) \neq l(y), \\ \sup_{h \colon (X, \Gamma^i(\bot)) \to ([0,1], d_e)} \tau(x) \cdot h - \tau(y) \cdot h, & \text{o.w.} \end{cases}$$

If $l(x) \neq l(y)$, we can prove $x \#_1 y$ using the *(lab)* rule. Otherwise, by Lemma 10, we have

$$\sup_{h \colon (X, \Gamma^i(\bot)) \to ([0,1], d_e)} \tau(x) \cdot h - \tau(y) \cdot h = \sup_{h \colon (\mathcal{S}, \Gamma^i(\bot)|_\mathcal{S}) \to ([0,1], d_e)} \tau(x) \cdot_\mathcal{S} h - \tau(y) \cdot_\mathcal{S} h$$

As explained in Remark 11, we can find an optimal h_0 by solving a linear program. This will in particular be a map $h_0 \colon (\mathcal{S}, \Gamma^i(\bot)|_\mathcal{S}) \to ([0,1]_\mathbb{Q}, d_e)$ (note the restriction to rationals) because all the coefficients in the problem are rational, by induction. We thus have

$$\sup_{h \colon (\mathcal{S}, \Gamma^i(\bot)|_\mathcal{S}) \to ([0,1], d_e)} \tau(x) \cdot_\mathcal{S} h - \tau(y) \cdot_\mathcal{S} h = \tau(x) \cdot_\mathcal{S} h_0 - \tau(y) \cdot_\mathcal{S} h_0$$

which is rational, because the transition probabilities given by $\tau(x)$ and $\tau(y)$ are also rational. We can now construct the following proof, in which recursive proofs are given by induction, and the above discussion allows us to choose $\varepsilon := \Gamma^{i+1}(\bot)(x, y)$:

$$\frac{\begin{array}{cc} h_0 \colon \mathcal{S} \to [0,1]_\mathbb{Q} \\ \forall x', y' \in \mathcal{S}.\, h_0(x') > h_0(y') \implies x' \#_{h_0(x') - h_0(y')} y' & \tau(x) \cdot h_0 - \tau(y) \cdot h_0 \geq \varepsilon \end{array}}{x \#_\varepsilon y} \quad ◀$$

We see that completeness in this sense only requires the use of the *(zero)*, *(lab)*, and *(exp)* rules. In fact, it can be shown for only the latter two, with the *(zero)* rule being essentially an instance of the *(exp)* rule where the map h is taken to be constant. The approximate completeness of the system is now a simple consequence of the above lemma and Proposition 5, with no further applications of the rules needed.

▶ **Theorem 14** (Approximate Completeness). *For any (real) $\delta > 0$, and any $x, y \in X$ there is a proof tree with $x \mathrel{\#_\varepsilon} y$ at the root, so that $0 \le \mathsf{bd}(x, y) - \varepsilon < \delta$.*

4 Logic

The previous sections have given us a way to inductively derive lower bounds on the behavioural distance between states of an LMC, and shown soundness and approximate completeness of the proof system with respect to the behavioural distance bd. In this sense, the system gives finite evidence or *witnesses* for (lower bounds on) behavioural distances.

Another approach to giving such evidence is to construct formulas in some logic which, in the terminology of [44], "explain" the difference. Ideally, given states $x, y \in X$, this would be a formula φ such that $|[\![\varphi]\!](x) - [\![\varphi]\!](y)| = \mathsf{bd}(x, y)$, i.e., the difference in interpretations of φ on the states is exactly equal to their behavioural distance. However, as is shown in *op. cit.*, such a formula can not be given in general (cf. Example 9). Instead, a construction is given of formulas corresponding to finite approximations of bd. In our notation, these are formulas φ such that $|[\![\varphi]\!](x) - [\![\varphi]\!](y)| = \Gamma^i(\bot)(x, y)$ (for some $i \in \mathbb{N}$). As discussed in Section 2, bd can be obtained as the countable limit of these approximations, so that the construction of [44] gives formulas explaining the behavioural distance of states up to an arbitrarily small error.

In this section, we give analogous constructions between proofs and formulas in the same logic used to characterise the behavioural distance bd in [44]. We start by recalling this logic and its interpretation on LMCs. Its semantics is given in terms of a real-valued interpretation function (first suggested in [38]) and is a slight variation of the logic studied in relation to behavioural distances in [16]. We then move onto the related constructions, the first of which is a straightforward inductive construction of a proof that the distance $|[\![\varphi]\!](x) - [\![\varphi]\!](y)|$ is a lower bound. The second, again inductively, constructs a formula witnessing some proved lower bound. This is based on constructions in [44] and relies on a non-trivial lemma in the case where the distance of states arises from the supremum case in the definition of Γ.

▶ **Definition 15.** *Define the syntax of the logic $\mathcal{L}$ by the following grammar:*

$$\varphi ::= a \mid \bigcirc\varphi \mid \neg\varphi \mid \varphi \ominus q \mid \varphi \vee \varphi$$

where $a \in L$ and $q \in [0, 1]_{\mathbb{Q}}$. Further, given an LMC (X, L, τ, l), the quantitative semantics of $\mathcal{L}$ is given by the interpretation function $[\![\cdot]\!] : \mathcal{L} \to X \to [0, 1]_{\mathbb{Q}}$ defined recursively by the following equations:

$$[\![a]\!](x) = \begin{cases} 1, & \text{if } l(x) = a, \\ 0, & \text{o.w.} \end{cases} \qquad [\![\varphi \ominus q]\!](x) = \max(0, [\![\varphi]\!](x) - q)$$

$$[\![\bigcirc\varphi]\!](x) = \tau(x) \cdot [\![\varphi]\!] \qquad\qquad [\![\varphi \vee \psi]\!](x) = \max([\![\varphi]\!](x), [\![\psi]\!](x))$$

$$[\![\neg\varphi]\!](x) = 1 - [\![\varphi]\!](x)$$

▶ **Remark 16.** From the connectives in the logic $\mathcal{L}$, it is possible to define also $\wedge$ and $\oplus$ as $\varphi \wedge \psi := \neg(\neg\varphi \vee \neg\psi)$ and $\varphi \oplus q := \neg(\neg\varphi \ominus q)$, which then have the expected semantics. We also write false for the formula $a \ominus 1$ whose interpretation is everywhere zero.

▶ **Example 17.** Consider the LMC from Example 9, and the formulas $\varphi_i := \bigcirc^i b$ for $i \in \mathbb{N}$. We can show that $[\![\varphi_i]\!](x_1) = 1$ for any i, so that $[\![\varphi_i]\!](x) = \sum_{n=1}^{i} \left(\frac{1}{2}\right)^i$, while $[\![\varphi_i]\!](y) = 0$. The formula φ_i captures the probability of reaching a state with label b after i steps.

The logic and its interpretation induce new distances between states, namely the difference in interpretations $|[\![\varphi]\!](x) - [\![\varphi]\!](y)|$. Our first correspondence result shows that this distance can be shown to be a lower bound on $\mathsf{bd}(x, y)$ by a proof in our system.

▶ **Theorem 18.** *For any LMC (X, L, τ, l), formula $\varphi \in \mathcal{L}$, and $x, y \in X$, there exists a proof tree with $x \#_\varepsilon y$ as its root, where $\varepsilon = |[\![\varphi]\!](x) - [\![\varphi]\!](y)|$.*

Proof. By induction on the structure of φ, where we write $\mathfrak{T}(\psi, x, y)$ for the proof tree constructed for a formula ψ and states x, y. We show only the *(exp)* case here.

We thus assume that $\varphi = \bigcirc\psi$. Then $\varepsilon = |\tau(x) \cdot [\![\psi]\!] - \tau(y) \cdot [\![\psi]\!]|$ and by induction, we have for all $x', y' \in \mathcal{S}$ with $[\![\psi]\!](x') > [\![\psi]\!](y')$ trees $\mathfrak{T}(\psi, x', y')$. Now we must distinguish two cases. If $\tau(x) \cdot [\![\psi]\!] \geq \tau(y) \cdot [\![\psi]\!]$, we take $h = [\![\psi]\!]|_{\mathcal{S}}$ and construct

$$\frac{[\![\psi]\!]|_{\mathcal{S}} \colon \mathcal{S} \to [0,1]_{\mathbb{Q}} \qquad \{\mathfrak{T}(\psi, x', y') \mid h(x') > h(y')\} \qquad \varepsilon = \tau(x) \cdot [\![\psi]\!] - \tau(y) \cdot [\![\psi]\!]}{x \#_\varepsilon y} \text{ (exp)}$$

Otherwise, we take $h = [\![\neg\psi]\!]|_{\mathcal{S}}$, and have

$$\frac{[\![\neg\psi]\!]|_{\mathcal{S}} \colon \mathcal{S} \to [0,1]_{\mathbb{Q}} \qquad \{\mathfrak{T}(\psi, x', y') \mid h(x') > h(y')\} \qquad \begin{array}{c}\varepsilon = \tau(x) \cdot [\![\neg\psi]\!] - \tau(y) \cdot [\![\neg\psi]\!] \\ = \tau(y) \cdot [\![\varphi]\!] - \tau(x) \cdot [\![\varphi]\!]\end{array}}{x \#_\varepsilon y} \text{ (exp)} \blacktriangleleft$$

▶ **Remark 19.** The depth of the constructed proof matches the modal depth of the formula, i.e., the maximum number of nested $\bigcirc$ modalities. In this sense, the proof does not grow unexpectedly compared to the formula we start with. Due to the branching in the *(exp)* rule, the number of rules we apply will be larger than the number of operators in a given formula in general. For an example, recall the LMC from Example 4, and consider the formula $\bigcirc a$. This has just two operators, but the proof generated by the above procedure will contain five recursive proof obligations generated by $[\![\bigcirc a]\!]$ in the application of *(exp)*, each requiring at least one rule application.

5 Constructing formulas from proofs

To complete the correspondence between proofs and modal formulas, in this section we provide a construction going from a proof of $x \#_\varepsilon y$, to a formula φ such that $|[\![\varphi]\!](x) - [\![\varphi]\!](y)| = \varepsilon$. In fact, we construct formulas whose interpretation on x is equal to the lower bound, and whose interpretation on y is zero. This is required for our recursive construction.

Our construction is inspired by that of [44], which relies on the following lemma:

▶ **Lemma 20.** *Let $f \colon X \to [0, 1]$. If for any $x, y \in X$, we have a function $g_{xy} \colon X \to [0, 1]$ such that $g_{xy}(x) = f(x)$ and $g_{xy}(y) = f(y)$ then $f = \max_x \min_y g_{xy} = \min_x \max_y g_{xy}$.*

A proof of a more general (continuous) version of this result can be found in [1, Lemma A7.2], where it is used in the proof of the Stone-Weierstrass Theorem.

The usefulness of this lemma is perhaps not very intuitive; the idea is that given an application of the *(exp)* rule, we may recursively assume formulas $\varphi_{x'y'}$ capturing the distance between successor states x', y', i.e., $\varphi_{x'y'}(x') = h(x') - h(y')$ and $\varphi_{x'y'}(y') = 0$. We thus have formulas whose interpretations match the function h, but only for pairs of states, so that a lemma similar to the above is required to combine them into a single formula.

Note that, due to the form of our *(exp)* rule, we can not assume in an inductive proof that formulas are given for all successors; only those x', y' for which $h(x') > h(y')$. It is possible to recover all other pairs using the *(zero)* and *(symm)* rules, but we wish to keep the formulas as small as we can. For this, we prove the following stronger version of Lemma 20.

▶ **Lemma 21.** *Let $f \colon X \to [0, 1]$. If for any $x, y \in X$ such that $f(x) \geq f(y)$, we have a function $g_{xy} \colon X \to [0, 1]$ such that: **1.** $g_{xy}(x) = f(x)$ **2.** $g_{xy}(y) = f(y)$ **3.** $\forall z \in X. g_{xy}(z) \geq f(y)$ **4.** $\forall z \in X. g_{xx}(z) = f(x)$, then $f = \max_x \min_{y: f(x) \geq f(y)} g_{xy}$.*

The proof is a little involved and not very informative, however it allows us to construct (potentially) smaller formulas witnessing distances as compared to [44]. This is because we do not require formulas distinguishing all pairs of reachable states in the construction of the next theorem. We discuss the improvement in size at the end of this section, and illustrate it in Examples 24 and 25.

▶ **Theorem 22.** *For any LMC (X, L, τ, l), any $\varepsilon \in [0, 1]_{\mathbb{Q}}$, and any $x, y \in X$, if we have a proof of $x \#_{\varepsilon} y$ using the rules of Definition 6, then there is a formula $\varphi_{xy} \in \mathcal{L}$ such that $[\![\varphi_{xy}]\!](x) = \varepsilon$ and $[\![\varphi_{xy}]\!](y) = 0$.*

Note that we abuse notation, and use $x \#_{\varepsilon} y$ to refer to both a judgment in a proof, and a proof tree with this judgment at the root.

Proof. This is by induction on the structure of the proof. We show only the case of the *(exp)* rule. Suppose we have a proof of the form

$$\frac{h \colon \mathcal{S} \to [0, 1]_{\mathbb{Q}} \qquad\qquad \forall x', y' \in \mathcal{S}.\, h(x') > h(y') \implies x' \#_{h(x')-h(y')} y' \qquad \tau(x) \cdot h - \tau(y) \cdot h \geq \varepsilon}{x \#_{\varepsilon} y}$$

By induction, we have formulas $\varphi_{x'y'}$ such that $[\![\varphi_{x'y'}]\!](x') = h(x') - h(y')$ and $[\![\varphi_{x'y'}]\!](y') = 0$ for all $x', y' \in \mathcal{S}$ with $h(x') > h(y')$. For those x', y' such that $h(x') = h(y')$ we define $\varphi_{x'y'} := \mathsf{false}$ (any formula which is everywhere zero can be used). Using these we construct, for x', y' such that $h(x') \geq h(y')$, the formulas $\psi_{x'y'}^{h} := \varphi_{x'y'} \oplus h(y')$.

We now claim that the interpretations $[\![\psi_{x'y'}^{h}]\!]$ satisfy the conditions of Lemma 21. The first two clearly hold. For the third, note that $[\![\varphi_{x'y'}]\!](z) \geq 0$ for any z, so that indeed

$$[\![\psi_{x'y'}^{h}]\!](z) = [\![\varphi_{x'y'}]\!](z) \oplus h(y') \geq 0 \oplus h(y') = h(y')$$

For the fourth, we see that for any $z \in X$:

$$[\![\psi_{x'x'}^{h}]\!](z) = [\![\mathsf{false} \oplus h(x')]\!](z) = h(x')$$

We now define

$$\varphi_{xy}^{h} := \bigvee_{x'} \left[\left[\bigwedge_{y':h(x')>h(y')} \psi_{x'y'}^{h} \right] \wedge (\mathsf{false} \oplus h(x')) \right]$$

This has the same interpretation as $\bigvee_{x'} \bigwedge_{y':h(x') \geq h(y')} \psi_{x'y'}^{h}$, because for pairs (x', y') with $h(x') = h(y')$, the formula $\psi_{x'y'}^{h}$ will be equal to $\mathsf{false} \oplus h(x')$ by definition. Note also that these formulas are finite, as we quantify over $\mathcal{S}$. Thus letting $\varphi_{xy} := \bigcirc \varphi_{xy}^{h} \ominus (\tau(y) \cdot h)$, yields a (finitary) formula with the desired property.   ◀

One may wonder why we have constructed the formulas φ_{xy}^{h} as a conjunction over a disjunction, and not vice versa. It turns out that this order matters in the case of our *(exp)* rule, as the following example shows.

▶ **Example 23.** Consider the following LMC:

For this example, we can prove the bound $x_0 \#_{\frac{1}{2}} y_0$ using the *(exp)* rule and the map $h(x) = \mathbf{if}\ x = x_1\ \mathbf{then}\ 0\ \text{else}\ 1$, defined for $x \in \{x_1, x_2, y_1, y_2\}$. The required recursive proofs generated by this h are $u \#_1 x_1$ for $u \in \{x_2, y_1, y_2\}$, which can all be proved using the *(lab)* rule. Constructing the $\psi^h_{x'y'}$ for the pairs occurring in these recursive proofs yields

$$\psi^h_{x_2 x_1} = b \oplus 0 \quad \psi^h_{y_1 x_1} = c \oplus 0 \quad \psi^h_{y_2 x_1} = b \oplus 0$$

We may now try to construct φ^h_{xy} as $\bigwedge_{x'} \bigvee_{y:h(x) \geq h(y)} \psi^h_{x'y'}$. In the example, this gives $\varphi^h_{x_0 y_0} \equiv \mathsf{false}$, i.e., the interpretation is zero everywhere thereby not matching the map h. This also shows that we can not interchange the max and min in Lemma 21.

Size of constructed formulas As discussed in the introduction, our constructions together yield an algorithm going from an approximation $\Gamma^i(\bot)(x, y)$ of the behavioural distance of states, via a proof tree, to a formula φ_{xy}. This is an alternative to the construction given as an algorithm in [44, Sec. 7]. In the worst case, this procedure will yield formulas whose size is exponential in the size of the corresponding LMC and the depth i of the approximation. We thus achieve the same asymptotic size complexity as the construction of *op. cit.*, however there are large classes of examples for which the optimisations in our proof system lead to smaller formulas (when counting the total number of connectives). The first example, taken from *op. cit.*, will show a notable improvement in size, and will allow us to discuss the shapes of LMCs leading to these improvements. Our final example applies to an LMC modelling random walks on the natural numbers. This is an infinite state example, which we are still able to capture as it is finitely branching.

▶ **Example 24.** We will compare the size of formulas obtained via our construction with those obtained in an example of [44, Sec. 5]. The LMC involved can be represented as follows:

It can be computed that $\Gamma^3(\bot)(x_0, y_0) = \frac{1}{8}$. The construction applied in Theorem 14 gives the map $h_0\colon x_1, y_1 \mapsto 0, x_2, y_2 \mapsto 1$ for which $\tau(x_0) \cdot h_0 - \tau(y_0) \cdot h_0 = \frac{1}{8}$ so that we have:

$$\frac{\begin{array}{c}\vdots\\\hline x_2 \#_1 x_1\end{array} \quad \begin{array}{c}\vdots\\\hline x_2 \#_1 y_1\end{array} \quad \begin{array}{c}\vdots\\\hline y_2 \#_1 x_1\end{array} \quad \begin{array}{c}\vdots\\\hline y_2 \#_1 y_1\end{array}}{x_0 \#_{\frac{1}{8}} y_0}$$

The recursive lower bounds, given by $\Gamma^2(\bot)$, all have the same proof tree, up to renaming of states. For $x_2 \#_1 y_1$, we get $h_0\colon x_4 \mapsto 1, y_3 \mapsto 0$ giving $\tau(x_4) \cdot h_0 - \tau(y_3) \cdot h_0 = 1$ so that:

$$\frac{\overline{x_4 \#_1 y_3}}{x_2 \#_1 y_1}$$

The formulas generated from such proofs are

$$\varphi_{x_2 x_1} = \varphi_{x_2 x_1} = \varphi_{x_2 x_1} = \varphi_{x_2 x_1} = \bigcirc[[(a \oplus 0) \wedge (\mathsf{false} \oplus 1)] \vee [(\mathsf{false} \oplus 0)]] \ominus 0$$

Putting everything together, the formula $\varphi_{x_0 y_0}$ is

$$\varphi_{x_0 y_0} = \bigcirc[[(\varphi_{x_2 x_1} \oplus 0) \wedge (\varphi_{x_2 y_1} \oplus 0) \wedge (\mathsf{false} \oplus 1)]\vee$$
$$[(\varphi_{y_2 x_1} \oplus 0) \wedge (\varphi_{y_2 y_1} \oplus 0) \wedge (\mathsf{false} \oplus 1)]\vee$$
$$[(\mathsf{false} \oplus 0)] \vee [(\mathsf{false} \oplus 0)]] \ominus \frac{3}{8}$$

This has 8 recursive subformulas compared to the 100 occurring in the formula constructed in [44] and could all be written out within around 5 lines. Clearly, the formula is still not minimal; it can be simplified to $\bigcirc(\bigcirc a)) \ominus \frac{3}{8}$. However, we see a clear improvement in size.

The main features which allow us to achieve such an improvement in the size of formulas witnessing lower bounds are: the number of states reachable at each step being less than the size of the entire state space; and those successors having non-zero behavioural distance so that the map h takes many different values. The combination of restricting to supports and omitting symmetric pairs from the recursive proof obligations in the *(exp)* rule, gives smaller proofs in these cases, which in turn are transformed into smaller formulas.

▶ **Example 25.** We finish with an infinite state example based on random walks on the natural numbers. We model this as an LMC with state space $\mathbb{N}$ and transitions $\tau(n) = \frac{1}{2}|n-1\rangle + \frac{1}{2}|n+1\rangle$ for $n > 0$ and $\tau(0) = 1 \cdot |0\rangle$. Further, we have labels $\{a, b\}$ and labelling function $l(n) = \mathbf{if}\ n = 0\ \mathbf{then}\ b\ \mathbf{else}\ a$.

States $n < m$ can clearly be distinguished by the probability to reach the state 0 with unique label b in n steps. In fact, this turns out to completely determine the distance between states. For example, $\Gamma^5(\bot)(4, 6) = \frac{1}{2^4}$. This corresponds to the interpretation of the formula $\bigcirc^4 b$ on these states. In general, we have for $n < m$ and $i > n$, $\Gamma^i(\bot)(n, m) = \frac{1}{2^n}$. We will show the proof constructed for one such bound, as well as the formula constructed from this.

We have $\Gamma^3(\bot)(2, 3) = \frac{1}{4}$ which can be proved as a lower bound using the map $h_0 \colon 1 \mapsto \frac{1}{2}, 3 \mapsto 0, 2 \mapsto 0, 4 \mapsto 0$ for which $\tau(2) \cdot h_0 - \tau(3) \cdot h_0 = \frac{1}{4}$ as follows

$$\frac{\dfrac{\vdots}{1 \mathrel{\#_{\frac{1}{2}}} 2} \qquad \dfrac{\vdots}{1 \mathrel{\#_{\frac{1}{2}}} 3} \qquad \dfrac{\vdots}{1 \mathrel{\#_{\frac{1}{2}}} 4}}{2 \mathrel{\#_{\frac{1}{4}}} 3}$$

The recursive proofs are all essentially the same, we show only the one for $1 \mathrel{\#_{\frac{1}{2}}} 2$, which uses the map $h_0 \colon 0 \mapsto 1, 2 \mapsto 0, 1 \mapsto 0, 3 \mapsto 0$ yielding $\tau(1) \cdot h_0 - \tau(2) \cdot h_0 = \frac{1}{2}$ and

$$\frac{\dfrac{\vdots}{0 \mathrel{\#_1} 1} \qquad \dfrac{\vdots}{0 \mathrel{\#_1} 2} \qquad \dfrac{\vdots}{0 \mathrel{\#_1} 3}}{1 \mathrel{\#_{\frac{1}{2}}} 2}$$

For these recursive proofs, the corresponding formulas are

$$\varphi_{12} = \varphi_{13} = \varphi_{14} = \bigcirc[[(b \oplus 0) \wedge (b \oplus 0) \wedge (b \oplus 0) \wedge (\mathsf{false} \oplus 1)]\vee$$
$$[\mathsf{false} \oplus 0] \vee [\mathsf{false} \oplus 0] \vee [\mathsf{false} \oplus 0]] \ominus 0$$

From these, we obtain

$$\varphi_{23} = \bigcirc\left[\left[(\varphi_{12} \oplus 0) \wedge (\varphi_{13} \oplus 0) \wedge (\varphi_{14} \oplus 0) \wedge \left(\mathsf{false} \oplus \frac{1}{2}\right)\right] \vee \right.$$
$$[\mathsf{false} \oplus 0] \vee [\mathsf{false} \oplus 0] \vee [\mathsf{false} \oplus 0]] \ominus 0$$

This can be simplified to $\bigcirc[\bigcirc b \wedge (\mathsf{false} \oplus \frac{1}{2})]$, which is not in general equivalent to $\bigcirc \bigcirc b$, but does have the required property. We thus obtain evidence for differences in the behaviour of states even in a system with an infinite state space.

6   Conclusions and Future Work

We have given a derivation system for lower bounds on behavioural distances between states in labelled Markov chains, with proofs of soundness and approximate completeness with respect to a least fixed point definition of the behavioural distance. The choice of the definition based on non-expansive maps was made specifically to allow the definition of a proof system, with the commonly used alternative definition based on couplings not immediately yielding a proof principle for lower bounds. The definitions are equivalent, by Kantorovich-Rubinstein duality [31]. This duality arises more generally when defining equivalences and their apartness counterparts via liftings, as was noted in earlier work on *behavioural apartness* [51]. We further showed a close correspondence between proofs in our system and formulas in a modal logic, and compared this to the constructions in [44] going between finite approximations of distances and formulas in the same logic. We see quite some avenues for future work, and sketch some of them here.

Definitions of behavioural distances have been given for a variety of other system types, both metric and probabilistic, e.g.: Metric LTSs [52]; and Markov decision processes with finite [21] and infinite state spaces [22]. A natural extension would be to generalise our results in case we change the system type while keeping a similar definition of distance between distributions; however we may also consider adapted notions of distance, such as the total variation distance studied for LMCs in [12] or the MICo distance on MDPs [10]. Further statistical metrics/divergences such as the Lévy-Prokhorov metric [43] or Kullback-Leibler divergence [39] may also be of interest. The former has recently [19] been shown to define a functional on pseudometric spaces, whose greatest fixed point is the ε-distance [18].

A more general option is to take a coalgebraic view and use the definition of codensity lifting [32, 48] and its suitability for capturing quantitative notions of equivalence to provide a sound and complete derivation system for many of these systems at once. Existing work on corresponding expressive logics [35] then gives a starting point for providing a general version of the construction in Section 5. Another approach in the same vein is to use the theory of Kantorovich functors [27], used to obtain characteristic logics also in the quantitative setting. We would also like to investigate the connection to strategies in quantitative bisimulation games studied in [36, 59] and developed for codensity bisimulations in [34].

The efficient computation of proofs and distinguishing formulas would also be an interesting extension. There are a number of works in the qualitative setting (beyond the already discussed [44]) which could provide inspiration. For LTSs and branching bisimulation, computing (minimal) distinguishing formulas has been investigated by Martens and Groote [41, 42]. König, Mika-Michalski and Schröder use coalgebraic techniques to develop algorithms for computing strategies in bisimulation games and transforming these into distinguishing formulas [37]. Wißmann, Milius and Schröder give a coalgebraic algorithm related to partition refinement which constructs modal formulas characterising behavioural equivalence classes [60].

An alternative approach to improving the robustness of probabilistic bisimilarity, is the use of approximate bisimulations, such as: ε-bisimilarity [18]; ε-APB [20]; and ε-lumpability [9]. These are often close to existing qualitative definitions, with some degree of error introduced. For a recent overview, and extension to weak and branching bisimulation, see [47]. It would be interesting to compare these approximate notions to distances and relate them to proofs and logics.

References

1 Robert B Ash. *Real Analysis and Probability: Probability and Mathematical Statistics: A Series of Monographs and Textbooks.* Academic press, 1972.

2 Giorgio Bacci, Giovanni Bacci, Kim G. Larsen, and Radu Mardare. On-the-fly computation of bisimilarity distances. *Log. Methods Comput. Sci.*, 13(2), 2017. `doi:10.23638/LMCS-13(2:13)2017`.

3 Giorgio Bacci, Giovanni Bacci, Kim G. Larsen, and Radu Mardare. A complete quantitative deduction system for the bisimilarity distance on Markov chains. *Log. Methods Comput. Sci.*, 14(4), 2018. `doi:10.23638/LMCS-14(4:15)2018`.

4 Giorgio Bacci, Giovanni Bacci, Kim G. Larsen, Radu Mardare, Qiyi Tang, and Franck van Breugel. Computing probabilistic bisimilarity distances for probabilistic automata. *Log. Methods Comput. Sci.*, 17(1), 2021. URL: `https://lmcs.episciences.org/7147`.

5 Paolo Baldan, Filippo Bonchi, Henning Kerstan, and Barbara König. Coalgebraic behavioral metrics. *Log. Methods Comput. Sci.*, 14(3), 2018. `doi:10.23638/LMCS-14(3:20)2018`.

6 Paolo Baldan, Richard Eggert, Barbara König, and Tommaso Padoan. Fixpoint theory - upside down. *Log. Methods Comput. Sci.*, 19(2), 2023. `doi:10.46298/LMCS-19(2:15)2023`.

7 Filippo Bonchi, Barbara König, and Daniela Petrisan. Up-to techniques for behavioural metrics via fibrations. *Math. Struct. Comput. Sci.*, 33(4-5):182–221, 2023. `doi:10.1017/S0960129523000166`.

8 Filippo Bonchi, Daniela Petrisan, Damien Pous, and Jurriaan Rot. A general account of coinduction up-to. *Acta Informatica*, 54(2):127–190, 2017. `doi:10.1007/S00236-016-0271-4`.

9 Peter Buchholz. Exact and ordinary lumpability in finite Markov chains. *Journal of applied probability*, 31(1):59–75, 1994.

10 Pablo Samuel Castro, Tyler Kastner, Prakash Panangaden, and Mark Rowland. Mico: Improved representations via sampling-based state similarity for markov decision processes. In *NeurIPS*, pages 30113–30126, 2021. URL: `https://proceedings.neurips.cc/paper/2021/hash/fd06b8ea02fe5b1c2496fe1700e9d16c-Abstract.html`.

11 Di Chen, Franck van Breugel, and James Worrell. On the complexity of computing probabilistic bisimilarity. In *FoSSaCS*, volume 7213 of *Lecture Notes in Computer Science*, pages 437–451. Springer, 2012. `doi:10.1007/978-3-642-28729-9_29`.

12 Taolue Chen and Stefan Kiefer. On the total variation distance of labelled Markov chains. In *CSL-LICS*, pages 33:1–33:10. ACM, 2014. `doi:10.1145/2603088.2603099`.

13 Josée Desharnais, Abbas Edalat, and Prakash Panangaden. A logical characterization of bisimulation for labeled Markov processes. In *LICS*, pages 478–487. IEEE Computer Society, 1998. `doi:10.1109/LICS.1998.705681`.

14 Josée Desharnais, Abbas Edalat, and Prakash Panangaden. Bisimulation for labelled Markov processes. *Inf. Comput.*, 179(2):163–193, 2002. `doi:10.1006/INCO.2001.2962`.

15 Josée Desharnais, Vineet Gupta, Radha Jagadeesan, and Prakash Panangaden. Metrics for labeled Markov systems. In *CONCUR*, volume 1664 of *Lecture Notes in Computer Science*, pages 258–273. Springer, 1999. `doi:10.1007/3-540-48320-9_19`.

16 Josée Desharnais, Vineet Gupta, Radha Jagadeesan, and Prakash Panangaden. Metrics for labelled Markov processes. *Theor. Comput. Sci.*, 318(3):323–354, 2004. `doi:10.1016/J.TCS.2003.09.013`.

17 Josée Desharnais, Radha Jagadeesan, Vineet Gupta, and Prakash Panangaden. The metric analogue of weak bisimulation for probabilistic processes. In *LICS*, pages 413–422. IEEE Computer Society, 2002. `doi:10.1109/LICS.2002.1029849`.

18 Josée Desharnais, François Laviolette, and Mathieu Tracol. Approximate analysis of probabilistic processes: Logic, simulation and games. In *QEST*, pages 264–273. IEEE Computer Society, 2008. `doi:10.1109/QEST.2008.42`.

19 Josée Desharnais and Ana Sokolova. ϵ-distance via lévy-prokhorov lifting. *CoRR*, abs/2507.10732, 2025. `doi:10.48550/arXiv.2507.10732`.

20   Alessandro D'Innocenzo, Alessandro Abate, and Joost-Pieter Katoen. Robust PCTL model checking. In *HSCC*, pages 275–286. ACM, 2012. `doi:10.1145/2185632.2185673`.

21   Norm Ferns, Prakash Panangaden, and Doina Precup. Metrics for finite Markov decision processes. In *AAAI*, pages 950–951. AAAI Press/The MIT Press, 2004. URL: `http://www.aaai.org/Library/AAAI/2004/aaai04-124.php`.

22   Norm Ferns, Prakash Panangaden, and Doina Precup. Metrics for Markov decision processes with infinite state spaces. In *UAI*, pages 201–208. AUAI Press, 2005. URL: `https://dslpitt.org/uai/displayArticleDetails.jsp?mmnu=1&smnu=2&article_id=1175&proceeding_id=21`.

23   Nathanaël Fijalkow, Bartek Klin, and Prakash Panangaden. Expressiveness of probabilistic modal logics, revisited. In *ICALP*, volume 80 of *LIPIcs*, pages 105:1–105:12. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2017. `doi:10.4230/LIPIcs.ICALP.2017.105`.

24   Herman Geuvers. Apartness and distinguishing formulas in Hennessy-Milner logic. In *A Journey from Process Algebra via Timed Automata to Model Learning*, volume 13560 of *Lecture Notes in Computer Science*, pages 266–282. Springer, 2022. `doi:10.1007/978-3-031-15629-8_14`.

25   Herman Geuvers and Bart Jacobs. Relating apartness and bisimulation. *Log. Methods Comput. Sci.*, 17(3), 2021. `doi:10.46298/LMCS-17(3:15)2021`.

26   Alessandro Giacalone, Chi-Chang Jou, and Scott A. Smolka. Algebraic reasoning for probabilistic concurrent systems. In *Programming Concepts and Methods*, pages 443–458. North-Holland, 1990.

27   Sergey Goncharov, Dirk Hofmann, Pedro Nora, Lutz Schröder, and Paul Wild. Kantorovich functors and characteristic logics for behavioural distances. In *FoSSaCS*, volume 13992 of *Lecture Notes in Computer Science*, pages 46–67. Springer, 2023. `doi:10.1007/978-3-031-30829-1_3`.

28   Claudio Hermida and Bart Jacobs. Structural induction and coinduction in a fibrational setting. *Inf. Comput.*, 145(2):107–152, 1998. `doi:10.1006/INCO.1998.2725`.

29   Arend Heyting. *Intuitionism: an introduction*, volume 41. Elsevier, 1966.

30   Bart Jacobs. Structured probabilistic reasoning. Unpublished draft. URL: `http://www.cs.ru.nl/B.Jacobs/PAPERS/ProbabilisticReasoning.pdf`.

31   Leonid Vasilevich Kantorovich and SG Rubinshtein. On a space of totally additive functions. *Vestnik of the St. Petersburg University: Mathematics*, 13(7):52–59, 1958.

32   Shin-ya Katsumata, Tetsuya Sato, and Tarmo Uustalu. Codensity lifting of monads and its dual. *Log. Methods Comput. Sci.*, 14(4), 2018. `doi:10.23638/LMCS-14(4:6)2018`.

33   Stephen C. Kleene. *Introduction to Metamathematics*. D. van Nostrand, Princeton, New Jersey, 1952.

34   Yuichi Komorida, Shin-ya Katsumata, Nick Hu, Bartek Klin, Samuel Humeau, Clovis Eberhart, and Ichiro Hasuo. Codensity games for bisimilarity. *New Gener. Comput.*, 40(2):403–465, 2022. `doi:10.1007/S00354-022-00186-Y`.

35   Yuichi Komorida, Shin-ya Katsumata, Clemens Kupke, Jurriaan Rot, and Ichiro Hasuo. Expressivity of quantitative modal logics: Categorical foundations via codensity and approximation. In *LICS*, pages 1–14. IEEE, 2021. `doi:10.1109/LICS52264.2021.9470656`.

36   Barbara König and Christina Mika-Michalski. (Metric) bisimulation games and real-valued modal logics for coalgebras. In *CONCUR*, volume 118 of *LIPIcs*, pages 37:1–37:17. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2018. `doi:10.4230/LIPIcs.CONCUR.2018.37`.

37   Barbara König, Christina Mika-Michalski, and Lutz Schröder. Explaining non-bisimilarity in a coalgebraic approach: Games and distinguishing formulas. In *CMCS*, volume 12094 of *Lecture Notes in Computer Science*, pages 133–154. Springer, 2020. `doi:10.1007/978-3-030-57201-3_8`.

38   Dexter Kozen. A probabilistic PDL. *J. Comput. Syst. Sci.*, 30(2):162–178, 1985. `doi:10.1016/0022-0000(85)90012-1`.

39   S. Kullback and R. A. Leibler. On Information and Sufficiency. *The Annals of Mathematical Statistics*, 22(1):79–86, 1951. `doi:10.1214/aoms/1177729694`.

40   Radu Mardare, Prakash Panangaden, and Gordon D. Plotkin. Quantitative algebraic reasoning. In *LICS*, pages 700–709. ACM, 2016. `doi:10.1145/2933575.2934518`.

41   Jan Martens and Jan Friso Groote. Computing minimal distinguishing Hennessy-Milner formulas is NP-hard, but variants are tractable. In *CONCUR*, volume 279 of *LIPIcs*, pages 32:1–32:17. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2023. `doi:10.4230/LIPIcs.CONCUR.2023.32`.

42   Jan Martens and Jan Friso Groote. Minimal depth distinguishing formulas without until for branching bisimulation. In *Logics and Type Systems in Theory and Practice*, volume 14560 of *Lecture Notes in Computer Science*, pages 188–202. Springer, 2024. `doi:10.1007/978-3-031-61716-4_12`.

43   Yu V Prokhorov. Convergence of random processes and limit theorems in probability theory. *Theory of Probability & Its Applications*, 1(2):157–214, 1956.

44   Amgad Rady and Franck van Breugel. Explainability of probabilistic bisimilarity distances for labelled Markov chains. In *FoSSaCS*, volume 13992 of *Lecture Notes in Computer Science*, pages 285–307. Springer, 2023. `doi:10.1007/978-3-031-30829-1_14`.

45   Wojciech Rozowski. A complete quantitative axiomatisation of behavioural distance of regular expressions. In *ICALP*, volume 297 of *LIPIcs*, pages 149:1–149:20. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2024. `doi:10.4230/LIPIcs.ICALP.2024.149`.

46   Jan J. M. M. Rutten. Relators and metric bisimulations. In *CMCS*, volume 11 of *Electronic Notes in Theoretical Computer Science*, pages 252–258. Elsevier, 1998. `doi:10.1016/S1571-0661(04)00063-5`.

47   Timm Spork, Christel Baier, Joost-Pieter Katoen, Jakob Piribauer, and Tim Quatmann. A spectrum of approximate probabilistic bisimulations. In *CONCUR*, volume 311 of *LIPIcs*, pages 37:1–37:19. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2024. `doi:10.4230/LIPIcs.CONCUR.2024.37`.

48   David Sprunger, Shin-ya Katsumata, Jérémy Dubut, and Ichiro Hasuo. Fibrational bisimulations and quantitative reasoning: Extended version. *J. Log. Comput.*, 31(6):1526–1559, 2021. `doi:10.1093/LOGCOM/EXAB051`.

49   Qiyi Tang. *Computing Probabilistic Bisimilarity Distances*. PhD thesis, York University, Toronto, 2018.

50   Kathleen Trustrum. *Linear programming*. Springer, 1971.

51   Ruben Turkenburg, Harsh Beohar, Clemens Kupke, and Jurriaan Rot. Proving behavioural apartness. In *CMCS*, volume 14617 of *Lecture Notes in Computer Science*, pages 156–173. Springer, 2024. `doi:10.1007/978-3-031-66438-0_8`.

52   Franck van Breugel. A behavioural pseudometric for metric labelled transition systems. In *CONCUR*, volume 3653 of *Lecture Notes in Computer Science*, pages 141–155. Springer, 2005. `doi:10.1007/11539452_14`.

53   Franck van Breugel. On behavioural pseudometrics and closure ordinals. *Inf. Process. Lett.*, 112(19):715–718, 2012. `doi:10.1016/J.IPL.2012.06.019`.

54   Franck van Breugel. Probabilistic bisimilarity distances. *ACM SIGLOG News*, 4(4):33–51, 2017. `doi:10.1145/3157831.3157837`.

55   Franck van Breugel, Claudio Hermida, Michael Makkai, and James Worrell. Recursively defined metric spaces without contraction. *Theor. Comput. Sci.*, 380(1-2):143–163, 2007. `doi:10.1016/J.TCS.2007.02.059`.

56   Franck van Breugel, Babita Sharma, and James Worrell. Approximating a behavioural pseudometric without discount for probabilistic systems. *Log. Methods Comput. Sci.*, 4(2), 2008. `doi:10.2168/LMCS-4(2:2)2008`.

57   Franck van Breugel and James Worrell. A behavioural pseudometric for probabilistic transition systems. *Theor. Comput. Sci.*, 331(1):115–142, 2005. `doi:10.1016/J.TCS.2004.09.035`.

58   Cédric Villani et al. *Optimal transport: old and new*, volume 338. Springer, 2008.

59   Emily Vlasman, Anto Nanah Ji, James Worrell, and Franck van Breugel. Explainability is a game for probabilistic bisimilarity distances. In *CONCUR*, volume 348 of *LIPIcs*, pages 36:1–36:20. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2025. `doi:10.4230/LIPIcs.CONCUR.2025.36`.

60   Thorsten Wißmann, Stefan Milius, and Lutz Schröder. Quasilinear-time computation of generic modal witnesses for behavioural inequivalence. *Log. Methods Comput. Sci.*, 18(4), 2022. `doi:10.46298/LMCS-18(4:6)2022`.

A   Omitted Proofs

▶ **Lemma 10.** *For (X, L, τ, l) an LMC, $d\colon X \times X \to [0,1]$ a pseudometric and $x, y \in X$:*

$$\sup_{h\colon (X,d)\to([0,1],d_e)} \tau(x) \cdot h - \tau(y) \cdot h = \sup_{h\colon (\mathcal{S},d|_{\mathcal{S}})\to([0,1],d_e)} \tau(x) \cdot_{\mathcal{S}} h - \tau(y) \cdot_{\mathcal{S}} h$$

where $d|_{\mathcal{S}} = d \circ (\iota_{\mathcal{S}} \times \iota_{\mathcal{S}})$ with $\iota_{\mathcal{S}}\colon \mathcal{S} \hookrightarrow X$ the inclusion map.

Proof. We prove two inequalities:

$\leq$.   This holds because any $h\colon (X, d) \to ([0,1], d_e)$ restricts to a map $h|_{\mathcal{S}} = h \circ \iota_{\mathcal{S}}\colon (\mathcal{S}, d|_{\mathcal{S}}) \to ([0,1], d_e)$, and we can show that

$$\tau(x) \cdot h - \tau(y) \cdot h = \tau(x) \cdot_{\mathcal{S}} h|_{\mathcal{S}} - \tau(y) \cdot_{\mathcal{S}} h|_{\mathcal{S}}$$

$\geq$.   For this direction, we show that for any $h\colon (\mathcal{S}, d|_{\mathcal{S}}) \to ([0,1], d_e)$, there is an $h'\colon (X, d) \to ([0,1], d_e)$ such that

$$\tau(x) \cdot h' - \tau(y) \cdot h' = \tau(x) \cdot_{\mathcal{S}} h - \tau(y) \cdot_{\mathcal{S}} h$$

We use an existing construction of extensions of non-expansive maps, to extend h along the inclusion $\iota_{\mathcal{S}}\colon \mathcal{S} \hookrightarrow X$. Namely, we define $h'(x) := \inf_{z \in \mathcal{S}} h(z) \oplus d(x, z)$, where $\oplus$ is truncated addition on the unit interval. This is an extension in the sense that $h' \circ \iota_{\mathcal{S}} = h$, so that the above equality indeed holds.   ◀

▶ **Theorem 12** (Soundness). *For any LMC (X, L, τ, l), any proof tree built from the rules of Definition 6, any $\varepsilon \in [0,1]_{\mathbb{Q}}$, and any $x, y \in X$, if the proof tree has $x \#_\varepsilon y$ at the root, then $\mathsf{bd}(x,y) \geq \varepsilon$.*

Proof. We proceed by induction on the structure of the proof tree.

Case *(zero)*:   By its definition, bd takes values in $[0,1]$, so that $\mathsf{bd}(x,y) \geq 0$ holds.

Case *(label)*:   We have a proof tree

$$\frac{l(x) \neq l(y)}{x \#_1 y}$$

By definition of Γ, we must have $\mathsf{bd}(x,y) = 1$, so that indeed $\mathsf{bd}(x,y) \geq 1$.

Case *(symm)*:   We have a proof tree

$$\frac{y \#_\varepsilon x}{x \#_\varepsilon y}$$

By induction, we have $\mathsf{bd}(y,x) \geq \varepsilon$, but bd is symmetric, so that also $\mathsf{bd}(x,y) \geq \varepsilon$.

Case *(weak)*:   We have a proof tree

$$\frac{x \#_{\varepsilon'} y \qquad \varepsilon \leq \varepsilon'}{x \#_\varepsilon y}$$

By induction, we have $\mathsf{bd}(x,y) \geq \varepsilon' \geq \varepsilon$.

Case *(exp)*:   We have a proof tree

$$\frac{\begin{array}{c} h \colon \mathcal{S} \to [0,1]_{\mathbb{Q}} \\ \forall x', y' \in \mathcal{S}.\, h(x') > h(y') \implies x' \#_{h(x')-h(y')} y' \qquad \tau(x) \cdot h - \tau(y) \cdot h \geq \varepsilon \end{array}}{x \#_\varepsilon y} \text{(exp)}$$

By induction, we have for all $x', y' \in \mathcal{S}$ with $h(x') > h(y')$ that $\mathsf{bd}(x', y') \geq |h(x') - h(y')|$. For $x', y' \in \mathcal{S}$ with $h(x') < h(y')$, we have $\mathsf{bd}(x', y') = \mathsf{bd}(y', x') \geq |h(y') - h(x')| = |h(x') - h(y')|$. For the remaining pairs, $|h(x') - h(y')| = 0 \leq \mathsf{bd}(x', y')$. In other words, $h \colon \mathcal{S} \to [0,1]_{\mathbb{Q}}$ is a non-expansive map $h \colon (\mathcal{S}, \mathsf{bd}) \to ([0,1]_{\mathbb{Q}}, d_e)$. As bd is defined as the least fixed point of Γ, we have

$$\mathsf{bd}(x,y) = \Gamma(\mathsf{bd})(x,y)$$

$$= \begin{cases} 1, & \text{if } l(x) \neq l(y), \\ \sup_{h \colon (X, \mathsf{bd}) \to ([0,1], d_e)} \tau(x) \cdot h - \tau(y) \cdot h, & \text{o.w.} \end{cases}$$

In case $l(x) \neq l(y)$, we have $\mathsf{bd}(x,y) = 1 \geq \varepsilon$.

In the remaining case, we have

$$\mathsf{bd}(x,y) = \sup_{k \colon (X, \mathsf{bd}) \to ([0,1], d_e)} \tau(x) \cdot k - \tau(y) \cdot k$$

$$= \sup_{h \colon (\mathcal{S}, d|_{\mathcal{S}}) \to ([0,1], d_e)} \tau(x) \cdot_{\mathcal{S}} h - \tau(y) \cdot_{\mathcal{S}} h$$

$$\geq \tau(x) \cdot h - \tau(y) \cdot h \geq \varepsilon$$

where the second equality is shown in Lemma 10 and the first inequality holds because h is one of the non-expansive maps ranged over in the sup. This covers all cases, so soundness follows by induction.   ◀

▶ **Theorem 14** (Approximate Completeness). *For any (real) $\delta > 0$, and any $x, y \in X$ there is a proof tree with $x \#_\varepsilon y$ at the root, so that $0 \leq \mathsf{bd}(x,y) - \varepsilon < \delta$.*

Proof. Let $\delta > 0$. By Proposition 5, we have $\mathsf{bd}(x,y) = \sup_{i < \omega} \Gamma^i(\bot)(x,y)$, so there exists $i \in \mathbb{N}$ such that

$$0 \leq \mathsf{bd}(x,y) - \Gamma^i(\bot)(x,y) < \delta$$

Lemma 13 exactly gives us a proof of $x \#_{\Gamma^i(\bot)(x,y)} y$, and we are done.   ◀

▶ **Theorem 18.** *For any LMC (X, L, τ, l), formula $\varphi \in \mathcal{L}$, and $x, y \in X$, there exists a proof tree with $x \#_\varepsilon y$ as its root, where $\varepsilon = |[\![\varphi]\!](x) - [\![\varphi]\!](y)|$.*

Proof. By induction on the structure of φ, where we write $\mathcal{T}(\psi, x, y)$ for the proof tree constructed for a formula ψ and states x, y.

Case $\varphi = a$:   We have $\varepsilon = 0$ or $\varepsilon = 1$ with the following proofs:

$$\frac{}{x \#_0 y} \text{(zero)} \qquad \frac{l(x) \neq l(y)}{x \#_1 y} \text{(lab)}$$

Case $\varphi = \neg\psi$: We have $|[\![\varphi]\!](x) - [\![\varphi]\!](y)| = |[\![\psi]\!](x) - [\![\psi]\!](y)|$ so simply let $\mathcal{T}(\varphi, x, y) = \mathcal{T}(\psi, x, y)$.

Case $\varphi = \psi \ominus q$: In this case we will have $|[\![\psi]\!](x) - [\![\psi]\!](y)| \leq |[\![\varphi]\!](x) - [\![\varphi]\!](y)|$ (truncation may give an inequality) so that we have

$$\frac{x \#_{|[\![\varphi]\!](x) - [\![\varphi]\!](y)|} y \qquad |[\![\psi]\!](x) - [\![\psi]\!](y)| \leq |[\![\varphi]\!](x) - [\![\varphi]\!](y)|}{x \#_{|[\![\psi]\!](x) - [\![\psi]\!](y)|} y} \text{ (weak)}$$

Case $\varphi = \varphi_1 \vee \varphi_2$: We have

$$
\begin{aligned}
\varepsilon &= |[\![\varphi_1 \vee \varphi_2]\!](x) - [\![\varphi_1 \vee \varphi_2]\!](y)| \\
&= |\max([\![\varphi_1]\!](x), [\![\varphi_2]\!](x)) - \max([\![\varphi_1]\!](y), [\![\varphi_2]\!](y))| \\
&\leq \max(|[\![\varphi_1]\!](x) - [\![\varphi_1]\!](y)|, |[\![\varphi_2]\!](x) - [\![\varphi_2]\!](y)|)
\end{aligned}
$$

so that we take $\mathcal{T}(\varphi, x, y)$ to be

$$\frac{\mathcal{T}(\varphi_i, x, y) \qquad \varepsilon \leq \varepsilon'}{x \#_\varepsilon y} \text{ (weak)}$$

with φ_i the formula yielding the above maximum, which we have called ε'.

Case $\varphi = \bigcirc\psi$: See the main text. ◀

▶ **Lemma 21.** *Let $f \colon X \to [0,1]$. If for any $x, y \in X$ such that $f(x) \geq f(y)$, we have a function $g_{xy} \colon X \to [0,1]$ such that:* **1.** $g_{xy}(x) = f(x)$ **2.** $g_{xy}(y) = f(y)$ **3.** $\forall z \in X. \, g_{xy}(z) \geq f(y)$ **4.** $\forall z \in X. \, g_{xx}(z) = f(x)$, *then $f = \max_x \min_{y : f(x) \geq f(y)} g_{xy}$.*

Proof. We first define

$$k_{xy} = \begin{cases} g_{xy} & \text{if } f(x) \geq f(y) \\ g_{yx} & \text{if } f(x) < f(y) \end{cases}$$

Note that these k_{xy} satisfy the conditions of Lemma 20 so that $f = \max_x \min_y k_{xy}$. It thus suffices to prove that

$$\max_x \min_y k_{xy} = \max_x \min_{y : f(x) \geq f(y)} g_{xy}$$

We prove this by proving two inequalities.

$\leq$**.** Consider the inequality and simplify as follows:

$$
\begin{aligned}
&\forall z \in X. \max_x \min_y k_{xy}(z) \leq \max_x \min_{y \,:\, f(x) \geq f(y)} g_{xy}(z) \\
\Longleftrightarrow \; &\forall z \in X. \forall u_1 \in X. \min_y k_{u_1 y}(z) \leq \max_x \min_{y \,:\, f(x) \geq f(y)} g_{xy}(z) \\
\Longleftrightarrow \; &\forall z \in X. \forall u_1 \in X. \exists u_3 \in X. \min_y k_{u_1 y}(z) \leq \min_{y \,:\, f(u_3) \geq f(y)} g_{u_3 y}(z) \\
\Longleftrightarrow \; &\forall z \in X. \forall u_1 \in X. \exists u_3 \in X. \forall u_4 \in X. f(u_3) \geq f(u_4) \implies \min_y k_{u_1 y}(z) \leq g_{u_3 u_4}(z) \\
\Longleftrightarrow \; &\forall z \in X. \forall u_1 \in X. \exists u_3 \in X. \forall u_4 \in X. f(u_3) \geq f(u_4) \implies \exists u_2 \in X. k_{u_1 u_2}(z) \leq g_{u_3 u_4}(z)
\end{aligned}
$$

So, let $z, u_1 \in X$ and take $u_3 = z$. Further, let $u_4 \in X$ such that $f(u_3) \geq f(u_4)$ and take $u_2 = z$. Then

$$
\begin{aligned}
k_{u_1 u_2}(z) &= k_{u_1 z}(z) = f(z) \\
g_{u_3 u_4}(z) &= g_{z u_4}(z) = f(z)
\end{aligned}
$$

This first inequality thus holds.

$\geq$. Consider the inequality and simplify as follows:

$$\forall z \in X. \max_x \min_y k_{xy}(z) \geq \max_x \min_{y:f(x)\geq f(y)} g_{xy}(z)$$

$$\iff \forall z \in X. \forall u_3 \in X. \max_x \min_y k_{xy}(z) \geq \min_{y:f(u_3)\geq f(y)} g_{u_3 y}(z)$$

$$\iff \forall z \in X. \forall u_3 \in X. \exists u_1 \in X. \min_y k_{u_1 y}(z) \geq \min_{y:f(u_3)\geq f(y)} g_{u_3 y}(z)$$

$$\iff \forall z \in X. \forall u_3 \in X. \exists u_1 \in X. \forall u_2 \in X. k_{u_1 u_2}(z) \geq \min_{y:f(u_3)\geq f(y)} g_{u_3 y}(z)$$

$$\iff \forall z \in X. \forall u_3 \in X. \exists u_1 \in X. \forall u_2 \in X. \exists u_4 \in X. f(u_3) \geq f(u_4) \wedge k_{u_1 u_2}(z) \geq g_{u_3 u_4}(z).$$

So, we let $z, u_3 \in X$ and take $u_1 = u_3$. Now let $u_2 \in X$. We distinguish two further cases:

$f(u_2) > f(u_3)$.   Here we take $u_4 = u_3$ and have

$$k_{u_1 u_2}(z) = k_{u_3 u_2}(z) = g_{u_2 u_3}(z) \overset{(3)}{\geq} f(u_3) \overset{(4)}{=} g_{u_3 u_3}(z) = g_{u_3 u_4}(z)$$

$f(u_2) \leq f(u_3)$.   Here we take $u_4 = u_2$ and see that

$$k_{u_1 u_2}(z) = k_{u_3 u_2}(z) = g_{u_3 u_2}(z) = g_{u_3 u_4}(z)$$

This concludes the case distinctions.   ◄

▶ **Theorem 22.** *For any LMC (X, L, τ, l), any $\varepsilon \in [0,1]_{\mathbb{Q}}$, and any $x, y \in X$, if we have a proof of $x \#_\varepsilon y$ using the rules of Definition 6, then there is a formula $\varphi_{xy} \in \mathcal{L}$ such that $[\![\varphi_{xy}]\!](x) = \varepsilon$ and $[\![\varphi_{xy}]\!](y) = 0$.*

Proof. This is by induction on the structure of the proof.

Case (zero):   In this case we take $\varphi_{xy} = \mathsf{false}$. We have $[\![\varphi_{xy}]\!](z) = 0$ for any $z \in X$, yielding the desired interpretations.

Case (lab):   Here, we take $\varphi_{xy} = l(x)$. We must have $[\![l(x)]\!](y) = 0$ as $l(x) \neq l(y)$, so that the interpretations are as required.

Case (symm):   The induction hypothesis gives a formula φ_{yx}. Taking $\varphi_{xy} = \neg\varphi_{yx} \ominus (1-\varepsilon)$, we have $[\![\varphi_{xy}]\!](x) = (1 - [\![\varphi_{yx}]\!](x)) \ominus (1-\varepsilon) = \varepsilon$ and $[\![\varphi_{xy}]\!](y) = (1 - [\![\varphi_{yx}]\!](y)) \ominus (1-\varepsilon) = 0$ as desired.

Case (weak):   The induction hypothesis here gives φ'_{xy} with $[\![\varphi'_{xy}]\!](x) = \varepsilon'$ and $[\![\varphi'_{xy}]\!](y) = 0$ and $\varepsilon' \geq \varepsilon$. This means we can take $\varphi_{xy} = \varphi'_{xy} \ominus (\varepsilon' - \varepsilon)$ and have $[\![\varphi_{xy}]\!](x) = \varepsilon' \ominus (\varepsilon' - \varepsilon) = \varepsilon$ and $[\![\varphi_{xy}]\!](y) = 0 - (\varepsilon' - \varepsilon) = 0$.

Case (exp):   See the main text.   ◄