Generalized bisimulation metrics

Konstantinos Chatzikokolakis, Daniel Gebler, Catuscia Palamidessi, Lili Xu

Generalized bisimulation metrics*

Konstantinos Chatzikokolakis[1,2], Daniel Gebler[3],
Catuscia Palamidessi[4,2], and Lili Xu[2,5]

[1] CNRS
[2] LIX, Ecole Polytechnique
[3] VU University Amsterdam
[4] INRIA
[5] Institute of Software, Chinese Academy of Science

Abstract. The pseudometric based on the Kantorovich lifting is one of the most popular notion of distance between probabilistic processes proposed in the literature. However, its application in verification is limited to linear properties. We propose a generalization which allows to deal with a wider class of properties, such as those used in security and privacy. More precisely, we propose a family of pseudometrics, parametrized on a notion of distance which depends on the property we want to verify. Furthermore, we show that the members of this family still characterize bisimilarity in terms of their kernel, and provide a bound on the corresponding distance between trace distributions. Finally, we study the instance corresponding to differential privacy, and we show that it has a dual form, easier to compute. We also prove that the typical process-algebra constructs are non-expansive, thus paving the way to a modular approach to verification.

1 Introduction

Originally proposed in the seminal works of van Breugel and Worrel [5,4] and of Desharnais et al. [19,20,21], the pseudometric based on the Kantorovich lifting has become very popular in the process algebra community. One reason for its success is that, when dealing with probabilistic processes, distances are more suitable than equivalences, since the latter are not robust wrt small variation of probabilities. Another important reason is that, thanks to the dual presentation of the Kantorovich lifting in terms of the mass transportation problem, the distance can be efficiently computed using linear programming algorithms [4,7,8,2]. Furthermore, this pseudometric is an extension of probabilistic bisimilarity, in the sense that two states have distance distance 0 if and only if they are bisimilar. In fact, this pseudometric also shares with bisimilarity a similar coinductive definition. More precisely, it is defined as the greatest fixpoint of a

* This work has been partially supported by the project ANR-12-IS02-001 PACE, the project ANR-11-IS02-0002 LOCALI, the INRIA Equipe Associée PRINCESS, the INRIA Large Scale Initiative CAPPRIS, and the EU grant 295261 MEALS.

transformation that has the same structure as the one used for bisimilarity.[6] This allows to transfer some of the concepts and methods that have been extensively explored in process algebra, and to use lines of reasoning which the process algebra community is familiar with. Along the same lines, a nice property of this pseudometric is that the standard operators of process algebra are non-expansive wrt it. This generalizes the result that bisimulation is a congruence, and can be used in a similar way, for compositional reasoning and verification.

Last but not least, the Kantorovich bisimilarity metric provides a bound on the corresponding distance on probabilistic traces [13] (corresponding in the sense that the definition is based on the same Kantorovich lifting). This means that it can be used to verify certain probabilistic properties on traces. More specifically, it can be used to verify properties that are expressed in terms of difference between probabilities of sets of traces. These properties are linear, in the sense that the difference increases linearly wrt variations on the distributions.

Many properties, however, such as several privacy and security ones, are not linear. This is the case of the popular property of differential privacy [22], which is expressed in terms of ratios of probabilities. In fact, there are processes that have small Kantorovich distance, and which are not ϵ-differentially private for any finite ϵ. Another example are the properties used in quantitative information flow, which involve logarithmic functions on probabilities.

The purpose of this work is to generalize the Kantorovich lifting to obtain a family of pseudometrics suitable for the verification of a wide class of properties, following the principles that:

i. the members of this family should depend on a parameter related to the class of properties (on traces) that we wish to verify,
ii. each member should provide a bound on the corresponding distance on trace distributions,
iii. the kernel of each member should correspond to probabilistic bisimilarity,
iv. the general construction should be coinductive,
v. the typical process-algebra operators should be non-expansive,
vi. each member should be feasible to compute.

In this paper we have achieved the first four desiderata. Regarding the last two, so far we have studied a particular case (hereafter called multiplicative variant of the Kantorovich lifting) based on the notion of distance used in the definition of differential privacy. We were able to find a dual form of the lifting, which allows to reduce the problem of its computation to a linear optimization problem solvable with standard algorithms. We have also proved that several typical process-algebra operators are non-expansive, and we have given explicitly the expression of the bound. For some of them we were able to prove this result in a general form, i.e., non-expansiveness wrt all the metrics of the family, and with the bound represented by the same expression.

[6] In the original definition the Kantorovich bisimilarity pseudometric was defined as the greatest fixpoint, but such definition requires using the reverse order on metrics. More recently, authors tend to use the natural order, and define the bisimilarity metric as the least fixpoint, see [1,2,13]. Here we follow the latter approach.

As an example of application of our framework, we show how to instantiate our construction to obtain the multiplicative variant of the Kantorovich pseudometric, and how to use it to verify the property of differential privacy.

All proofs are given in the report version of this paper [12].

Related Work Bisimulation metrics based on the standard Kantorovich distance have been used in various applications, such as systems biology [31], games [10], planning [15] and security [9]. We consider in this paper discrete state spaces. Bisimulation metrics on uncountable state spaces have been explored in [21,24,25]. We define bisimulation metrics as fixed point of an appropriate transformation. Alternative characterizations were provided in terms of coalgebras [5,6] and real-valued modal logics [19,21].

Verification of differential privacy has been itself an active area of research. Prominent approaches based on formal methods are those based on type systems [28] and logical formulations [3]. Earlier papers [32,33] defined bisimulation metrics suitable for proving differential privacy, however they suffered from the fact that the respective kernel relations do not fully characterize probabilistic bisimilarity.

2 Preliminaries

2.1 Probabilistic automata

Given a set X, we denote by $Prob(X), Disc(X)$ the set of all and discrete probability measures over X respectively; the support of a measure μ is defined as $supp(\mu) = \{x \in X | \mu(x) > 0\}$. A *Probabilistic automaton* (henceforth PA) $\mathcal{A}$ is a tuple (S, A, D) where S is a countable set of *states*, A is a countable set of action *labels*, and $D \subseteq S \times A \times Disc(S)$ is a *transition relation.* We write $s \xrightarrow{a} \mu$ for $(s, a, \mu) \in D$. We say that $\mathcal{A}$ is *finitely branching* iff $supp(\mu)$ is finite for all $s \xrightarrow{a} \mu$ (note that non-deterministic branching is not constrained).

An *execution* α is a (possibly infinite) sequence $s_0 a_1 s_1 a_2 s_2 \ldots$ of alternating states and labels, such that for each $i : s_i \xrightarrow{a_{i+1}} \mu_{i+1}$ and $\mu_{i+1}(s_{i+1}) > 0$. We use $lstate(\alpha)$ to denote the last state of a finite execution α. We use $Exec^*(\mathcal{A})$ and $Exec(\mathcal{A})$ to represent the set of finite executions and of all executions of $\mathcal{A}$, respectively. A *trace* is a sequence of labels in $A^* \cup A^\omega$ obtained from executions by removing the states. We use $[\,]$ to represent the empty trace, and $^\frown$ to concatenate two traces.

A *fully probabilistic automaton* (henceforth FPA) $\mathcal{A}$ is a PA where from each state of $\mathcal{A}$ there is at most one transition available. We denote by $L(s)$ and $\pi(s)$ the label and distribution of the unique transition starting from s (if any).

In a FPA $\mathcal{A}$, a state s of $\mathcal{A}$ induces a probability measure over traces as follows. The basic measurable events are the cones of finite traces, where the cone of a finite trace $\boldsymbol{t}$, denoted by $C_{\boldsymbol{t}}$, is the set $\{\boldsymbol{t'} \in A^* \cup A^\omega | \boldsymbol{t} \le \boldsymbol{t'}\}$, where $\le$ is the standard prefix preorder on sequences. The probability induced by s on a cone $C_{\boldsymbol{t}}$, denoted by $\Pr[s \rhd C_{\boldsymbol{t}}]$, is defined recursively as follows:

$$\Pr[s \rhd C_{\boldsymbol{t}}] = \begin{cases} 1 & \text{if } \boldsymbol{t} = [\,] \\ 0 & \text{if } \boldsymbol{t} = a^\frown \boldsymbol{t'} \text{ and } a \neq L(s) \\ \sum_{s_i} \mu(s_i) \Pr[s_i \rhd C_{\boldsymbol{t'}}] & \text{if } \boldsymbol{t} = a^\frown \boldsymbol{t'} \text{ and } s \xrightarrow{a} \mu \end{cases} \tag{1}$$

This probability measure is extended to arbitrary measurable sets in the σ-algebra of traces in the standard way. We write $\Pr[s \rhd \sigma]$ to represent the probability induced by s on the measurable set of traces σ.

2.2 Pseudometrics

A pseudometric is a relaxed notion of a normal metric in which distinct elements can have distance zero. We consider here a generalized notion where the distance can also be infinite, and we use $[0, +\infty)$ to denote the non-negative fragment of the real numbers $\mathbb{R}$ enriched with $+\infty$. Formally, an (extended) pseudometric on a set X is a function $m : X^2 \to [0, +\infty)$ with the following properties: $m(x, x) = 0$ (reflexivity), $m(x, y) = m(y, x)$ (symmetry), and $m(x, y) \leq m(x, z) + m(z, y)$ (triangle inequality). A metric has the extra condition that $m(x, y) = 0$ implies $x = y$. Let $\mathcal{M}_X$ denote the set of all pseudometrics on X with the ordering $m_1 \preceq m_2$ iff $\forall x, y . m_1(x, y) \leq m_2(x, y)$. It can be shown that $(\mathcal{M}_X, \preceq)$ is a complete lattice with bottom element $\bot$ such that $\forall x, y. \bot(x, y) = 0$ and top element $\top$ such that $\forall x, y. \top(x, y) = +\infty$.

The *ball* (wrt m) of radius r centered at $x \in X$ is defined as $B^m_r(x) = \{x' \in X : m(x, x') \leq r\}$. A point $x \in X$ is called *isolated* iff there exists $r > 0$ such that $B^m_r(x) = \{x\}$; m is called *discrete* if all points are isolated. The *diameter* (wrt m) of $A \subseteq X$ is defined as $\operatorname{diam}_m(A) = \sup_{x, x' \in A} m(x, x')$. A *geodesic* is a curve on which paths have minimum distance, i.e. a curve $\gamma : I \to X$, where I is an interval of reals, such that $m(\gamma(a), \gamma(b)) = |a - b|$ for all $a, b \in I$. The *kernel* $\ker(m)$ is an equivalence relation on X defined as

$$(x, x') \in \ker(m) \quad \text{iff } m(x, x') = 0$$

3 A general family of Kantorovich liftings

We introduce here a family of liftings from pseudometrics on a set X to pseudometrics on $\mathit{Prob}(X)$. This family is obtained as a generalization of the Kantorovich lifting, in which the Lipschitz condition plays a central role.

Given two pseudometric spaces $(X, d_X), (Y, d_Y)$, we say that $f : X \to Y$ is 1-Lipschitz wrt d_X, d_Y iff $d_Y(f(x), f(x')) \leq d_X(x, x')$ for all $x, x' \in X$. We denote by 1-Lip$[(X, d_X), (Y, d_Y)]$ the set of all such functions.

A function $f : X \to \mathbb{R}$ can be lifted to a function $\hat{f} : \mathit{Prob}(X) \to \mathbb{R}$ by taking its expected value. For discrete distributions (countable X) it can be written as:

$$\hat{f}(\mu) = \textstyle\sum_{x \in X} \mu(x) f(x) \tag{2}$$

while for continuous distributions we need to restrict f to be measurable wrt the corresponding σ-algebra on X, and take $\hat{f}(\mu) = \int f d\mu$.

Given a pseudometric m on X, the *standard Kantorovich lifting* of m is a pseudometric $K(m)$ on $Prob(X)$, defined as:

$$K(m)(\mu, \mu') = \sup\{|\hat{f}(\mu) - \hat{f}(\mu')| : \ f \in 1\text{-Lip}[(X, m), (\mathbb{R}, d_{\mathbb{R}})]\}$$

where $d_{\mathbb{R}}$ denotes the standard metric on reals. For continuous distributions we implicitly take the sup to range over measurable functions.

Generalization. A generalization of the Kantorovich lifting can be naturally obtained by extending the range of f from $(\mathbb{R}, d_{\mathbb{R}})$ to a generic metric space (V, d_V), where $V \subseteq \mathbb{R}$ is a convex subset of the reals[7], and d_V is a metric on V. A function $f : X \to V$ can be lifted to a function $\hat{f} : Prob(X) \to V$ in the same way as before (cfr. (2)); the requirement that V is convex ensures that $\hat{f}(\mu) \in V$.

Then, similarly to the standard case, given a pseudometric space (X, m), we can define a lifted pseudometric $K_V(m)$ on $Prob(X)$ as:

$$K_V(m)(\mu, \mu') = \sup\{d_V(\hat{f}(\mu), \hat{f}(\mu')) : \ f \in 1\text{-Lip}[(X, m)(V, d_V)]\} \qquad (3)$$

The subscript V in K_V is to emphasize the fact that for each choice of (V, d_V) we may get a different lifting. We should also point out the difference between m, the pseudometric on X being lifted, and d_V, the metric (not pseudo) on V which parametrizes the lifting.

The constructed $K_V(m)$ can be shown to be an extended pseudometric for any choice of (V, d_V), i.e. it is non-negative, symmetric, identical elements have distance zero, and it satisfies the triangle inequality. However, without extra conditions, it is not guaranteed to be bounded (even if m itself is bounded). For the purposes of this paper this is not an issue. In the appendix we show that under the condition that d_V is *ball-convex* (i.e. all its balls are convex sets, which holds for all metrics in this paper), the following bound can be obtained:

$$K_V(m)(\mu, \mu') \leq \text{diam}_m(supp(\mu) \cup supp(\mu'))$$

Examples The standard Kantorovich lifting is obtained by taking $(V, d_V) = (\mathbb{R}, d_{\mathbb{R}})$. When 1-bounded pseudometrics are used, like in the construction of the standard bisimilarity metric, then we can equivalently take $V = [0, 1]$.

Moreover, a multiplicative variant of the Kantorovich lifting can be obtained by taking $(V, d_V) = ([0, 1], d_\otimes)$ (or equivalently $([0, \infty), d_\otimes)$) where $d_\otimes(x, y) = |\ln x - \ln y|$. The resulting lifting is discussed in detail in Section 5 and its relation to differential privacy is shown in Section 5.1.

[7] V could be further generalized to be a convex subset of a vector space. It is unclear whether such a generalization would be useful, hence it is left as future work.

4 A general family of bisimilarity pseudometrics

In this section we define a general family of pseudometrics on the states of a PA which have the property of extending probabilistic bisimilarity in the usual sense. Following standard lines, we define a transformation on state pseudometrics by first lifting a state pseudometric to a pseudometric on distributions (over states), using the generalized Kantorovich lifting defined in previous section. Then we apply the standard Hausdorff lifting to obtain a pseudometric on sets of distributions. This last step is to take into account the nondeterminism of the PA, i.e., the fact that in general, from a state, we can make transitions to different distributions. The resulting pseudometric naturally corresponds to a state pseudometric, obtained by associating each set of distributions to the states which originate them. Finally, we define the intended bisimilarity pseudometric as the least fixpoint of this transformation wrt the ordering $\preceq$ on the state pseudometrics (or equivalently, as the greatest fixpoint wrt the reverse of $\preceq$). We recall that $m \preceq m'$ means that $m(s,s') \leq m'(s,s')$ for all $s, s' \in S$.

Let $\mathcal{A} = (S, A, D)$ be a PA, let (V, d_V) be a metric space (for some convex $V \subseteq \mathbb{R}$), and let $\mathcal{M}$ be the set of pseudometrics m on S such that $\mathrm{diam}_m(S) \leq \mathrm{diam}_{d_V}(V)$. Recall that $\inf \emptyset = \mathrm{diam}_{d_V}(V)$ and $\sup \emptyset = 0$.

Definition 1. *The transformation $F_V : \mathcal{M} \to \mathcal{M}$ is defined as follows.*

$$F_V(m)(s,t) = \max_{a \in A}\{ \sup_{s \xrightarrow{a} \mu} \inf_{t \xrightarrow{a} \nu} K_V(m)(\mu,\nu), \sup_{t \xrightarrow{a} \nu} \inf_{s \xrightarrow{a} \mu} K_V(m)(\nu,\mu)\}$$

We can also characterize F_V in terms of the following zigzag formulation:

Proposition 1. *For any $\epsilon \geq 0$, $F_V(m)(s,t) \leq \epsilon$ if and only if:*

- *if $s \xrightarrow{a} \mu$, then there exists ν such that $t \xrightarrow{a} \nu$ and $K_V(m)(\mu,\nu) \leq \epsilon$,*
- *if $t \xrightarrow{a} \nu$, then there exists μ such that $s \xrightarrow{a} \mu$ and $K_V(m)(\nu,\mu) \leq \epsilon$.*

The following result states that K_V and F_V are monotonic wrt $(\mathcal{M}, \preceq)$.

Proposition 2. *Let $m, m' \in \mathcal{M}$. If $m \preceq m'$ then:*

$$F_V(m)(s,s') \leq F_V(m')(s,s') \quad \textit{for all states } s, s'$$
$$K_V(m)(\mu,\mu') \leq K_V(m')(\mu,\mu') \quad \textit{for all distributions } \mu, \mu'$$

Since $(\mathcal{M}, \preceq)$ is a complete lattice and F_V is monotone on $\mathcal{M}$, by Tarski's theorem [30] F_V has a least fixpoint, which coincides with the least pre-fixpoint. We define the *bisimilarity pseudometric* bm_V as this least fixpoint:

Definition 2. *The bisimilarity pseudometric bm_V is defined as:*

$$bm_V = \min\big\{m \in \mathcal{M} \mid F_V(m) = m\big\} = \min\big\{m \in \mathcal{M} \mid F_V(m) \preceq m\big\}$$

In addition, if the states of $\mathcal{A}$ are finite, then the closure ordinal of F_V is ω (cf: [20], Lemma 3.10). Hence we can approximate bm_V by iterating the function F_V from the bottom element:

Proposition 3. *Assume that S is finite. Let $m_0 = \bot$ and $m_{i+1} = F_V(m_i)$. Then $bm_V = \sup_i m_i$.*

Next section shows that bm_V is indeed a bisimilarity metric, in the sense that its kernel coincides with probabilistic bisimilarity.

4.1 Bisimilarity as 0-distance

We now show that under certain conditions, the pseudometric constructed from $K_V(m)$ characterizes bisimilarity at its kernel. Recall that the kernel $\ker(m)$ of m is an equivalence relation relating states at distance 0.

Given an equivalence relation R on S, its lifting $\mathcal{L}(R)$ is an equivalence relation on $Disc(S)$, defined as

$$(\mu, \mu') \in \mathcal{L}(R) \quad \text{iff} \quad \forall s \in S : \mu([s]_R) = \mu'([s]_R)$$

where $[s]_R$ denotes the equivalence class of s wrt R.

To obtain the characterization result we assume that (a) the PA is finitely branching, and (b) there exists a geodesic in (V, d_V). The main result is that, under condition (b), the kernel operator and the lifting operators $\mathcal{L}, K_V$ commute on distributions with finite support.[8] This is then sufficient to obtain the characterization result due to condition (a).

Lemma 1. *If (V, d_V) has a geodesic then $\mathcal{L}(\ker(m))$ and $\ker(K_V(m))$ coincide on all distributions of finite support.*

If S is finite, the same result can be obtained under the weaker condition that (V, d_V) is non-discrete. We also expect the result to be extensible to distributions with infinite support.

We recall the notions of probabilistic bisimulation and bisimilarity, following the formulation in terms of post-fixpoints of a transformation on state relations:

Definition 3.

- *The transformation $B : S \times S \to S \times S$ is defined as: $(s, s') \in B(R)$ iff*
  - *if $s \xrightarrow{a} \mu$, then there exists μ' such that $t \xrightarrow{a} \mu'$ and $(\mu, \mu') \in \mathcal{L}(R)$,*
  - *if $s' \xrightarrow{a} \mu'$, then there exists μ such that $s \xrightarrow{a} \mu$ and $(\mu', \mu) \in \mathcal{L}(R)$.*
- *A relation $R \subseteq S \times S$ is called a bisimulation if it is a post-fixpoint of R, i.e. $R \subseteq B(R)$.*

It is easy to see that B is monotonic on $(2^{S \times S}, \subseteq)$ and that the latter is a complete lattice, hence by Tarski's theorem there exists the greatest fixpoint of B, and it coincides with the greatest bisimulation:

Definition 4. *The bisimilarity relation $\sim \subseteq S \times S$ is defined as:*

$$\sim \; = \; \max\{R \mid R = B(R)\} \; = \; \max\{R \mid R \subseteq B(R)\} \; = \; \bigcup\{R \mid R \subseteq B(R)\}$$

[8] cfr. [17] for the analogous property for the standard Kantorovich lifting.

We are now ready to show the correspondence between pre-fixpoint metrics and bisimulations. Using Lemma 1, we can see that the definition of B corresponds to the characterization of F_V in Proposition 1, for $\epsilon = 0$. Hence we have:

Proposition 4. *Assume a finitely branching PA and that (V, d_V) has a geodesic. For every $m \in \mathcal{M}$, if $F_V(m) \preceq m$ then $\ker(m) \subseteq B(\ker(m))$, i.e., $\ker(m)$ is a bisimulation.*

As a consequence, $\ker(bm_V) \subseteq \sim$. The converse of Proposition 4 does not hold, because the fact that $\ker(m) \subseteq B(\ker(m))$ does not say anything about the effect of F_V on the distance between elements that are not on the kernel. However, in the case of bisimilarity we can make a connection: consider the greatest metric $m_\sim$ whose kernel coincides with bisimilarity, namely, $m_\sim(s, s') = 0$ if $s \sim s'$ and $m_\sim(s, s') = \mathrm{diam}_{d_V}(V)$ otherwise. We have that $F_V(m_\sim) \preceq m_\sim$, and therefore $\sim = \ker(m_\sim) \subseteq bm_V$. Therefore we can conclude that the kernel of the bisimilarity pseudometrics coincides with bisimilarity.

Theorem 1. *Assume a finitely branching PA and that (V, d_V) has a geodesic. Then $\ker(bm_V) = \sim$.*

4.2 Relation with trace distributions

In this section, we show the relation between the bisimilarity metric bm_V and the corresponding metric on traces, in the case of FPAs (fully probabilistic automata). Note that we restrict to the fully probabilistic case here, where probabilities on traces can defined in the way shown in the preliminaries. The full case of PAs can be treated by using schedulers, but a proper treatment involves imposing scheduler restrictions which complicate the formalism. Since these problems are orthogonal to the goals of this paper, we keep the discussion simple by restricting to the fully probabilistic case.

The distance between trace distributions (i.e. distributions over A^ω) will be measured by the Kantorovich lifting of the *discrete metric*. Given (V, d_V), let $\delta_V = \mathrm{diam}_{d_V}(V)$. Then let dm_{δ_V} be the δ_V-valued discrete metric on A^ω, defined as $dm_{\delta_V}(\boldsymbol{t}, \boldsymbol{t'}) = 0$ if $\boldsymbol{t} = \boldsymbol{t'}$, and $dm_{\delta_V}(\boldsymbol{t}, \boldsymbol{t'}) = \delta_V$ otherwise.

Then $K_V(dm_{\delta_V})(\mu, \mu')$ is a pseudometric on $Prob(A^\omega)$, whose kernel coincides with probabilistic trace equivalence.

Proposition 5. *$K_V(dm_{\delta_V})(\mu, \mu') = 0$ iff $\mu(\sigma) = \mu'(\sigma)$ for all measurable $\sigma \subseteq A^\omega$.*

The following theorem expresses that our bisimilarity metric bm_V is a bound on the distance on traces, which extends the standard relation between probabilistic bisimilarity and probabilistic trace equivalence.

Theorem 2. *Let $\mu = \Pr[s \rhd \cdot]$ and $\mu' = \Pr[s' \rhd \cdot]$. Then $K_V(dm_{\delta_V})(\mu, \mu') \leq bm_V(s, s')$.*

	Standard $K(m)(\mu, \mu')$	Multiplicative $K_\otimes(m)(\mu, \mu')$
Primal	$\max_f \lvert \hat{f}(\mu) - \hat{f}(\mu') \rvert$ subject to $\forall s, s'.\ \lvert f(s) - f(s') \rvert \leq m(s, s')$	$\max_f \lvert \ln \hat{f}(\mu) - \ln \hat{f}(\mu') \rvert$ subject to $\forall s, s'.\ \lvert \ln f(s) - \ln f(s') \rvert \leq m(s, s')$
Dual	$\min_\ell \sum_{i,j} \ell_{ij} m(s_i, s_j)$ subject to $\forall i, j.\ \ell_{ij} \geq 0$ $\forall i.\ \sum_j \ell_{ij} = \mu(s_i)$ $\forall j.\ \sum_i \ell_{ij} = \mu'(s_j)$	$\min \ln z$ subject to $\forall i, j.\ \ell_{ij}, r_i \geq 0$ $\forall i. \sum_j \ell_{ij} - r_i = \mu(s_i)$ $\forall j.\ \sum_i \ell_{ij} e^{m(s_i, s_j)} - r_j \leq z \cdot \mu'(s_j)$

Table 1: The standard Kantorovich metric and its multiplicative variant.

It should be noted that, although the choice of $K_V(dm_{\delta_V})$ as our trace distribution metric might seem arbitrary, this metric is in fact of great interest. In the case of the standard bisimilarity pseudometric, i.e. when $(V, d_V) = ([0,1], d_\mathbb{R})$, this metric is equal to the well-known *total variation* distance (also known as *statistical distance*), defined as $tv(\mu, \mu') = \sup_\sigma |\mu(\sigma) - \mu'(\sigma)|$:

$$K(dm_{\delta_V}) \ = \ tv \tag{4}$$

Theorem 2 reduces to the result of [13] relating the total variation distance to the bisimilarity pseudometric. Moreover, in the case of the multiplicative pseudometric, discussed in the next section, $K_V(dm_{\delta_V})$ is the same as the multiplicative distance between distributions, discussed in Section 5.1, which plays a central role in differential privacy.

5 The multiplicative variant

In this section we investigate the multiplicative variant of the Kantorovich pseudometric, obtained by considering as distance d_V the ratio between two numbers instead than their difference. This is the distance used to define differential privacy. We show that this variant has a dual form, which can be used to compute the metric by using linear programming techniques. In the next section, we will show how to use it to verify differential privacy.

Definition 5. *The multiplicative variant $K_\otimes$ of the Kantorovich lifting is defined as the instantiation of K_V with $([0,1], d_\otimes)$ where $d_\otimes(x, y) = |\ln x - \ln y|$.*

It is well known that the standard Kantorovich metric has a dual form which can be interpreted in terms of *the Transportation Problem*, namely, the lowest total cost of transporting the mass of one distribution μ to the other distribution μ' given the distance m between locations (in our case, states). The dual form is shown in Table 1. Note that both the primal and the dual forms are linear optimization problems. The dual form is particularly suitable for computation, via standard linear programming techniques.

For our multiplicative variant, the objective function of the primal form is not a linear expression, hence the linear programming techniques cannot be applied directly. However, since $\ln \hat{f}(\mu) - \ln \hat{f}(\mu') = \ln \hat{f}(\mu)/\hat{f}(\mu')$ and ln is a monotonically increasing function, the primal problem is actually a linear-fractional program. It is known that such kind of program can be converted to an equivalent linear programming problem and then to a dual program. The dual form of the multiplicative variant obtained in this way is shown in Table 1. (For the sake of simplicity, the table shows only the dual form of $\ln \hat{f}(\mu) - \ln \hat{f}(\mu')$. The dual form of $\ln \hat{f}(\mu') - \ln \hat{f}(\mu)$ can be obtained by simply switching the roles of μ and μ'.) Hence, the multiplicative pseudometric can be computed by using linear programming techniques.

Finally, note that the curve $\gamma : [0, a] \to [0, 1]$, for $a > 0$, defined by $\gamma(t) = e^{-t}$ is a geodesic of $([0, 1], d_\otimes)$, since $d_\otimes(\gamma(a), \gamma(b)) = |\ln e^{-a} - \ln e^{-b}| = |a - b|$. Hence, the conditons of Theorem 1 are satisfied, which means that $bm_\otimes$, i.e. the bisimulation metric constructed by $K_\otimes$, characterizes bisimulation at its kernel.

5.1 Application to differential privacy

Differential privacy [22] is a notion of privacy originating from the area of statistical databases, which however has been recently applied to several other areas. The standard context is that of an analyst who wants to perform a statistical query to a database. Although obtaining statistical information is permitted, privacy issues arise when this information can be linked to that of an individual in the database. In order to hide this link, differentially private mechanisms add noise to the outcome of the query, in a way such that databases differing in a single individual have similar probability of producing the same observation.

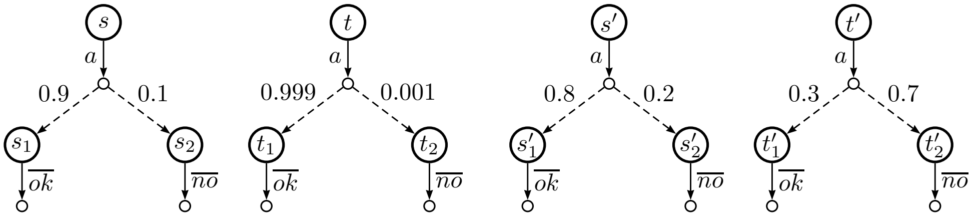
More concretely, let $\mathcal{X}$ be the set of all databases; two databases $x, x' \in \mathcal{X}$ are *adjacent*, written $x \smile x'$, if they differ in the value of a single individual. A *mechanism* is a function $M : \mathcal{X} \to Prob(\mathcal{Z})$ where $\mathcal{Z}$ is some set of reported values. Intuitively, $M(x)$ gives the outcome of the query when applied to database x, which is a probability distribution since noise is added.

Let $tv_\otimes$ be a multiplicative variant of the total variation distance on $Prob(\mathcal{Z})$ (simply called "multiplicative distance" in [29]), defined as:

$$tv_\otimes(\mu, \mu') = \sup_Z |\ln \frac{\mu(Z)}{\mu'(Z)}|$$

Then differential privacy can be defined as follows.[9]

[9] The definition can be generalized to an arbitrary set of secrets $\mathcal{X}$ equipped with a "distinguishability metric" $d_\mathcal{X}$ [11]. The results of this section extend to this setting.



(a) $bm(s,t) = 0.099$ while $\epsilon = \ln 100$.

(b) $bm(s',t') = 0.5$ while $\epsilon' = \ln 3.5$.

Fig. 1: The bisimilarity pseudometric bm does not imply differential privacy.

Definition 6. *A mechanism $M : \mathcal{X} \to Prob(\mathcal{Z})$ is ϵ-differentially private iff*

$$tv_{\otimes}(M(x), M(x')) \leq \epsilon \qquad \forall x \smile x'$$

Intuitively, the definition requires that, when run on adjacent databases, the mechanism should produce similar results, since the distance between the corresponding distributions should be bounded by ϵ (a privacy parameter).

In our setting, we assume that the mechanism M is modelled by a FPA, and the result of the mechanism running on x is the trace produced by the execution of the FPA starting from some corresponding state s_x. That is, $\mathcal{Z} = A^{\omega}$ and

$$M(x) = \Pr[s_x \rhd \cdot] \tag{5}$$

The relation between differential privacy and the multiplicative bisimilarity metric comes from the fact that $tv_{\otimes}$ can be obtained as the $K_{\otimes}$ lifting of the discrete metric on A^{ω}.

Lemma 2. *Let $\delta_V = \mathrm{diam}_{d_{\otimes}}([0,1]) = +\infty$ and let dm_{δ_V} be the discrete metric on A^{ω}. Then $tv_{\otimes} = K_{\otimes}(dm_{\delta_V})$.*

Let $bm_{\otimes}$ be the instantiation of the bisimilarity metric bm_V with $K_{\otimes}$. The above Lemma, together with Theorem 2, imply the following result, which makes $bm_{\otimes}$ useful to verify differential privacy:

Theorem 3. *Let M be the mechanism defined by (5), and assume that*

$$bm_{\otimes}(s_x, s_{x'}) \leq \epsilon \qquad \textit{for all } x \smile x'$$

Then M satisfies ϵ-differential privacy.

Note that the use of the multiplicative $bm_{\otimes}$ is crucial in the above result. The following example shows that the standard bisimilarity metric bm (generated by the original Kantorovich lifting) may be very different from the level of differential privacy, which is expected, since bm bounds the additive total variation metric (Theorem 2 and (4)) instead of the multiplicative $tv_{\otimes}$.

Example 1. Consider the processes s, t shown in Fig. 1 (a). We have that $bm(s,t) = 0.1 - 0.001 = 0.099$ while their level of differential privacy is $\epsilon = \ln {}^{0.1}/_{0.001} = \ln 100$. Moreover, for the processes s', t' shown in Fig. 1 (b) we have $bm(s',t') = 0.7 - 0.2 = 0.5$ while their level of differential privacy is $\epsilon' = \ln {}^{0.7}/_{0.2} = \ln 3.5$. Using the original Kantorovich metric, s and t are considered more indistinguishable than s' and t', in sharp contrast to the corresponding differential privacy levels.

Approximate differential privacy. An approximate, also known as (ϵ, δ) version of differential privacy is also widely used [23], relaxing the definition by an additive factor δ. It requires that:

$$M(x)(Z) \leq e^{\epsilon} M(x')(Z) + \delta \qquad \forall x \smile x', Z \subseteq \mathcal{Z}$$

The α-distance on distributions is proposed in [3] to capture (ϵ, δ)-differential privacy. For two real numbers a, b and a skew parameter $\alpha \geq 1$, the α-distance between a and b is $\max\{a - \alpha b, b - \alpha a, 0\}$. An instantiation of the Kantorovich lifting based on the α-distance seems promising for extending Theorem 3 to the approximate case; we leave this extension as future work.

Weak probabilistic anonymity. Weak probabilistic anonymity was proposed in [18] as a measure of the degree of protection of user's identities. It is defined in a way similar to differential privacy, with the crucial difference (apart from the lack of an adjacency relation) that it uses the (additive) total variation instead of the multiplicative one. Formally, let $\mathcal{X}$ contain the users' identities, and let $M : \mathcal{X} \to Prob(\mathcal{Z})$ be the system in which users operate. We say that M is *ϵ-weakly probabilistically anonymous* iff $tv(M(x), M(x')) \leq \epsilon$ for all $x, x' \in \mathcal{X}$.

For systems modelled by FPAs, by (4) and Theorem 2, we have that if $bm(s_x, s_{x'}) \leq \epsilon$ for all $x, x' \in \mathcal{X}$, then M satisfies ϵ-weak probabilistic anonymity. Hence bm can be used to verify this anonymity property.

6 Process algebra

Process algebras allow to syntactically describe probabilistic processes in terms of a small set of well-understood operators. The operational semantics of a process term is a PA with transitions derived from SOS rules.

In order to specify and verify systems in a compositional manner, it is necessary that the behavioral semantics is compatible with all operators of the language that describe these systems. For behavioral equivalence semantics there is the common agreement that compositional reasoning requires that the considered behavioral equivalence is a congruence wrt all operators. On the other hand, for behavioral metric semantics there are several proposals of properties that operators should satisfy in order to facilitate compositional reasoning [21,1]. In this section we will show that the standard non-recursive process algebra operators are non-expansiveness [21] (as most prominent compositionality property) with respect to the bisimilarity metric.

$$\frac{}{\varepsilon \xrightarrow{\surd} \delta(0)} \qquad \frac{}{a.\bigoplus_{i=1}^{n}[p_i]x_i \xrightarrow{a} \bigoplus_{i=1}^{n} p_i\delta(x_i)}$$

$$\frac{x \xrightarrow{a} \mu \quad a \neq \surd}{x;y \xrightarrow{a} \mu;\delta(y)} \qquad \frac{x \xrightarrow{\surd} \mu \quad y \xrightarrow{a} \nu}{x;y \xrightarrow{a} \nu} \qquad \frac{x \xrightarrow{a} \mu}{x+y \xrightarrow{a} \mu} \qquad \frac{y \xrightarrow{a} \nu}{x+y \xrightarrow{a} \nu}$$

$$\frac{x \xrightarrow{a} \mu \quad y \xrightarrow{a} \nu}{x \mid y \xrightarrow{a} \mu \mid \nu} \qquad \frac{x \xrightarrow{a} \mu}{x \parallel y \xrightarrow{a} \mu \parallel \delta(y)} \qquad \frac{y \xrightarrow{a} \nu}{x \parallel y \xrightarrow{a} \delta(x) \parallel \nu}$$

$$\frac{x \xrightarrow{a} \mu \quad y \stackrel{a}{\not\longrightarrow}}{x +_p y \xrightarrow{a} \mu} \qquad \frac{x \stackrel{a}{\not\longrightarrow} \quad y \xrightarrow{a} \nu}{x +_p y \xrightarrow{a} \nu} \qquad \frac{x \xrightarrow{a} \mu \quad y \xrightarrow{a} \nu}{x +_p y \xrightarrow{a} \mu \oplus_p \nu}$$

$$\frac{x \xrightarrow{a} \mu \quad y \stackrel{a}{\not\longrightarrow}}{x \parallel_p y \xrightarrow{a} \mu \parallel_p \delta(y)} \qquad \frac{x \stackrel{a}{\not\longrightarrow} \quad y \xrightarrow{a} \nu}{x \parallel_p y \xrightarrow{a} \delta(x) \parallel_p \nu} \qquad \frac{x \xrightarrow{a} \mu \quad y \xrightarrow{a} \nu}{x \parallel_p y \xrightarrow{a} \mu \parallel_p \delta(y) \oplus_p \delta(x) \parallel_p \nu}$$

Table 2: Probabilistic process algebra operators

We introduce a simple probabilistic process algebra that comprises the following operators i) constants 0 (stop process) and ϵ (skip process); ii) a family of n-ary prefix operators $a.([p_1]\_\oplus\ldots\oplus[p_n]\_)$ with $a \in Act$, $n \geq 1$, $p_1, \ldots, p_n \in (0,1]$ and $\sum_{i=1}^{n} p_i = 1$; iii) binary operators $\_;\_$ (sequential composition), $\_+\_$ (alternative composition), $\_+_p\_$ (probabilistic alternative composition), $\_\mid\_$ (synchronous parallel composition), $\_\parallel\_$ (asynchronous parallel composition), and $\_\parallel_p\_$ (probabilistic parallel composition). We assume a set of actions Act with the distinguished action $\surd \in A$ to denote successful termination. The operational semantics of all operators is specified by the rules in Table 2.

We use distribution terms in the target of rules (right hand side of the conclusion of the rules) in order to describe distributions. We briefly recall the semantics of distribution terms of [26,16]. The expression $\delta(x)$ denotes a Dirac distribution on x. The expression $\mu;\delta(y)$ denotes a distribution such that $(\mu;\delta(y))(x;y) = \mu(x)$, the expression $\mu \oplus_p \nu$ denotes a distribution such that $(\mu \oplus_p \nu)(x) = p\mu(x) + (1-p)\nu(x)$, and $(\mu \parallel \nu)(s \parallel t) = \mu(s)\nu(t)$.

The probabilistic prefix operator expresses that the process $a.([p_1]t_1 \oplus \ldots \oplus_p lus[p_n]t_n)$ can perform action a and evolves to process t_i with probability p_i. The sequential composition and the alternative composition are as usual. The synchronous parallel composition $s \mid t$ describes the simultaneous evolution of processes s and t, while the asynchronous parallel composition $t \parallel t$ describes the interleaving of s and t where both processes can progress by alternating at any rate the execution of their actions. The probabilistic alternative and probabilistic parallel composition replaces the nondeterministic choice of their non-probabilistic variants by a probabilistic choice. The probabilistic alternative composition $s +_p t$ evolves to the probabilistic choice between a distribution reached by s (with probability p) and a distribution reached by t (with probability $1-p$) for actions which can be performed by both processes. For actions that can be performed by either only s or only t, the probabilistic alternative composition $s +_p t$ behaves just like the nondeterministic alternative composition $s+t$. Similarly, the probabilistic parallel composition $s \parallel_p t$ evolves to a probabilistic choice between the nondeterministic choices of asynchronous parallel composition of s and t.

We start by showing an important auxiliary property how the distance between convex combinations of probability distributions relates to the distance between the combined probability distributions.

Proposition 6. *Let $\mu_1, \mu_2, \mu'_1, \mu'_2 \in Disc(X)$ and $p \in [0,1]$. Then*

$$K_\otimes(bm_\otimes)(p\mu_1+(1-p)\mu_2, p\mu'_1+(1-p)\mu'_2) \leq \max(K_\otimes(bm_\otimes)(\mu_1,\mu'_1), K_\otimes(bm_\otimes)(\mu_2,\mu'_2))$$

Non-expansiveness is the most wildly studied compositionality property stating that the distance between composed processes is at most the sum of the distance between its parts.

Definition 7. *A n-ary operator f is non-expansive wrt a pseudometric m if*

$$m(f(s_1,\ldots,s_n), f(t_1,\ldots,t_n)) \leq \sum_{i=1}^{n} m(s_i,t_i)$$

Now we can show that all (non-recursive) operators of the probabilistic process algebra introduced above are non-expansive. In fact, we will provide upper bounds on distance between the composed processes which are in case of the (nondeterministic and probabilistic) alternative composition even stricter than the non-expansiveness condition.

Theorem 4. *Let s, t, s', t' be probabilistic processes. Then*

1. $bm_\otimes(s;t, s';t') \leq bm_\otimes(s,s') + bm_\otimes(t,t')$
2. $bm_\otimes(s+t, s'+t') \leq \max(bm_\otimes(s,s'), bm_\otimes(t,t'))$
3. $bm_\otimes(s+_p t, s'+_p t') \leq \max(bm_\otimes(s,s'), bm_\otimes(t,t'))$
4. $bm_\otimes(s \mid t, s' \mid t') \leq bm_\otimes(s,s') + bm_\otimes(t,t')$
5. $bm_\otimes(s \parallel t, s' \parallel t') \leq bm_\otimes(s,s') + bm_\otimes(t,t')$
6. $bm_\otimes(s \parallel_p t, s \parallel_p t') \leq bm_\otimes(s,s') + bm_\otimes(t,t')$

A similar result can be gained for the bisimilarity metric bm based on the standard Kantorovich lifting. This generalizes a similar result of [21] which considered only PTSs without nondeterministic branching and only a small set of process combinators.

For the generalized bisimilarity metric bm_V we can formulate a similar result for the nondeterministic alternative composition.

Theorem 5. *Let s, t, s', t' be probabilistic processes. Then*

$$bm_V(s+t, s'+t') \leq \max(bm_V(s,s'), bm_V(t,t'))$$

7 Conclusion and future work

We have proposed a family of Kantorovich pseudometrics depending on the notion of distance used to specify properties over traces. We have developed the theory of this notion, and showed how we can use it to verify the corresponding kind of properties. We have also showed that for the multiplicative variant, which is an interesting case because of its relation with differential privacy, it is possible to give a dual form that makes the metric computable by standard techniques.

Future work include the investigation of methods to compute other members of this family, and of conditions that make possible a general dual form.

References

1. Bacci, G., Bacci, G., Larsen, K.G., Mardare, R.: Computing Behavioral Distances, Compositionally. In: Proc. MFCS'13, pp. 74–85. Springer (2013)
2. Bacci, G., Bacci, G., Larsen, K.G., Mardare, R.: On-the-fly exact computation of bisimilarity distances. In: TACAS. LNCS, vol. 7795, pp. 1–15. Springer (2013)
3. Barthe, G., Köpf, B., Olmedo, F., Béguelin, S.Z.: Probabilistic relational reasoning for differential privacy. In: Proc. of POPL. ACM (2012)
4. van Breugel, F., Worrell, J.: An algorithm for quantitative verification of probabilistic transition systems. In: Proc. of CONCUR'01. pp. 336–350. Springer (2001)
5. van Breugel, F., Worrell, J.: Towards quantitative verification of probabilistic transition systems. In: Proc. of ICALP. LNCS, vol. 2076, pp. 421–432. Springer (2001)
6. van Breugel, F., Worrell, J.: A behavioural pseudometric for probabilistic transition systems. Theor. Comp. Sci. 331(1), 115–142 (2005)
7. van Breugel, F., Worrell, J.: Approximating and computing behavioural distances in probabilistic transition systems. Theor. Comp. Sci. 360(1-3), 373 – 385 (2006)
8. van Breugel, F., Worrell, J.: The complexity of computing a bisimilarity pseudometric on probabilistic automata. In: Horizons of the Mind. LNCS, vol. 8464, pp. 191–213. Springer (2014)
9. Cai, X., Gu, Y.: Measuring anonymity. In: ISPEC, LNCS, vol. 5451, pp. 183–194. Springer (2009)
10. Chatterjee, K., de Alfaro, L., Majumdar, R., Raman, V.: Algorithms for Game Metrics. In: FSTTCS. vol. 2, pp. 107–118. Leibniz-Zentrum fuer Informatik (2008)
11. Chatzikokolakis, K., Andrés, M.E., Bordenabe, N.E., Palamidessi, C.: Broadening the scope of Differential Privacy using metrics. In: Proc. of PETS. LNCS, vol. 7981, pp. 82–102. Springer (2013)
12. Chatzikokolakis, K., Gebler, D., Palamidessi, C., Xu, L.: Generalized bisimulation metrics. Tech. rep., INRIA (2014)
13. Chen, D., van Breugel, F., Worrell, J.: On the complexity of computing probabilistic bisimilarity. In: FOSSACS. LNCS, vol. 7213, pp. 437–451. Springer (2012)
14. Cheng, S.: A crash course on the lebesgue integral and measure theory (2008)
15. Comanici, G., Precup, D.: Basis function discovery using spectral clustering and bisimulation metrics. In: AAAI, LNCS, vol. 7113, pp. 85–99. Springer (2012)
16. D'Argenio, P.R., Gebler, D., Lee, M.D.: Axiomatizing Bisimulation Equivalences and Metrics from Probabilistic SOS Rules. In: Proc. FoSSaCS'14. LNCS, vol. 8412, pp. 289–303. Springer (2014)
17. Deng, Y., Du, W.: The kantorovich metric in computer science: A brief survey. ENTCS 253(3), 73–82 (2009)

18. Deng, Y., Palamidessi, C., Pang, J.: Weak probabilistic anonymity. In: Proc. of SecCo. ENTCS, vol. 180 (1), pp. 55–76. Elsevier (2007)
19. Desharnais, J., Gupta, V., Jagadeesan, R., Panangaden, P.: Metrics for labeled markov systems. In: Proc. of CONCUR. LNCS, vol. 1664, pp. 258–273. Springer (1999)
20. Desharnais, J., Jagadeesan, R., Gupta, V., Panangaden, P.: The metric analogue of weak bisimulation for probabilistic processes. In: Proc. of LICS. pp. 413–422. IEEE (2002)
21. Desharnais, J., Jagadeesan, R., Gupta, V., Panangaden, P.: Metrics for labelled Markov processes. Theor. Comp. Sci. 318(3), 323–354 (2004)
22. Dwork, C.: Differential privacy. In: Proc. of ICALP. LNCS, vol. 4052, pp. 1–12. Springer (2006)
23. Dwork, C., Kenthapadi, K., Mcsherry, F., Mironov, I., Naor, M.: Our data, ourselves: Privacy via distributed noise generation. In: In EUROCRYPT. pp. 486–503. Springer (2006)
24. Ferns, N., Panangaden, P., Precup, D.: Metrics for markov decision processes with infinite state spaces. In: Proc. of UAI. pp. 201–208. AUAI Press (2005)
25. Ferns, N., Panangaden, P., Precup, D.: Bisimulation metrics for continuous markov decision processes. SIAM J. Comput 40(6), 1662–1714 (2011)
26. Lee, M.D., Gebler, D., D'Argenio, P.R.: Tree Rules in Probabilistic Transition System Specifications with Negative and Quantitative Premises. In: Proc. EXPRESS/SOS'12. EPTCS, vol. 89, pp. 115–130 (2012)
27. Norfolk, T.: When does a metric generate convex balls? Tech. rep. (1991), `http://www.math.uakron.edu/~norfolk/convex.ps`
28. Reed, J., Pierce, B.C.: Distance makes the types grow stronger: a calculus for differential privacy. In: Proc. of ICFP. pp. 157–168. ACM (2010)
29. Smith, A.: Efficient, differentially private point estimators. arXiv preprint arXiv:0809.4794 (2008)
30. Tarski, A.: A lattice-theoretical fixpoint theorem and its applications. Pacific Journal of Mathematics 5(2), 285—309 (1955)
31. Thorsley, D., Klavins, E.: Approximating stochastic biochemical processes with wasserstein pseudometrics. Systems Biology, IET 4(3), 193–211 (May 2010)
32. Tschantz, M.C., Kaynar, D., Datta, A.: Formal verification of differential privacy for interactive systems (extended abstract). ENTCS 276, 61–79 (sep 2011)
33. Xu, L., Chatzikokolakis, K., Lin, H.: Metrics for differential privacy in concurrent systems. In: Proc. of FORTE. LNCS, vol. 8461, pp. 199–215. Springer (2014)

A Proofs and additional results

We include here all proofs omitted from the main paper due to space constraints, as well as some additional results about the generalized Kantorovich lifting.

A.1 Bounding the distance between distributions

In this section we show that under the *ball-convexity* constraint for d_V, the distance $K_V(m)(\mu, \mu')$ is bounded by the support of μ, μ'.

We say that (V, d_V) is *ball-convex* if $B_r^{d_V}(x)$ is convex for all $r > 0, x \in V$. Not all metrics have this property, in fact in [27] it is shown that (V, d_V) is ball-convex iff

$$d_V(x, \lambda y_1 + \bar{\lambda} y_2) \leq \max\{d_V(x, y_1), d_V(x, y_2)\} \qquad \forall x, y_1, y_2 \in V, \lambda \in [0, 1]$$

i.e. iff $d_V(x, \cdot)$ is a quasi-convex function for any fixed $x \in V$. Many standard metrics (for instance all norms) satisfy this property. Moreover the metric $d_\otimes$ used in the multiplicative Kantorovich variant (Section 5) also satisfies it.

The usefulness of ball-convexity is given by the following proposition, stating that on such metrics, convex combinations cannot increase distances. We denote by $\mathrm{ch}(A)$ the convex hull of A.

Proposition 7. *Let (V, d_V) be ball-convex and $A \subseteq V$. Then* $\mathrm{diam}_{d_V}(\mathrm{ch}(A)) = \mathrm{diam}_{d_V}(A)$.

Proof. From $A \subseteq \mathrm{ch}(A)$ we get $\mathrm{diam}_{d_V}(A) \leq \mathrm{diam}_{d_V}(\mathrm{ch}(A))$. We now show that $\mathrm{diam}_{d_V}(\mathrm{ch}(A)) \leq \mathrm{diam}_{d_V}(A)$.

Let $\delta = \mathrm{diam}_{d_V}(A)$ and assume that $\mathrm{diam}_{d_V}(\mathrm{ch}(A)) > \delta$, i.e. $\exists x, y \in \mathrm{ch}(A)$ s.t. $d_V(x, y) > \delta$. If $A \subseteq B_\delta^{d_V}(x)$ then $\mathrm{ch}(A) \subseteq \mathrm{ch}(B_\delta^{d_V}(x)) = B_\delta^{d_V}(x)$ (balls are convex) which is a contradiction since $y \notin B_\delta^{d_V}(x)$. Hence it must hold that $A \not\subseteq B_\delta^{d_V}(x)$, that is $\exists z \in A$ with $d_V(x, z) > \delta$.

Finally, from $z \in A$ we get $A \subseteq B_\delta^{d_V}(z)$, hence $\mathrm{ch}(A) \subseteq \mathrm{ch}(B_\delta^{d_V}(z)) = B_\delta^{d_V}(z)$ which is a contradiction since $a \notin B_\delta^{d_V}(z)$. □

As a corollary of the previous result, we can bound the Kantorovich lifting of a pseudometric m.

Proposition 8. *Let (V, d_V) be ball-convex. Then*

$$K_V(m)(\mu, \mu') \leq \mathrm{diam}_m(\mathit{supp}(\mu) \cup \mathit{supp}(\mu'))$$

Proof. Let $f \in$ 1-Lip$[(X, m)(V, d_V)]$, let $A = \mathit{supp}(\mu) \cup \mathit{supp}(\mu')$, and let $f(A)$ denote the set $\{f(x) : x \in A\}$. We have that

$$\begin{aligned} d_V(\hat{f}(\mu), \hat{f}(\mu')) &\leq \mathrm{diam}_{d_V}(\mathrm{ch}(f(A))) && \hat{f}(\mu), \hat{f}(\mu') \in \mathrm{ch}(f(A)) \\ &= \mathrm{diam}_{d_V}(f(A)) && \text{Prop 7} \\ &\leq \mathrm{diam}_m(A) && \text{1-Lipschitz} \end{aligned}$$

This holds for all 1-Lipschitz functions, hence $K_V(m)(\mu, \mu') \leq \mathrm{diam}_m(A)$. □

A.2 Proofs of Section 4 A general family of bisimilarity pseudometrics

Proposition 1. *For any $\epsilon \geq 0$, $F_V(m)(s,t) \leq \epsilon$ if and only if:*

- *if $s \xrightarrow{a} \mu$, then there exists ν such that $t \xrightarrow{a} \nu$ and $K_V(m)(\mu,\nu) \leq \epsilon$,*
- *if $t \xrightarrow{a} \nu$, then there exists μ such that $s \xrightarrow{a} \mu$ and $K_V(m)(\nu,\mu) \leq \epsilon$.*

Proof. The proposition can be proved by directly checking the definition of F_V. □

Proposition 2. *Let $m, m' \in \mathcal{M}$. If $m \preceq m'$ then:*

$$F_V(m)(s,s') \leq F_V(m')(s,s') \quad \textit{for all states } s, s'$$
$$K_V(m)(\mu,\mu') \leq K_V(m')(\mu,\mu') \quad \textit{for all distributions } \mu, \mu'$$

Proof. The essence of the proof is the observation that

$$\text{1-Lip}[(V,d_V),(S,m)] \subseteq \text{1-Lip}[(V,d_V),(S,m')]$$

whenever $m \preceq m'$. □

Proposition 3. *Assume that S is finite. Let $m_0 = \bot$ and $m_{i+1} = F_V(m_i)$. Then $bm_V = \sup_i m_i$.*

Proof. Since the closure ordinal of F_V is ω, following the standard way, one can approximate the least fixpoint bm_V by iterating the function F_V from the bottom element. □

Proofs of Section 4.1 Bisimilarity as 0-distance

Lemma 1. *If (V, d_V) has a geodesic then $\mathcal{L}(\ker(m))$ and $\ker(K_V(m))$ coincide on all distributions of finite support.*

Proof. Direction $\subseteq$: let $(\mu,\mu') \in L(\ker(m))$ and let $f : S \to V$ be 1-Lipschitz wrt m, d_V. Every such function needs to map equivalent elements of S to the same element of V, since $(s,s') \in \ker(m)$ implies $m(s,s') = 0$ which, from 1-Lipschitz, means that $d_V(f(s), f(s')) = 0$ which in turn implies $f(s) = f(s')$.

For simplicity, we write $[s]$ for $[s]_{\ker(m)}$. Let S_r be a set of representatives of each class, i.e. $S = \uplus_{s \in S_r}[s]$. Then

$$\begin{aligned}
\hat{f}(\mu) &= \sum_{s\in S_r} \sum_{s'\in[s]} \mu(s')f(s') \\
&= \sum_{s\in S_r} f(s)\mu([s]) && f(s') \text{ is common for the class} \\
&= \sum_{s\in S_r} f(s)\mu'([s]) && (\mu,\mu') \in L(\ker(m)) \\
&= \hat{f}(\mu')
\end{aligned}$$

Hence $d_V(\hat{f}(\mu), \hat{f}(\mu')) = 0$ and this happens for all such f, which implies $K_V(m)(\mu, \mu') = 0$, that is $(\mu, \mu') \in \ker(K_V(m))$. Note that this direction requires neither an assumption on (V, d_V), nor that μ, μ' have finite support.

Direction $\supseteq$: let $(\mu, \mu') \notin L(\ker(m))$ such that $S_+ = \mathit{supp}(\mu) \cup \mathit{supp}(\mu')$ is finite; we show that $(\mu, \mu') \notin \ker(K_V(m))$. Since μ, μ' are not equivalent, there exists $s_0 \in S$ such that $\mu([s_0]) \neq \mu'([s_0])$. Let $\zeta > 0$ be the minimum distance between s_0 and elements of S_+ not equivalent to s_0, that is

$$\zeta = \min_{s \in S_+ \setminus [s_0]} m(s, s_0)$$

Moreover, let $\gamma : [0, d] \to V$ be a geodesic[10] of (V, d_V), and take some $\epsilon > 0$ that is smaller than both ζ and d.

We define a function $f : S \to V$ as:

$$f = \gamma \circ g \qquad \text{where} \qquad g(s) = \min\{m(s, s_0), \epsilon\}$$

We first show that f is 1-Lipshitz wrt m, d_V. Let $s, s' \in S$ and assume wlog that $g(s) \geq g(s')$. From the definiton of g it follows that:

$$g(s) - g(s') \leq m(s, s_0) - m(s', s_0) \tag{6}$$

Then we have that:

$$\begin{aligned} d_V(f(s), f(s')) &= d_V(\gamma(g(s)), \gamma(g(s'))) && \text{Def. of } f \\ &= g(s) - g(s') && \gamma \text{ is a geodesic} \\ &\leq m(s, s_0) - m(s', s_0) && (6) \\ &\leq m(s, s') && \text{triangle ineq.} \end{aligned}$$

hence f is 1-Lipshitz wrt m, d_V.

Moreover, since $\epsilon < \zeta$, for all elements $s \in S_+$ we have that either $g(s) = 0$ (when $s \in [s_0]$) or $g(s) = \epsilon$, hence f maps all elements of $S_+ \cap [s_0]$ to $\gamma(0)$ and all elements of $S_+ \setminus [s_0]$ to $\gamma(\epsilon)$. Finally, for any $a \neq b \in \mathbb{R}, \lambda \neq \lambda' \in [0, 1]$ it holds that $a\lambda + b(1 - \lambda) \neq a\lambda' + b(1 - \lambda')$, as a consequence:

$$\begin{aligned} \hat{f}(\mu) &= \textstyle\sum_{s \in S_+} \mu(s) f(s) \\ &= \gamma(0)\mu\left([s_0]\right) + \gamma(\epsilon)(1 - \mu\left([s_0]\right)) \\ &\neq \gamma(0)\mu'([s_0]) + \gamma(\epsilon)(1 - \mu'([s_0])) \\ &= \hat{f}(\mu') \end{aligned}$$

Hence $d_V(\hat{f}(\mu), \hat{f}(\mu')) > 0$ which implies $K_V(m)(\mu, \mu') > 0$, that is $(\mu, \mu') \notin \ker(K_V(m))$. □

Note that in the above proof we need a geodesic γ since in general there might be elements of S arbitrarily close to s_0, and we need to map such elements to V

[10] Wlog we can take γ's domain to be of the form $[0, d]$.

while preserving the 1-Lipshitz condition. However, if S is finite, we can always find an $\epsilon > 0$ smaller than the distance between s_0 and any $s \notin [s_0]$. In this case it is enough that (V, d_V) has a *non-isolated point* a, so we can find $b \in V$ s.t. $d_V(a, b) < \epsilon$, then define f as $f(s) = a$ iff $s \in [s_0]$ and $f(s) = b$ otherwise, and continue the proof in the same way.

Proposition 4. *Assume a finitely branching PA and that (V, d_V) has a geodesic. For every $m \in \mathcal{M}$, if $F_V(m) \preceq m$ then $\ker(m) \subseteq B(\ker(m))$, i.e., $\ker(m)$ is a bisimulation.*

Proof. Let $(s, t) \in \ker(m)$, i.e. $m(s, t) = 0$. Since $F_V(m) \preceq m$, we have that if $s \xrightarrow{a} \mu$, then there exists ν such that $t \xrightarrow{a} \nu$ and $K_V(m)(\mu, \nu) = 0$. Clearly $(\mu, \nu) \in \ker(K_V(m))$, by Lemma 1, it follows that $(\mu, \nu) \in \mathcal{L}(\ker(m))$. A similar condition holds for the converse direction where t initiates transitions. Hence, we have $(s, t) \in B(\ker(m))$. □

Theorem 1. *Assume a finitely branching PA and that (V, d_V) has a geodesic. Then $\ker(bm_V) = \sim$.*

Proof. Since bm_V is a fixpoint of F_V, then $\ker(bm_V)$ is a probabilistic bisimulation relation. Viceversa, let R be a probabilistic bisimulation relation. Define $m(s, t) = 0$ if $(s, t) \in R$, and $m(s, t) = \mathrm{diam}_{d_V}(V)$ otherwise. Due to Lemma 1 we have $F_V(m) \preceq m$, hence $bm_V \preceq m$, which means that $\ker(bm_V) \subseteq \ker(m) = R$. □

A.3 Proofs of Section 4.2 Relation with trace distributions

Proposition 5. *$K_V(dm_{\delta_V})(\mu, \mu') = 0$ iff $\mu(\sigma) = \mu'(\sigma)$ for all measurable $\sigma \subseteq A^\omega$.*

Proof. We have

$$\begin{aligned}
&K_V(dm_{\delta_V})(\mu, \mu') = 0 \\
\text{iff}\ \ &\text{for any } f \in \text{1-Lip}[(A^\omega, dm_{\delta_V})(V, d_V)], d_V(\hat{f}(\mu), \hat{f}(\mu')) = 0 \\
\text{iff}\ \ &\text{for any } f \in \text{1-Lip}[(A^\omega, dm_{\delta_V})(V, d_V)], \hat{f}(\mu) = \hat{f}(\mu') \qquad d_V \text{ is a metric.} \qquad (7)
\end{aligned}$$

We shall show that the right hand part (7) of the above relation is equivalent to the right hand part of Proposition 5.

($\Leftarrow$) If $\mu(\sigma) = \mu'(\sigma)$ for all measurable $\sigma \subseteq A^\omega$, by checking the definition of $\hat{f}$, it is straightforward that $\hat{f}(\mu) = \hat{f}(\mu')$ for any f.

($\Rightarrow$) For the converse direction, we assume that there exists a measurable $\sigma \subseteq A^\omega$ such that $\mu(\sigma) \neq \mu'(\sigma)$. We construct a function $f \in \text{1-Lip}[(A^\omega, dm_{\delta_V})(V, d_V)]$: $f(\boldsymbol{t}) = c$ for $\boldsymbol{t} \in \sigma$, 0 otherwise, where c is a constant in V. We get that $\hat{f}(\mu) = c \cdot \mu(\sigma)$ and $\hat{f}(\mu') = c \cdot \mu'(\sigma)$. Due to the assumption, $\hat{f}(\mu) \neq \hat{f}(\mu')$, which contradicts (7). □

A.4 Proofs of Section 5 The multiplicative variant

Transformations of the linear-fractional program In case S is finite, and since ln is a monotonically increasing function, the multiplicative Kantorovich distance can be computed by solving the following linear-fractional program:

$$\begin{array}{ll} \text{maximize} & \dfrac{\sum_i \mu(s_i)x_i}{\sum_i \mu'(s_i)x_i} \\[2ex] \text{subject to:} & \forall i,j.\ x_i \leq e^{m(s_i,s_j)}x_j. \end{array}$$

We now show how the above program can be first converted to a linear one, and then written in dual form.

Following the techniques in [4], we extend the dimensions of the feasible region by adding new decision variables y_i for $i \in [1,|S|]$. The extension does not affect the optimal value. This is justified by the new constraints ensuring that in fact $x_i = y_i$ for $i \in [1,|S|]$ (because $m(s_i,s_i) = 0$).

$$\begin{array}{ll} \text{maximize} & \dfrac{\sum_i \mu(s_i)x_i}{\sum_j \mu'(s_j)y_j} \\[2ex] \text{subject to:} & \forall i,j.\ x_i - e^{m(s_i,s_j)}y_j \leq 0 \\[1ex] & \forall i,\ y_i - x_i \leq 0. \end{array}$$

Now let

$$\alpha_i = \frac{x_i}{\sum_j \mu'(s_j)y_j} \quad \beta_i = \frac{y_i}{\sum_j \mu'(s_j)y_j} \quad t = \frac{1}{\sum_j \mu'(s_j)y_j}$$

The above linear-fractional problem can be transformed to the equivalent linear program.

$$\begin{array}{ll} \text{maximize} & \sum_i \mu(s_i)\alpha_i \\[2ex] \text{subject to:} & \forall i,j.\ \alpha_i - e^{m(s_i,s_j)}\beta_j \leq 0 \\[1ex] & \forall i,\ \beta_i - \alpha_i \leq 0 \\[1ex] & \sum_i \mu'(s_i)\beta_i = 1 \\[1ex] & \forall i.\ \alpha_i, \beta_i \geq 0. \end{array}$$

Assign the new variables l_{ij}, r_i, z to the first three kinds of constraints respectively, then dualizing the above (primal) problem yields:

$$\begin{array}{ll} \text{minmize} & z \\[1ex] \text{subject to:} & \forall i.\ \sum_j l_{ij} - r_i \geq \mu(s_i) \\[1ex] & \forall j.\ \sum_i l_{ij}e^{m(s_i,s_j)} - r_j \leq z \cdot \mu'(s_j) \\[1ex] & \forall i,j.\ l_{ij}, r_i \geq 0 \end{array}$$

which is equivalent to the following program where the first kind of constraints becomes an equation:

$$\begin{array}{ll} \text{minmize} & z \\ \text{subject to:} & \forall i.\, \sum_j l_{ij} - r_i = \mu(s_i) \\ & \forall j.\, \sum_i l_{ij} e^{m(s_i,s_j)} - r_j \le z \cdot \mu'(s_j) \\ & \forall i,j.\, l_{ij}, r_i \ge 0. \end{array}$$

Proofs of Section 5.1 Application to differential privacy

The following lemma is used later for the proof of Lemma 2.

Lemma 3. *Let $a, a', b, b' \in [0,1]$ s.t. $a + b \le 1, a' + b' \le 1$, and let*

$$g(x) = d_{\otimes}(a + bx, a' + b'x)$$

Then $g(x) \le \max\{g(0), g(1)\}$ for all $x \in [0,1]$.

Proof. Wlog assume $a, a', b, b' > 0$, we can extend to the case 0 by continuity. Define

$$h(x) = \frac{a\ + b\ x}{a' + b'x}$$

The derivative of h is $h'(x) = \frac{a'b - ab'}{(a'+b'x)^2}$, hence h is monotonically increasing when $a'b \ge ab'$ and monotonically decreasing otherwise. This implies that

$$h(x) \le \max\{h(0), h(1)\} \quad \text{and} \quad h^{-1}(x) \le \max\{h^{-1}(0), h^{-1}(1)\} \tag{8}$$

We have:

$$\begin{aligned} g(x) &= |\ln h(x)| && \text{Def. of } d_{\otimes} \\ &= \max\{\ln h(x), \ln h^{-1}(x)\} \\ &\le \max\{\ln h(0), \ln h(1), \ln h^{-1}(0), \ln h^{-1}(0)\} && \text{(8), monot. of ln} \\ &= \max\{g(0), g(1)\} \end{aligned}$$

□

Lemma 2. *Let $\delta_V = \mathrm{diam}_{d_{\otimes}}([0,1]) = +\infty$ and let dm_{δ_V} be the discrete metric on A^{ω}. Then $tv_{\otimes} = K_{\otimes}(dm_{\delta_V})$.*

Proof. Any function $f : A^{\omega} \to [0,1]$ is 1-Lipschitz wrt $d_{\otimes}, dm_{\delta_V}$, hence

$$K_{\otimes}(dm_{\delta_V})(\mu, \mu') = \sup_f d_{\otimes}(\hat{f}(\mu), \hat{f}(\mu'))$$

where f ranges over all measurable functions and $\hat{f}(\mu) = \int f d\mu$. Recall that

$$tv_{\otimes}(\mu, \mu') = \sup_Z |\ln \frac{\mu(Z)}{\mu(Z')}| = \sup_Z d_{\otimes}(\mu(Z), \mu'(Z))$$

Let $\mathbf{1}_Z$ be the indicator function, defined as $\mathbf{1}_Z(x) = 1$ iff $x \in Z$ and $\mathbf{1}_Z(x) = 0$ otherwise. We have that $\hat{\mathbf{1}_Z}(\mu) = \mu(Z)$, hence

$$d_{\otimes}(\hat{\mathbf{1}_Z}(\mu), \hat{\mathbf{1}_Z}(\mu')) = d_{\otimes}(\mu(Z), \mu'(Z)) \tag{9}$$

Direction $\leq$) This is the easy case, since (9) implies that every Z in the definition of $tv_{\otimes}(\mu, \mu')$ can be matched by an f in the definition of $K_{\otimes}(dm_{\delta_V})(\mu, \mu')$.

Direction $\geq$) A function ϕ is called *simple* if its image $\mathrm{img}(\phi)$ is a finite set. Let Φ be the set of all measurable simple functions from A^ω to $[0,1]$. Any $\phi \in \Phi$ can be represented as $\phi = \sum_{v \in \mathrm{img}(f)} v \cdot \mathbf{1}_{f^{-1}(v)}$ hence $\hat{\phi}(\mu) = \sum_{v \in \mathrm{img}(f)} v \cdot \mu(f^{-1}(a))$. A simple function ϕ is an indicator function iff $\mathrm{img}(\phi) \subseteq \{0, 1\}$.

We are going to show that

$$d_{\otimes}(\hat{\phi}(\mu), \hat{\phi}(\mu')) \leq tv_{\otimes}(\mu, \mu') \qquad \forall \phi \in \Phi \tag{10}$$

The intuition is that we can bound $d_{\otimes}(\hat{\phi}(\mu), \hat{\phi}(\mu'))$ from above by changing ϕ's values to either 0 or 1. After replacing all values we end up with an indicator function, for which the distance is bounded by $tv_{\otimes}(\mu, \mu')$ because of (9).

Formally, we show (10) by induction on $n = |\mathrm{img}(\phi) \setminus \{0,1\}|$, i.e. the (finite) number of ϕ's values that are neither 0 nor 1. For the base case $n = 0$, ϕ is an indicator function and (10) follows directly from (9). Now assume (10) holds for $n \leq k$ and let $\phi \in \Phi$ s.t. $n = k + 1$. Then there exists some $v \in \mathrm{img}(\phi)$ s.t. $0 < v < 1$.

Let $\phi_x \in \Phi$ be the function obtained from ϕ after mapping v to x (hence $\phi = \phi_v$). Note that $\hat{\phi_x}(\mu), \hat{\phi_x}(\mu')$ can be written as $a + bx$ and $a' + b'x$ respectively, with a, a', b, b' satisfying the conditions of Lemma 3. As a consequence we have that

$$d_{\otimes}(\hat{\phi_v}(\mu), \hat{\phi_v}(\mu')) \leq \max\{d_{\otimes}(\hat{\phi_0}(\mu), \hat{\phi_0}(\mu')), d_{\otimes}(\hat{\phi_1}(\mu), \hat{\phi_1}(\mu'))\}$$

From the induction hypothesis both $d_{\otimes}(\hat{\phi_0}(\mu), \hat{\phi_0}(\mu'))$ and $d_{\otimes}(\hat{\phi_1}(\mu), \hat{\phi_1}(\mu'))$ are bounded from above by $tv_{\otimes}(\mu, \mu')$, which concludes the proof of (10).

Having shown (10), it only remains to extend it to any non-simple measurable $f : A^\omega \to [0,1]$. This comes by approximating f using simple functions: there exist ϕ_n increasing pointwise and converging pointwise to f. From the Monotone Convergence Theorem we have that $\hat{f}(\mu) = \lim_{n\to\infty} \hat{\phi_n}(\mu)$ (see [14], Thm 2.4.10 and 3.1.1). We conclude by the continuity of $d_{\otimes}$, since $\lim_{n\to\infty} d_{\otimes}(\hat{\phi_n}(\mu), \hat{\phi_n}(\mu'))$ $= d_{\otimes}(\hat{f}(\mu), \hat{f}(\mu'))$. □

Theorem 3. *Let M be the mechanism defined by (5), and assume that*

$$bm_{\otimes}(s_x, s_{x'}) \leq \epsilon \qquad \textit{for all } x \smile x'$$

Then M satisfies ϵ-differential privacy.

Proof. We have that $M(x) = \Pr[s_x \rhd \cdot\,]$ and $M(x') = \Pr[s_{x'} \rhd \cdot\,]$, hence:

$$\begin{aligned} tv_{\otimes}(M(x), M(x')) &= K_{\otimes}(dm_{\delta_V})(M(x), M(x')) && \text{Lemma 2} \\ &\leq bm_{\otimes}(s_x, s_{x'}) && \text{Theorem 2} \\ &\leq \epsilon && \text{hypothesis} \end{aligned}$$

□

A.5 Proofs of Section 6 Process algebra

Proposition 6. *Let $\mu_1, \mu_2, \mu_1', \mu_2' \in Disc(X)$ and $p \in [0,1]$. Then*

$$K_\otimes(bm_\otimes)(p\mu_1+(1-p)\mu_2, p\mu_1'+(1-p)\mu_2') \leq \max(K_\otimes(bm_\otimes)(\mu_1,\mu_1'), K_\otimes(bm_\otimes)(\mu_2,\mu_2'))$$

Proof. By the definition of $K_\otimes(bm_\otimes)$,

$$K_\otimes(bm_\otimes)(p\mu_1+(1-p)\mu_2, p\mu_1'+(1-p)\mu_2') = \max|\ln \frac{\sum_i [p\mu_1(s_i)+(1-p)\mu_2(s_i)]x_i}{\sum_i [p\mu_1'(s_i)+(1-p)\mu_2'(s_i)]x_i}|$$

under the constraints: $\forall i,j.\ x_i \leq e^{bm_\otimes(s_i,s_j)}x_j$. Let x_i^*'s be the variables that realize the maximum value on the problem. We have:

$$\begin{aligned}
&K_\otimes(bm_\otimes)(p\mu_1+(1-p)\mu_2, p\mu_1'+(1-p)\mu_2')\\
&= \ln \frac{\sum_i [p\mu_1(s_i)+(1-p)\mu_2(s_i)]x_i^*}{\sum_i [p\mu_1'(s_i)+(1-p)\mu_2'(s_i)]x_i^*}\\
&= \ln \frac{p\sum_i \mu_1(s_i)x_i^* + (1-p)\sum_i \mu_2(s_i)x_i^*}{p\sum_i \mu_1'(s_i)x_i^* + (1-p)\sum_i \mu_2'(s_i)x_i^*}\\
&\leq \ln \frac{e^{K_\otimes(bm_\otimes)(\mu_1,\mu_1')}p\sum_i \mu_1'(s_i)x_i^* + e^{K_\otimes(bm_\otimes)(\mu_2,\mu_2')}(1-p)\sum_i \mu_2'(s_i)x_i^*}{p\sum_i \mu_1'(s_i)x_i^* + (1-p)\sum_i \mu_2'(s_i)x_i^*}\\
&\leq \max\{K_\otimes(bm_\otimes)(\mu_1,\mu_1'), K_\otimes(bm_\otimes)(\mu_2,\mu_2')\}
\end{aligned}$$

in which the first inequality is obtained by the definition of $K_\otimes(bm_\otimes)$:

$$\ln \frac{\sum_i \mu_1(s_i)x_i^*}{\sum_i \mu_1'(s_i)x_i^*} \leq K_\otimes(bm_\otimes)(\mu_1,\mu_1')$$

and

$$\ln \frac{\sum_i \mu_2(s_i)x_i^*}{\sum_i \mu_2'(s_i)x_i^*} \leq K_\otimes(bm_\otimes)(\mu_2,\mu_2')$$

□

Theorem 4. *Let s, t, s', t' be probabilistic processes. Then*

1. $bm_\otimes(s;t, s';t') \leq bm_\otimes(s,s') + bm_\otimes(t,t')$
2. $bm_\otimes(s+t, s'+t') \leq \max(bm_\otimes(s,s'), bm_\otimes(t,t'))$
3. $bm_\otimes(s+_p t, s'+_p t') \leq \max(bm_\otimes(s,s'), bm_\otimes(t,t'))$
4. $bm_\otimes(s \mid t, s' \mid t') \leq bm_\otimes(s,s') + bm_\otimes(t,t')$
5. $bm_\otimes(s \parallel t, s' \parallel t') \leq bm_\otimes(s,s') + bm_\otimes(t,t')$

6. $bm_{\otimes}(s \parallel_p t, s \parallel_p t') \leq bm_{\otimes}(s, s') + bm_{\otimes}(t, t')$

Proof. Case 1 (sequential composition): We show below the proof of an intermediate result:

$$bm_{\otimes}(s;t, s';t) \leq bm_{\otimes}(s, s') \tag{11}$$

Using the intermediate result again for $s';t$ and $s';t'$, we will get

$$bm_{\otimes}(s';t, s';t') \leq bm_{\otimes}(t, t')$$

Applying the triangle inequality property of $bm_{\otimes}$, we will obtain the first clause as required.

The proof of the inequality (11) proceeds as follows: we construct a metric m as:

$$m(P, Q) = \begin{cases} bm_{\otimes}(s, s') & \text{if } P = s;t \text{ and } Q = s';t \\ 0 & \text{if } P = Q \\ \infty & \text{otherwise.} \end{cases} \tag{12}$$

and show that it satisfies $F_{\otimes}(m) \preceq m$, namely, it is a pre-fixpoint of $F_{\otimes}$. Remember that $bm_{\otimes}$ is the least one, thus we have $bm_{\otimes}(s;t, s';t) \leq m(s;t, s';t) = bm_{\otimes}(s, s')$ as required.

Since the case in which $P = s;t$ and $Q = s';t$ is most interesting, thus we focus on showing $F_{\otimes}(m)(s;t, s';t) \leq m(s;t, s';t)$, namely, for any $\epsilon \geq 0$, if $m(s;t, s';t) \leq \epsilon$, then $F_{\otimes}(m)(s;t, s';t) \leq \epsilon$.

By the transition rule of sequential composition, if $s;t \xrightarrow{a} \nu$ is due to $s \xrightarrow{\surd} \mu$ and $t \xrightarrow{a} \nu$, there exists also a tick action in $s';t \xrightarrow{a} \nu$ which then proceeds in t, certainly $K_{\otimes}(m)(\nu, \nu) = 0 \leq \epsilon$.

If $s;t \xrightarrow{a} \mu;\delta(t)$ is due to $s \xrightarrow{a} \mu$ and $a \neq \surd$, by $bm_{\otimes}(s, s') = m(s;t, s';t) \leq \epsilon$, there exist also a transition $s' \xrightarrow{a} \mu'$, $a \neq \surd$ and $K_{\otimes}(m)(\mu, \mu') \leq \epsilon$.

By the definition of $K_{\otimes}(m)$,

$$K_{\otimes}(m)(\mu;t, \mu';t') = \max |\ln \frac{\sum_i (\mu;t)(s_i;t)x_i}{\sum_i (\mu';t)(s_i;t)x_i}|$$

under the constraints: $\forall i, j.\ x_i \leq e^{m(s_i;t, s_j;t)} x_j$. Since for all i, $(\mu;t)(s_i;t) = \mu(s_i)$, $(\mu';t)(s_i;t) = \mu'(s_i)$ and $m(s_i;t, s_j;t) = bm_{\otimes}(s_i, s_j)$ as defined by the equation (12). It turns out that the primal program of $K_{\otimes}(m)(\mu;t, \mu';t')$ is the same as the primal program of $K_{\otimes}(m)(\mu, \mu')$. Thus $K_{\otimes}(m)(\mu;t, \mu';t) = K_{\otimes}(m)(\mu, \mu') \leq \epsilon$. It is analogous for the converse direction.

By the definition that the function $F_{\otimes}$ is the Hausdorff distance between the transitions of the states $s;t$ and $s';t$, we obtain $F_{\otimes}(m)(s;t, s';t) \leq \epsilon$ as required.

Case 2 (nondeterministic alternative composition): This case is trivial. If s can perform an action that t cannot, or if t can perform an action that s cannot, then the distance $bm_{\otimes}(s, t) = \infty$. Symmetrically also for s' and t'. In this case the inequality is trivially satisfied. Hence, we focus on the case that s and t, and s' and t', agree on the actions they can perform initially.

The process $s+t$ evolves either $s+t \xrightarrow{a} \mu$ if $s \xrightarrow{a} \mu$, or $s+t \xrightarrow{a} \nu$ if $t \xrightarrow{a} \nu$. From the metric bisimulation condition we get that there are distributions μ', ν' and $s' \xrightarrow{a} \mu'$ and $t' \xrightarrow{a} \nu'$ such that $bm_\otimes(s,s') \leq K_\otimes(\mu,\mu')$ and $bm_\otimes(t,t') \leq K_\otimes(\nu,\nu')$. The specification of the alternative composition allows to derive now also the transitions $s'+t' \xrightarrow{a} \mu'$ and $s'+t' \xrightarrow{a} \nu'$.

Symmetrically also for $s'+t'$. Since $bm_\otimes(s,s') \leq K_\otimes(\mu,\mu')$ and $bm_\otimes(t,s') \leq K_\otimes(\mu,\mu')$

Case 3 (probabilistic alternative composition):

$$\begin{aligned}
& bm_\otimes(s +_p t, s' +_p t') \\
\leq & \max_{a\in A} K_\otimes(bm_\otimes)(pD(s,a)+(1-p)D(t,a), pD(s',a)+(1-p)D(t',a)) \\
\leq & \max_{a\in A} \max(K_\otimes(bm_\otimes)(D(s,a),D(s',a)), K_\otimes(bm_\otimes)(D(t,a),D(t',a))) \qquad (Prop.\ 6) \\
= & \max(\max_{a\in A} K_\otimes(bm_\otimes)(D(s,a),D(s',a)), \max_{a\in A} K_\otimes(bm_\otimes)(D(t,a),D(t',a))) \\
= & \max(bm_\otimes(s,s'), bm_\otimes(t,t'))
\end{aligned}$$

Case 5 (asynchronous parallel composition): Analogous to Case 1, we only have to show that

$$bm_\otimes(s \parallel t, s' \parallel t) \leq bm_\otimes(s,s') \tag{13}$$

We sketch the proof of the inequality (13) as follows: construct a metric m as:

$$m(P,Q) = \begin{cases} bm_\otimes(s,s') & \text{if } P = s \parallel t \text{ and } Q = s' \parallel t \\ 0 & \text{if } P = Q \\ \infty & \text{otherwise.} \end{cases} \tag{14}$$

It is routine to verify that the constructed metric m satisfies $F_\otimes(m) \preceq m$, namely, it is a pre-fixpoint of $F_\otimes$. Remember that $bm_\otimes$ is the least one, thus we have $bm_\otimes(s \parallel t, s' \parallel t) \leq m(s \parallel t, s' \parallel t) = bm_\otimes(s,s')$ as required by the inequality (13).

Case 4 (parallel composition) and Case 6 (probabilistic parallel composition) are not difficult by following analogous means used in the previous cases, we leave the two cases to readers. □

Theorem 5. *Let s, t, s', t' be probabilistic processes. Then*

$$bm_V(s+t, s'+t') \leq \max(bm_V(s,s'), bm_V(t,t'))$$

Proof. We sketch the proof as follows. By Def. 1 we get that

$$F_V(bm_V)(s+t, s'+t') \leq \max\{F_V(bm_V)(s,s'), F_V(bm_V)(t,t')\}$$

Using the fact $bm_V = F_V(bm_V)$ completes the proof of the required result. □