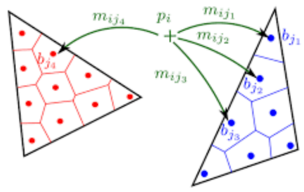
AGENCE NATIONALE DE LA RECHERCHE

ANR

REPAS

RELIABLE AND PRIVACY-AWARE SOFTWARE SYSTEMS



Deliverable D.4.a

Up-to techniques for Generalized Kantorovic Distances

1 Introduction

Bisimulation has played a fundamental role in the analysis and verification of traditional concurrent systems. In recent times, however, there is a growing tendency to consider probabilistic frameworks, partly to capture the random nature of interactions in distributed systems, partly to model and reason about protocols which make use of randomized mechanisms, such as those used in security and privacy. In this context, equivalences are not suitable, because they are not robust w.r.t. small variation of the transition probabilities, and they are usually replaced by (pseudo-)metrics: unlike an equivalence relation, a metric can vary smoothly as a function of the probabilities, and it can be used to measure the similarity of two systems in a more informative way than an equivalence relation.

Bisimulation metrics are particularly successful, especially in the area of concurrency, They can be defined by generalizing to metrics the bisimilarity "progress" relation; using a terminology introduced by Sangiorgi [12], we say that a relation between processes $\mathcal{R}$ *progresses to* $\mathcal{S}$ if for every pair of processes in $\mathcal{R}$, every transition from one process is matched by a transition from the other, and the derivative processes are related by $\mathcal{S}$. A bisimulation can then be defined as a relation that progresses to itself. Using the same terminology for probabilistic transitions, a metric d on states progresses to a metric l on distributions over states if, for all processes at d-distance ε, every transition from one process is matched by a transition from the other and the resulting distributions are at l-distance at most ε. Then d is a bisimulation metric if it progresses to its own *lifting* $K(d)$ on distributions.

Among the bisimulation metrics, those based on the Kantorovich lifting are the most popular. Originally proposed in the seminal works of Desharnais et al. [5, 6, 7] and of van Breugel and Worrel [13, 14], the traditional Kantorovich lifting has been extended in [3] so as to capture privacy properties such as *differential privacy* [8]. Part of their success is due to the *Kantorovich-Rubinstein duality*, which allows us to compute the lifting efficiently using linear programming algorithms [1, 13, 15, 16].

Analogously to the bisimilarity relation $\sim$, which is defined as the union of all bisimulations, the bisimilarity metric bm is defined coinductively as the smallest bisimulation metric. This means that we can extend the bisimulation proof method to metrics: given two processes P and Q, to prove $P \sim Q$ it is sufficient to find a bisimulation $\mathcal{R}$ such that $P \mathcal{R} Q$. Similarly, to show that $bm(P, Q) \leq \varepsilon$, it is sufficient to find a bisimulation metric d such that $d(P, Q) \leq \varepsilon$. The main difficulty in the bisimulation method is that the cost of naively checking that $\mathcal{R}$ is a bisimulation can be proportional to its size. Indeed, we need to prove that *for all pairs* of processes in $\mathcal{R}$, the derivatives of the matching transitions are still related by $\mathcal{R}$. Now, the size of bisimulations typically depends on the complexity of the underlying transition system, and if the transition system is unbounded, bisimulations are, in general, infinite sets. This difficulty translates immediately to the metric level: to prove that d is a bisimulation metric we need to prove that *for all pairs* of processes at d-distance ε, the distributions resulting from the matching transitions have $K(d)$-distance at most ε.

One well known and general approach, originally due to Milner [9], for reducing the sizes of bisimulations, is to represent them *up to* a different relation that identifies redundant pairs of process expressions. For instance, he showed that, when we consider the relation between the derivative processes, we can reason *modulo* bisimilarity. In other words, to prove $P \sim Q$ it is sufficient to find a relation $\mathcal{R}$ that relates P and Q, and that progresses to $\sim \mathcal{R} \sim$. In other words, if P' and Q' are the derivative processes, we do not need to show $P' \mathcal{R} Q'$, we only need to find a pair or processes P'' and Q'' such that $P' \sim P''$, $P'' \mathcal{R} Q''$, and $Q'' \sim Q'$. Such an $\mathcal{R}$ is called *bisimulation up to bisimilarity.* This technique was successively generalized by Sangiorgi [12], who introduced the notion of *bisimulation up to* $\mathcal{F}$, where $\mathcal{F}$ is a function from relations to relations. The idea is that $\mathcal{F}(\mathcal{R})$ contains the pairs of derivatives. The method is sound if, whenever $\mathcal{R}$ *progresses to* $\mathcal{F}(\mathcal{R})$, then $\mathcal{R} \subseteq \sim$. The paper also defines *respectfulness* for up-to techniques, later generalized as compatibility [11], which guarantees that it is sound to compose them with each other. The up-to techniques can be so effective that they may reduce the size of the relation to be checked from infinite to finite, and even, in some cases, to a singleton.

In this paper we aim at generalizing the up-to bisimulation method to the Kantorovich bisimulation metrics (in the extended version of [3]), thus enhancing the corresponding proof technique. The aim is to obtain a proof method that allows us to prove that $bm(P, Q) \leq \varepsilon$ by finding a metric d such that $d(P, Q) \leq \varepsilon$, and such that the set of pairs of processes for which we have to check the progress relation is relatively small. In other words, a metric d which gives maximal distance (and therefore the progress relation is verified trivially) between all processes except a small set. As an example, consider the following processes (from a probabilistic version of CCS):

$$A = a.([\tfrac{1}{2}]A \mid b \oplus [\tfrac{1}{2}]c) \qquad A' = a.([\tfrac{1}{2}]A' \mid b \oplus [\tfrac{1}{4}]c \oplus [\tfrac{1}{4}]d)$$

After performing an a-action, process A has one half probability of going back to itself, with the additional possibility of performing an action b in parallel, and one half probability of performing action c. Process A' behaves similarly to A, but with probability one fourth it performs action d instead of c. In order to prove that $bm(A, A') \leq \frac{1}{2}$, we should define a metric assigning distance one half not only to the pair (A, A'), but also to all pairs of the form $A \mid b^n$ and $A \mid b^n$, where b^n is the parallel composition of n instances of b, representing the pairs to be inspected after the action a is performed for the n-th time. Each of these pairs should then be proved to satisfy the bisimulation metric clauses. Using up-to techniques, we can prove that $bm(A, A') \leq \frac{1}{2}$ just by considering a (pre)metric assigning one half distance to (A, A'), and maximal distance to all other non-identical states. When A performs a, then A' replies with the same action and the (probabilistic) up-to-context technique guarantees that it is sound to directly use the distance on (A, A') in place of the distance on $(A \mid b, A' \mid b)$.

Plan of the paper Section 2 recalls some preliminary notions. Section 3 introduces some operators on premetrics and discusses some relevant properties of them. Section 4 presents the extension to metrics of the up-to techniques. Section 5 shows some examples of these techniques applied to probabilistic CCS and to the verification of differential privacy. Finally, Section 6 concludes. Some proofs were omitted for space reasons, they can be found in the appendix.

2 Preliminaries

Premetrics and metrics An (extended) *premetric* on a set X is a very relaxed form of metric, namely a function $m : X^2 \to [0, +\infty]$ satisfying only reflexivity ($m(x, x) = 0$). An (extended, pseudo) *metric* d on X is a premetric also satisfying symmetry ($d(x, y) = d(y, x)$) and the triangle inequality ($d(x, z) \leq d(x, y) + d(y, z)$). For simplicity we drop "extended" and "pseudo" but they are always implied; we denote by $\mathcal{M}(X), \mathcal{M}_d(X)$ the set of premetrics and metrics on X respectively. The *kernel* $\ker(m)$ of m is an equivalence relation on X relating elements at distance 0, i.e. $(x, y) \in \ker(m)$ iff $m(x, y) = 0$.

Premetrics $\mathcal{M}(X)$ bounded by some maximal distance $\top \in [0, \infty]$ form a complete lattice under element-wise ordering ($m \leq m'$ iff $m(x, y) \leq m'(x, y)$ for all x, y), with suprema and infima given by $(\bigvee A)(x, y) = \sup_{m \in A} m(x, y)$ and $(\bigwedge A)(x, y) = \inf_{m \in A} m(x, y)$. Note that the lattice depends on the choice of $\top$ – the value (possibly $+\infty$) assigned by the top premetric $\top_{\mathcal{M}(X)}$ to all distinct elements – which we generally leave implicit.

Metrics $\mathcal{M}_d(X)$ bounded by $\top$ also form a complete lattice under $\leq$, with the same supremum operator. On the other hand, the infimum operator, denoted by $\bigwedge_d$, is different since the inf of metrics is not necessarily a metric. Still, infima exist and can be obtained by $\bigwedge_d A = \bigvee(\downarrow_d A)$, where $\downarrow_d A = \{d \in \mathcal{M}_d(X) \mid \forall d' \in A : d \leq d' \}$.

Probabilistic automata, bisimilarity and metrics Let S be a countable set of *states*.[1] We denote by $\mathcal{P}(S)$ the set of all (discrete) probability measures Δ, Θ over S; the Dirac measure on s by $\delta(s)$. A *Probabilistic automaton* (henceforth PA) $\mathcal{A}$ is a tuple (S, A, D) where A is a countable set of action *labels*, and $D \subseteq S \times A \times \mathcal{P}(S)$ is a *transition relation.* We write $s \xrightarrow{\alpha} \Delta$ for $(s, \alpha, \Delta) \in D$, and define a family of functions $\to_\alpha : S \to 2^{\mathcal{P}(S)}$ as $\to_\alpha (s) = \{\Delta \mid s \xrightarrow{\alpha} \Delta\}$.

Let $R \subseteq S \times S$ be an equivalence relation on S; its lifting $\mathcal{L}(R)$ is an equivalence relation on $\mathcal{P}(S)$, defined as $(\Delta, \Theta) \in \mathcal{L}(R)$ iff Δ, Θ assign the same probability to all equivalence classes of R. *Probabilistic bisimilarity* $\sim$ can be defined as the largest equivalence relation R on S such that $(s, t) \in R$ and $s \xrightarrow{\alpha} \Delta$ imply $t \xrightarrow{\alpha} \Theta$ with $(\Delta, \Theta) \in \mathcal{L}(R)$.

Bisimilarity is a strong notion that often fails in probabilistic systems due to some "small" mismatch of probabilities. Hence, it is natural to define a

[1] A countable state space is assumed for simplicity; however, the proofs of several results do not rely on this assumption, and we expect those that do to be extendible to the continuous case.

metric that tells us "how much" different two states are, and such that its kernel coincides with $\sim$. Let $K : \mathcal{M}_d(S) \to \mathcal{M}_d(\mathcal{P}(S))$ be a lifting operator mapping metrics on S to metrics on distributions over S. A well known such operator is the *Kantorovich* lifting, but it is not unique: in fact, the Kantorovich itself can be generalized to a family of liftings, parametrized by an underlying distance (c.f. Section 3.2).

A metric $d \in \mathcal{M}_d(S)$ is a *bisimulation metric* if $d(s,t) < \top$ and $s \xrightarrow{\alpha} \Delta$ imply $t \xrightarrow{\alpha} \Theta$ with $K(d)(\Delta, \Theta) \leq d(s,t)$.[2] The *bisimilarity metric* bm can be defined as the $\bigwedge_d$ of all bisimulation metrics. Note that the lattice order of metrics has inverse meaning than the one of relations: a smaller metric corresponds to a larger relation.

It should be emphasized that, although $\sim$ is a uniquely defined relation, bm depends first on the choice of $\top$ and second, on the choice of the K operator. If $K, \mathcal{L}$ commute with ker, i.e. $\ker(K(d)) = \mathcal{L}(\ker(d))$ for all $d \in \mathcal{M}_d(S)$, it can be shown that $\sim = \ker(bm)$ [3]. In other words, we can have different metrics, all characterizing bisimilarity at their kernel, but which do not coincide on the distance they assign to non-bisimilar states.

Note that, although $\sim$ was defined as the union of all *equivalence* relations satisfying the bisimulation property, the "equivalence" requirement is only for convenience, so that the lifting $\mathcal{L}(R)$ has a simple form; we could obtain the same $\sim$ as the union of all *arbitrary* relations R satisfying the same property. The same is true for bm: although in the literature it is typically defined as the $\bigwedge_d$ of bisimulation *metrics*, we show in Section 4.1 that it can be constructed as the $\bigwedge$ of bisimulation *premetrics*. The advantage of using premetrics (resp. arbitrary relations) is that one has to construct a simpler bisimulation premetric m (resp. bisimulation relation R) not necessarily satisfying the triangle inequality (resp. transitivity), in order to bound the bisimilarity distance between two states.

3 Premetrics: operations and their properties

In this section we discuss various operations on premetrics and their properties. These will provide the technical building blocks for developing the up-to techniques in Section 4.

3.1 Lipschitz property and reverse maps

Lipschitz is a fundamental strong notion of continuity that plays a central role in all constructions of this work. A function $f : A \to B$ is *Lipschitz* (or nonexpansive) wrt the metrics m_A, m_B, written m_A, m_B-Lip, iff

$$m_B(f(a), f(a')) \leq m_A(a, a') \qquad \forall a, a' \in A$$

Tightly connected to this property is the *reverse map* on premetrics $f^{\leftarrow} : \mathcal{M}(B) \to \mathcal{M}(A)$ induced by $f : A \to B$, defined as $f^{\leftarrow}(m_B)(a, a') = m_B(f(a), f(a'))$.

Proposition 1. *The following hold:*

1. *f is m_A, m_B-Lip iff $m_A \geq f^{\leftarrow}(m_B)$.*

[2]Note that if $d(s,t) = \top$ (i.e. s, t are maximally "non-bisimilar") then $t \xrightarrow{a} \Theta$ is not required at all.

2. *$f^{\leftarrow}$ is monotone.*

3. *$f^{\leftarrow}$ preserves metrics: $m_B \in \mathcal{M}_d(B)$ implies $f^{\leftarrow}(m_B) \in \mathcal{M}_d(A)$.*

4. *$f^{\leftarrow}$ preserves $\bigwedge, \bigvee$, that is: $f^{\leftarrow}(\bigwedge M) = \bigwedge f^{\leftarrow}(M)$ and $f^{\leftarrow}(\bigvee M) = \bigvee f^{\leftarrow}(M)$.*

Note that, from the first property above, we have that $m_A = f^{\leftarrow}(m_B)$ is the smallest premetric such that f is m_A, m_B-Lip.

3.2 Generalized Kantorovich lifting

To construct metrics for probabilistic systems, as described in Section 2, one needs to lift (pre)metrics on the state space S to (pre)metrics on $\mathcal{P}(S)$. One well known such lifting is the Kantorovich metric, defined either via Lipschitz functions, or dually as a transportation problem. In [3] a generalization of this construction is given by extending the range of Lipschitz functions from $(\mathbb{R}, |\cdot|)$ to a generic metric space (V, d_V), where $V \subseteq \mathbb{R}$ is a convex subset of the reals and $d_V \in \mathcal{M}_d(V)$.

A function $f : S \to V$ can be lifted to a function $\hat{f} : \mathcal{P}(S) \to V$ by taking expectations: $\hat{f}(\Delta) = \int_S f d\Delta$. The requirement that V is convex ensures that $\hat{f}(\Delta) \in V$. Then, given a premetric $m \in \mathcal{M}(S)$, we can define a lifted metric $K(m) \in \mathcal{M}(\mathcal{P}(S))$ as:

$$K(m)(\Delta, \Theta) = \sup\{d_V(\hat{f}(\Delta), \hat{f}(\Theta)) \mid f \text{ is } m, d_V\text{-Lip}\}$$

The lifting K depends on the choice of (V, d_V) that we generally leave implicit: many results are given for any member of the family, while some state specific conditions on d_V. Note the difference between m, the premetric being lifted, and d_V, a parameter of the construction. Using the construction of Section 2, each member of the family gives rise to a different bisimilarity metric *bm*, and under mild assumptions it can be shown that all of them characterize bisimilarity at their kernel [3].[3]

Of particular interest is the classical Kantorovich $K_\oplus$, corresponding to $(V, d_V) = (\mathbb{R}, |\cdot|)$, and the *multiplicative* variant $K_\otimes$, corresponding to $(V, d_V) = ((0, +\infty), d_\otimes)$ where $d_\otimes(a, b) = |\ln a - \ln b|$. The corresponding bisimilarity metric obtained from the classical Kantorovich has been extensively studied; an important property of it is that $bm(s, t)$ is a bound on the total variation distance between the trace distributions originated from states s, t (a quantitative analogue of the fact that bisimilarity implies trace equivalence). The multiplicative Kantorovich provides the same bound, but for the multiplicative total variation distance, a metric of central importance to the area of differential privacy. Hence, the multiplicative variant provides a means for verifying privacy for concurrent systems.

Somewhat unexpectedly, it turns out that $K(m)$ is a proper metric, even if m itself is only a premetric: the metric properties of $K(m)$ come from those of d_V.

Proposition 2. *The following hold:*

1. *K is monotone.*

[3]Note that these "mild assumptions" are orthogonal to the results of this paper. If they are not satisfied, $\ker(bm)$ might be strictly included in $\sim$, without violating any of our results.

2. $K(m) \in \mathcal{M}_d(S)$ *(a proper metric) for all premetrics* $m \in \mathcal{M}(S)$*.*

Another interesting property of K concerns its relationship with $f^{\leftarrow}$. Given $f : A \to B$, let $f_* : \mathcal{P}(A) \to \mathcal{P}(B)$ denote the function mapping Δ to its *pushforward measure*, given by

$$f_*(\Delta)(Z) = \Delta(f^{-1}(Z)) \quad \text{for all measurable} \quad Z \subseteq B$$

Then, we can map metrics in $\mathcal{M}(B)$ to those in $\mathcal{M}(\mathcal{P}(A))$ by either applying $f^{\leftarrow}$ followed by K, or applying K followed by ${f_*}^{\leftarrow}$. The two options are related by the following result:

Proposition 3. *Let* $f : A \to B$ *and* $m_B \in \mathcal{M}(B)$*. Then* $(K \circ f^{\leftarrow})(m_B) \geq ({f_*}^{\leftarrow} \circ K)(m_B)$*.*

Due to the above result, K can be shown to preserve the Lip property (c.f. Section 3.4), which in turn is crucial for establishing the soundness of the up-to context techniques.

Dual form on premetrics The classical Kantorovich lifting can be dually expressed as a transportation problem. The primal and dual formulations are well-known to coincide on metrics; however, this is no longer the case when we work on premetrics. To see this, notice that in the transportation problem, the distance $K^d(m)(\delta(s), \delta(t))$ (where K^d denotes the dual Kantorovich) between two point distributions is exactly $m(s,t)$, in other words $\delta^{\leftarrow} \circ K^d = id_{\mathcal{M}(S)}$. On the other hand, $K(m)$ is always a metric, and it can be shown that $\delta^{\leftarrow} \circ K$ gives the metric closure operator.

Note that the dual forms of both the classical and the multiplicative Kantorovich are particularly useful since, in contrast to the primal form, they provide direct algorithms for computing the distance between finite distributions. Since the two forms no longer coincide, we should ensure that both of them are sound when used in the up-to techniques. For a general Kantorovich lifting K, let K^d be a monotone lifting that coincides with K on metrics. It can be shown that $K^d(m) \leq K(m)$ for all premetrics m, which in turn means that replacing K with K^d in the up-to techniques of Section 4 is sound.

3.3 Metric closure and chaining

A metric can be thought of as a generalization of an equivalence relation, since it satisfies reflexivity, symmetry and transitivity (in the form of the triangle inequality). Similarly to the equivalence closure, it is natural to define the *metric* closure $m^{\triangledown}$ of m: intuitively, the goal is to decrease m just enough to enforce the metric properties. Since $\mathcal{M}_d$ is a complete lattice, $m^{\triangledown}$ can be naturally defined as the greatest metric below m:

$$m^{\triangledown} = \bigvee(\mathcal{M}_d \cap {\downarrow} m)$$

It can be shown that $m \mapsto m^{\triangledown}$ is a closure operator whose fixpoints are exactly $\mathcal{M}_d(S)$.

Let $M^{\triangledown}$ denote the set $\{m^{\triangledown} \mid m \in M\}$. We can show that metric closure commutes with the infima of the two lattices.

Proposition 4. *Let $M \subseteq \mathcal{M}$. Then $\bigwedge_d(M^\triangledown) = (\bigwedge M)^\triangledown$.*

This, in turn, means that the metric infimum $\bigwedge_d$ can be obtained by the premetric infimum followed by metric closure, that is: $\bigwedge_d D = (\bigwedge D)^\triangledown$ for $D \subseteq \mathcal{M}_d(S)$. Based on this, we extend the $\bigwedge_d$ operator to premetrics, defined as $\bigwedge_d M = (\bigwedge M)^\triangledown$.

Finally, we can define the *chaining* $m_1 \curlywedge m_2$ of two premetrics as:

$$(m_1 \curlywedge m_2)(s_1, s_2) = \inf_{t \in S}(m_1(s_1, t) + m_2(t, s_2))$$

Chaining combines two premetrics by passing through some midway point, and will be used as a primitive block for constructing up-to techniques in Section 4.

Proposition 5. *The following hold:*

1. *$\curlywedge$ is associative and monotone on both arguments*
2. *$m_1 \wedge_d m_2 \leq m_1 \curlywedge m_2 \leq m_1 \wedge m_2$*
3. *$K(m_1 \curlywedge m_2) \leq K(m_1) \curlywedge K(m_2)$*

3.4 Operations that preserve Lipschitz

The Lipschitz property plays a central role in all constructions of this work, since both the Kantorovich lifting and the notion of progression depend on it. The following operations preserving this property will play a crucial role in the up-to techniques developed in Section 4.

Let $f : A \to B$ and assume it is m_A, m_B-Lip. Moreover, let $M_A = \{m_A^i\}_{i \in I}$ and $M_B = \{m_B^i\}_{i \in I}$ such that f is m_A^i, m_B^i-Lip for all $i \in I$. The following hold:

1. Inc/dec-reasing the source/target metric: f is m'_A, m'_B-Lip $\quad \forall m'_A \geq m_A, m'_B \leq m_B$
2. Infima and suprema: f is $\bigvee M_A, \bigvee M_B$-Lip and $\bigwedge M_A, \bigwedge M_B$-Lip
3. Metric closure: f is ${m_A}^\triangledown, {m_B}^\triangledown$-Lip
4. Kantorovich lifting: f_* is $K(m_A), K(m_B)$-Lip

Note that the property (3) above implies that $K(m) = K(m^\triangledown)$ since the sup in the definition of K for both sides ranges over the same set of functions.

3.5 Convex and quasiconvex premetrics

If X is a convex set then X^2 can be also viewed as a convex set of vectors (x, y), where $\sum_i \lambda_i(x_i, y_i) = (\sum_i \lambda_i x_i, \sum_i \lambda_i y_i)$ for all λ_i's such that $\sum_i \lambda_i = 1$. This allows us to talk about the convexity of a premetric jointly on both arguments. We say that $m \in \mathcal{M}(X)$ is:

- *convex* iff $m(\sum_i \lambda_i(x_i, y_i)) \leq \sum_i \lambda_i m(x_i, y_i)$
- *quasiconvex* iff $m(\sum_i \lambda_i(x_i, y_i)) \leq \max_i m(x_i, y_i)$

Note that there exist several distinct abstract notions of convexity for general metric spaces, here (quasi)convexity is used in the usual sense of (quasi)convex functions.

The set $\mathcal{P}(S)$ is convex and so is V used in the construction of the Kantorovich lifting. It can be shown that if d_V is convex (resp. quasiconvex) then $K(m)$ is also convex (resp. quasiconvex) for all $m \in \mathcal{M}(S)$. As a consequence, the classical Kantorovich $K_\oplus(m)$ is convex (since $|\cdot|$ is convex), while the multiplicative variant $K_\otimes(m)$ is quasiconvex (since $d_\otimes$ is quasiconvex).

4 Up-to techniques

In this section, we extend to the metric case the theory of up-to techniques presented in [12]. All the constructions assume some fixed underlying PA, which could be produced by a process calculus like the probabilistic CCS of Section 5. In what follows, we use l to denote premetrics on $\mathcal{P}(S)$.

4.1 Progressions

For a relation $\mathcal{R}$ on states of a non-probabilistic automaton, bisimulation can be defined in terms of progressions. A relation $\mathcal{R}$ progresses to $\mathcal{R}'$, denoted by $\mathcal{R} \rightarrowtail \mathcal{R}'$, if whenever $s \,\mathcal{R}\, t$ and $s \xrightarrow{\alpha} s'$ then $t \xrightarrow{\alpha} t'$ and $s' \,\mathcal{R}'\, t'$, and vice versa. A bisimulation can be thereby defined as a relation that progresses to itself, i.e. $\mathcal{R} \rightarrowtail \mathcal{R}$.

An important difference in the probabilistic case is that progressions have different source and target domains. A premetric m on S (the source premetric) progresses to a premetric l on $\mathcal{P}(S)$ (the target premetric).

Definition 1. *Given $m \in \mathcal{M}(S), l \in \mathcal{M}(\mathcal{P}(S))$ we say that m progresses to l, written $m \rightarrowtail l$, iff $m(s,t) < \top$ implies that:*

- *whenever $s \xrightarrow{\alpha} \Delta$ then $t \xrightarrow{\alpha} \Theta$ with $l(\Delta, \Theta) \leq m(s,t)$*
- *whenever $t \xrightarrow{\alpha} \Theta$ then $s \xrightarrow{\alpha} \Delta$ with $l(\Delta, \Theta) \leq m(s,t)$*

Using the Hausdorff metric, progression can be written as a Lipschitz property:[4]

$$m \rightarrowtail l \quad \textit{iff} \quad \forall \alpha : \rightarrow_\alpha \ \textit{is } m, H(l)\textit{-Lip}$$

From the results about operations preserving Lipschitz, and the fact that Hausdorff is monotone, we obtain the following useful properties of the progress relation:

- $m \rightarrowtail l$ implies $m' \rightarrowtail l'$ for all $m' \geq m, l' \leq l$.
- Let $d \in \mathcal{M}_d(\mathcal{P}(S))$. Then $m \rightarrowtail d$ implies $m^\nabla \rightarrowtail d$.
- Let $m = \bigwedge_i m_i$ and $l = \bigwedge l_i$ such that for all i: $m_i \rightarrowtail l_i$. Then $m \rightarrowtail l$.

[4] We could also define progression as a Lipschitz property of a *single* function $\rightarrow(s) = \{(\alpha, \Delta) \mid s \xrightarrow{\alpha} \Delta\}$.

From the definition of bisimulation (pre)metrics (Section 2), we have that $m \in \mathcal{M}(S)$ is a bisimulation (pre)metric iff $m \rightarrowtail K(m)$. The bisimilarity metric is traditionally defined as the $\bigwedge_d$ of all bisimulation metrics. Since metric closure preserves the Lip property, it also preserves the bisimulation property, which means that we can equivalently obtain bm as the $\bigwedge$ of all bisimulation premetrics. m is a bisimulation premetric iff $m^\triangledown$ is a bisimulation metric. Hence: $bm = \bigwedge_d \{d \in \mathcal{M}_d(S) \mid d \rightarrowtail K(d)\} = \bigwedge \{m \in \mathcal{M}(m) \mid m \rightarrowtail K(m)\}$

Proof. Assuming that m is a bisimulation premetric, we have that $\rightarrow_\alpha$ is $m, H(K(m))$-Lip for all α. Since $H(K(m))$ is a metric, from Theorem 3.4 we get that $\rightarrow_\alpha$ is $m^\triangledown, H(K(m))$-Lip and since $K(m^\triangledown) = K(m)$ we get that $\rightarrow_\alpha$ is $m^\triangledown, H(K(m^\triangledown))$-Lip which implies that $m^\triangledown$ is a bisimulation metric. $\square$

4.2 $\mathcal{F}$ functions, soundness, respectfulness

We can define an up-to technique using a function $\mathcal{F}$ on $\mathcal{M}(\mathcal{P}(S))$. Ideally, for a premetric m on states, we want to allow the distance $\mathcal{F}(K(m))(\Delta, \Theta)$ to be used instead of $K(m)(\Delta, \Theta)$ in a bisimulation proof, since a bound to $\mathcal{F}(K(m))$ could be easier to compute. Therefore, we consider progressions of the form $m \rightarrowtail \mathcal{F}(K(m))$, where $\mathcal{F} : \mathcal{M}(\mathcal{P}(S)) \to \mathcal{M}(\mathcal{P}(S))$.

Definition 2. *A function $\mathcal{F} : \mathcal{M}(\mathcal{P}(S)) \to \mathcal{M}(\mathcal{P}(S))$ is sound if $m \rightarrowtail \mathcal{F}(K(m))$ implies $bm \leq m$.*

Hence, if $\mathcal{F}$ is a sound function then a bisimulation premetric up-to $\mathcal{F}$ allows us to derive upper-bounds to the distance between two states. At the same time, using $\mathcal{F}$ in the target metric allows us to simplify the proof that the states actually satisfy these bounds.

Respectful functions Given a function $\mathcal{F} : \mathcal{M}(\mathcal{P}(S)) \to \mathcal{M}(\mathcal{P}(S))$, one can prove that it is a sound up-to technique by means of a direct proof. However, it is known that the composition of sound functions on relations is not necessarily a sound function, and the standard counterexamples apply to the metric setting as well. In the non-probabilistic case, this has led to the definition of "respectfulness": an up-to function $\mathcal{F}$ on relations is respectful if whenever $\mathcal{R} \rightarrowtail \mathcal{R}'$ and $\mathcal{R} \subseteq \mathcal{R}'$, then $\mathcal{F}(\mathcal{R}) \rightarrowtail \mathcal{F}(\mathcal{R}')$ and $\mathcal{F}(\mathcal{R}) \subseteq \mathcal{F}(\mathcal{R}')$. Respectfulness implies soundness and at the same time is closed under composition [12].
On metrics, the definition of respectfulness must take care of the fact that the source and target metrics have different domains, and that the function $\mathcal{F}$ is defined on the domain $\mathcal{P}(S)$ of the target metric. Hence, a "corresponding" function $\mathcal{G} : \mathcal{M}(S) \to \mathcal{M}(S)$ on the source metric has to be defined. Instead of constructing a specific such $\mathcal{G}$, we only assume its existence and that it "plays well" with $\mathcal{F}$ and K, meaning that $(K \circ \mathcal{G})(m) \leq (\mathcal{F} \circ K)(m)$. A concrete $\mathcal{G}$ is then chosen in the respectfulness proof of each up-to technique $\mathcal{F}$.

Definition 3. *A function $\mathcal{F} : \mathcal{M}(\mathcal{P}(S)) \to \mathcal{M}(\mathcal{P}(S))$ is* respectful *iff it is monotone and there exists $\mathcal{G} : \mathcal{M}(S) \to \mathcal{M}(S)$ such that for all $m, m' \in \mathcal{M}(S)$:*

- $(K \circ \mathcal{G})(m) \leq (\mathcal{F} \circ K)(m)$
- $m \rightarrowtail K(m')$ *and* $m \geq m'$ *imply* $\mathcal{G}(m) \rightarrowtail K(\mathcal{G}(m'))$ *and* $\mathcal{G}(m) \geq \mathcal{G}(m')$

Any respectful function is sound.

Proof. Let $\mathcal{F}$ be respectful and let $\mathcal{G}$ be its corresponding source map from the definition of respectfulness. Assume that $m \rightarrowtail \mathcal{F}(K(m))$. Analogously to the proof in [12], we define a sequence of metrics $m_n, n \geq 0$ as: $m_0 = m$ and $m_{n+1} = \mathcal{G}(m_n) \wedge m_n$. By construction, $m_n \geq m_{n+1}$ for all $n \geq 0$. We now show that $m_n \rightarrowtail K(m_{n+1})$ for all $n \geq 0$ For the base case $n = 0$, from the respectfulness of $\mathcal{F}$ and the monotonicity of K we have that $\mathcal{F}(K(m)) \geq K(\mathcal{G}(m)) \geq K(\mathcal{G}(m) \wedge m)$. Hence $m \rightarrowtail \mathcal{F}(K(m))$ implies $m_0 = m \rightarrowtail K(\mathcal{G}(m) \wedge m) = K(m_1)$. For the inductive step, we want to show that $m_{n+1} \rightarrowtail K(m_{n+2})$, that is, $\mathcal{G}(m_n) \wedge m_n \rightarrowtail K(\mathcal{G}(m_{n+1}) \wedge m_{n+1})$. We have that:

$$\begin{array}{rcr}
 & m_n \rightarrowtail K(m_{n+1}) & \text{induction hypothesis} \\
\Rightarrow & \mathcal{G}(m_n) \rightarrowtail K(\mathcal{G}(m_{n+1})) & \text{respectfulness, } m_n \geq m_{n+1} \\
\Rightarrow & \mathcal{G}(m_n) \wedge m_n \rightarrowtail K(\mathcal{G}(m_{n+1})) \wedge K(m_{n+1}) & \wedge \text{ preserves } \rightarrowtail \\
\Rightarrow & \mathcal{G}(m_n) \wedge m_n \rightarrowtail K(\mathcal{G}(m_{n+1}) \wedge m_{n+1}) & K(a \wedge b) \leq K(a) \wedge K(b)
\end{array}$$

Since progressions are closed under infima, $\bigwedge_{n\geq 0} m_n \rightarrowtail K(\bigwedge_{n\geq 0} m_n)$. Hence, $\bigwedge_{n\geq 0} m_n$ is a bisimulation metric, and $m \geq \bigwedge_{n\geq 0} m_n$, which concludes the proof. □

4.2.1 Composing up-to techniques

The advantage of the respectfulness condition is that it makes it possible to derive the soundness of a composed up-to function just by proving the respectfulness of its components. We present here three operations that preserve respectfulness: function composition, function chaining, and taking the infimum of a set of functions (these operations respectively correspond to composition, chaining and union in the relational case).

The composition of respectful functions is respectful.

The theorem is proved by showing that, given two respectful functions $\mathcal{F}_1, \mathcal{F}_2$ and their corresponding source maps $\mathcal{G}_1, \mathcal{G}_2$ from the definition of respectfulness, $\mathcal{F} = \mathcal{F}_1 \circ \mathcal{F}_2$ and $\mathcal{G} = \mathcal{G}_1 \circ \mathcal{G}_2$ satisfy the requirements of respectfulness.

The chaining of up-to functions is defined using the $\curlywedge$ operator from Section 4.2.1. Define the chaining of two functions $\mathcal{F}_1, \mathcal{F}_2$ as $(\mathcal{F}_1 \curlywedge \mathcal{F}_2)(m) = \mathcal{F}_1(m) \curlywedge \mathcal{F}_2(m)$. Using the properties of $\curlywedge$ proved in Proposition 5, we derive the following result.

The chaining of respectful functions is respectful.

Analogously to chaining, define the infimum of a countable set of functions $\bigwedge\{\mathcal{F}_i\}$ as $\bigwedge\{\mathcal{F}_i\}(m) = \bigwedge\{\mathcal{F}_i(m)\}$. Given a countable set $\{\mathcal{F}_i\}$ of respectful functions with corresponding source maps $\{\mathcal{G}_i\}$, we prove that the function $\bigwedge\{\mathcal{F}_i\}$ is respectful by using the source map $\bigwedge\{\mathcal{G}_i\}$.

The infimum of a set of respectful functions is respectful.

4.2.2 Up-to bisimilarity metric and up-to (quasi)convexity

The respectfulness (and soundness) of up-to techniques such as up-to-bisimilarity-metric can now be recovered by applying the operations presented in Section 4.2.1 to basic respectful functions. The identity $\mathcal{F}_{id}(l) = l$ and the constant-to-bm $\mathcal{F}_{bm}(l) = K(bm)$ functions are respectful. The result directly follows from the

definition: for the first we take $\mathcal{G}_{id}(m) = m$, for the second $\mathcal{G}_{bm}(m) = bm$. The up-to-bisimilarity-metric function can be now simply constructed as $\mathcal{F}_{bm} \curlywedge \mathcal{F}_{id} \curlywedge \mathcal{F}_{bm}$, and it is respectful as the chaining of respectful functions is (Theorem 4.2.1). By Theorem 4.2.1, we can also derive the respectfulness of the up-to-triangle-inequality function (corresponding to the up-to-transitive-closure technique on relations), defined as $\bigwedge\{\curlywedge^n \mathcal{F}_{id}\}_{n\geq 1}$, where $\curlywedge^n \mathcal{F}_{id}$ is the chaining of $\mathcal{F}_{id}$ with itself n-times.
Another useful proof technique consists in the possibility of splitting probability distributions into components with common factors, and then only consider the (possibly weighted) distances between the components. Define the up-to-quasiconvexity and the up-to-convexity functions as follows:

- $\mathcal{F}_{qcv}(l)(\Delta, \Theta) = \inf\{\max_i l(\Delta_i, \Theta_i) | \, \Delta = \sum_i p_i \Delta_i \text{ and } \Theta = \sum_i p_i \Theta_i\}$
- $\mathcal{F}_{cv}(l)(\Delta, \Theta) = \inf\{\sum_i p_i l(\Delta_i, \Theta_i) | \, \Delta = \sum_i p_i \Delta_i \text{ and } \Theta = \sum_i p_i \Theta_i\}$

The respectfulness of the above up-to techniques depends on the (quasi)convexity of the Kantorovich operator. The following result is derived using the identity $\mathcal{G}_{id}$ as a source map.

If K is quasiconvex (resp. convex) then $\mathcal{F}_{qcv}$ (resp. $\mathcal{F}_{cv}$) is respectful.

4.3 Faithful contexts

With up-to context techniques, common contexts in the probability distributions reached in the bisimulation game are allowed to be safely removed. Given a set of states S, a context is a function $C : S \to S$. As usual, we write $C[s]$ to denote the image of s under C. We look at states in S as defined by a language whose terms are syntactically finite expressions, which justifies the following assumption: for any class $\mathcal{C}$ of contexts, there is only a finite number of states s' such that $s = C[s']$ for some $C \in \mathcal{C}$.

Definition 4. *Given a class of contexts $\mathcal{C}$, a premetric m is closed under $\mathcal{C}$ iff C is m, m-Lip for all $C \in \mathcal{C}$. The closure of m under $\mathcal{C}$, denoted by $\mathcal{C}(m)$, is defined as the greatest premetric below m that is closed under $\mathcal{C}$:*

$$\mathcal{C}(m) = \bigvee\{m' \leq m \mid m' \textit{ is closed under } \mathcal{C}\}$$

Let $\mathcal{C}_* = \{C_* \mid C \in \mathcal{C}\}$. The up-to faithful context function $\mathcal{F}_{\mathcal{C}}$ is defined as: $\mathcal{F}_{\mathcal{C}}(l) = \mathcal{C}_*(l)$.

Since the Lipschitz property is preserved by $\bigvee$ (Thm 3.4), it is easy to show that $\mathcal{C}(m)$ itself is closed under $\mathcal{C}$, that is, $\mathcal{C}(m)(C[s], C[t]) \leq \mathcal{C}(m)(s, t) \leq m(s, t)$ for all $C \in \mathcal{C}$. Moreover, it follows from Thm 3.4 that K preserves the closure under $\mathcal{C}$. Hence, $K(\mathcal{C}(m))$ is always closed under $\mathcal{C}_*$: for all $C \in \mathcal{C}$, $K(\mathcal{C}(m))(C_*[\Delta], C_*[\Theta]) \leq K(\mathcal{C}(m))(\Delta, \Theta) \leq K(m)(\Delta, \Theta)$.
The function $\mathcal{C}(m)$ (respectively: $\mathcal{C}_*(l)$) can be alternatively characterized by considering the infimum value of m when a common context is removed from two terms (respectively: from two distributions). The context closure $(s, t)^{\mathcal{C}}$ of the pair (s, t) is the set of all pairs of terms of the form $(C[s], C[t])$, for $C \in \mathcal{C}$. The context closure $(\Delta, \Theta)^{\mathcal{C}_*}$ is extended to probability distributions using the set of contexts $C_* \in \mathcal{C}_*$.

The functions $\mathcal{C}$ and $\mathcal{C}_*$ can be alternatively characterized as follows:

1. $\mathcal{C}(m)(s,t) = \inf\{m(s',t') \mid (s,t) \in (s',t')^{\mathcal{C}}\}$

2. $\mathcal{C}_*(l)(\Delta,\Theta) = \inf\{l(\Delta',\Theta') \mid (\Delta,\Theta) \in (\Delta',\Theta')^{\mathcal{C}_*}\}$

In what follows, we often write $C[\Delta]$ to denote $C_*[\Delta]$.

Instead of directly proving soundness (or respectfulness) for up-to context functions $\mathcal{F}_{\mathcal{C}}$ where $\mathcal{C}$ are contexts of a specific language, we follow [12] and define the class of faithful contexts. Faithfulness only depends on general properties of the semantics of the contexts, and the up-to-faithful-context function is respectful whenever used with a quasiconvex Kantorovich operator (Theorem 4.3). In Section 5, the contexts of a probabilistic extension of CCS are proved to satisfy the condition of faithfulness.

Definition 5. *A context class $\mathcal{C}$ is faithful if whenever $C \in \mathcal{C}$, all transitions of $C[s]$ are of the form $C[s] \xrightarrow{\alpha} \sum_i p_i C_i[\Delta]$, where $C_i \in \mathcal{C}$ and either*

1. *$\Delta = \delta(s)$ and $\forall t$: $C[t] \xrightarrow{\alpha} \sum_i p_i C_i[\delta(t)]$, or*

2. *$s \xrightarrow{\alpha'} \Delta$ and $\forall t$: if $t \xrightarrow{\alpha'} \Theta$ then $C[t] \xrightarrow{\alpha} \sum_i p_i C_i[\Theta]$.*

We can now prove the respectfulness of $\mathcal{F}_{\mathcal{C}}$, assuming that the Kantorovich operator is quasiconvex. The reason for this extra condition is that faithfulness allows contexts to be probabilistic, meaning that when a transition is performed, the common context can be split into a weighted sum of contexts. Quasiconvexity then allows us to establish a bound to the distances between weighted sums of distributions with a common contexts (e.g., $\sum_i p_i C_i[\Delta']$ and $\sum_i p_i C_i[\Theta']$) based on the bounds of the components, which now are of the desired form ($C_i[\Delta']$ and $C_i[\Theta']$).

If K is quasiconvex then $\mathcal{F}_{\mathcal{C}}$ is respectful.

Proof. The monotonicity of $\mathcal{F}_{\mathcal{C}}$ comes directly from the definition of $\mathcal{C}(m)$. Let $\mathcal{G}(m) = \mathcal{C}(m)$, we prove that $\mathcal{G}$ is the source map required by the definition of respectfulness:

1. we prove $K(\mathcal{G}(m)) \leq \mathcal{F}_{\mathcal{C}}(K(m))$. From $\mathcal{G}(m) \leq m$ we derive $K(\mathcal{G}(m)) \leq K(m)$, and since $\mathcal{G}(m)$ is closed under $\mathcal{C}$ and K preserves closedness, then $K(\mathcal{G}(m))$ is closed under $\mathcal{C}_*$. Finally, $\mathcal{F}_{\mathcal{C}}(K(m))$ is the greatest premetric below $K(m)$ that is is closed under $\mathcal{C}_*$, from which the result follows;

2. suppose $m \rightarrowtail K(m')$ and $m \geq m'$. Then $\mathcal{G}(m) \geq \mathcal{G}(m')$ comes from the monotonicity of $\mathcal{C}(m)$, and it remains to prove that $\mathcal{G}(m) \rightarrowtail K(\mathcal{G}(m'))$. We first show that

   $\star$ for any faithful context C, $C[s] \xrightarrow{\alpha} \Delta$ implies that, for all t, if $m(s,t) < \top$ then $C[t] \xrightarrow{\alpha} \Theta$ with $K(\mathcal{G}(m'))(\Delta,\Theta) \leq m(s,t)$

   by considering the two cases of the definition of respectfulness and using quasiconvexity to derive the result. Since a term has only a finite number of subterms, by Theorem 4.3 we have $\mathcal{G}(m)(s,t) = m(s',t')$ for some s',t' and C faithful such that $s = C[s']$ and $t = C[t']$. Hence, by property $\star$ we have that $\mathcal{G}(m) \rightarrowtail K(\mathcal{G}(m'))$.

$\square$

$$
\frac{}{a.\oplus_i [p_i]P_i \xrightarrow{\alpha} \sum_i p_i \delta(P_i)} \qquad \frac{P \xrightarrow{\alpha} \Delta}{P + Q \xrightarrow{\alpha} \Delta} \qquad \frac{P \xrightarrow{\alpha} \Delta}{P \,|\, Q \xrightarrow{\alpha} \Delta \,|\, \delta(Q)}
$$

$$
\frac{P \xrightarrow{\alpha} \Delta \qquad Q \xrightarrow{\bar{\alpha}} \Theta}{P \,|\, Q \xrightarrow{\tau} \Delta \,|\, \Theta} \qquad \frac{P \xrightarrow{\alpha} \Delta \qquad \alpha \neq a, \bar{a}}{(\nu a)P \xrightarrow{\alpha} \Delta} \qquad \frac{P \xrightarrow{\alpha} \Delta \qquad A = P}{A \xrightarrow{\alpha} \Delta}
$$

Figure 1: Structured Operational Semantics for pCCS

5 Up-to techniques for probabilistic CCS

The conditions of faithfulness are quite general and can be instantiated by several varieties of probabilistic languages. We consider here CCS with a probabilistic choice operator and prove that its unary contexts (i.e., terms with a single hole, occurring only once) are faithful. The terms of pCCS are defined by the following grammar:

$$
P, Q ::= \mathbf{0} \;\Big|\; \alpha.\oplus_i [p_i]P_i \;\Big|\; P + Q \;\Big|\; P \,|\, Q \;\Big|\; (\nu a)P \;\Big|\; A
$$

where $\alpha ::= a, \bar{a}, \tau$ is an action label, for some underlying set of labels such that $a \in Act$ iff $\bar{a} \in Act$, and $\bar{\bar{\alpha}} = \alpha$ for $\alpha \in Act$, where $\tau \notin Act$. The semantics is given by the rules in Figure 1, where the parallel composition of distributions Δ, Θ on pCCS terms is defined by $\Delta \,|\, \Theta(P) = \Delta(P_1) \cdot \Theta(P_2)$ if $P = P_1 \,|\, P_2$, and 0 otherwise. The symmetric rules for the nondeterministic choice and parallel composition are omitted. We assume that every constant A of the language is defined by an equation $A = P$ for some pCCS process P where A may occur guarded. When the distribution following an action label is a point distribution, the $\oplus_i$ is omitted.

The (unary) contexts of pCCS are faithful.

Theorem 5 is proved by induction on the structure of the contexts. Since the up-to context technique is respectful for faithful contexts (Theorem 4.3), it follows from Theorem 5 that the up-to context function $\mathcal{F}_\mathcal{C}$ where $\mathcal{C}$ is the set of pCCS contexts is respectful.

Example 1. *Let A and A' be the* pCCS *constants defined in the introduction. We prove that their distance in the bisimilarity metric $bm_\oplus$, based on the standard Kantorovich lifting $K_\oplus$ and with $\top = 1$, is bounded by $\frac{1}{2}$. Define the premetric m on* pCCS *terms as follows: $m(A, A') = \frac{1}{2}$ and, for all P, Q different from A, A', m is the discrete metric, i.e., $m(P, Q) = 0$ if $P = Q$ and $m(P, Q) = 1$ otherwise. We prove that m is a bisimulation premetric up-to $(\mathcal{F}_{cv} \circ \mathcal{F}_\mathcal{C}) \curlywedge \mathcal{F}_{id}$, i.e., the chaining of the up-to-convexity-and-context function with the up-to-identity function.*
Suppose that A moves (the case when A' moves is symmetrical). If $A \xrightarrow{a} \Delta = \frac{1}{2} \cdot \delta(A) + \frac{1}{2} \cdot \delta(c)$, then $A' \xrightarrow{a} \Delta' = \frac{1}{2} \cdot \delta(A') + \frac{1}{4} \cdot \delta(c) + \frac{1}{4} \cdot \delta(d)$. Define $\Delta'' = \frac{1}{2} \cdot \delta(A') + \frac{1}{2} \cdot \delta(c)$. Then:

$$
\begin{aligned}
((\mathcal{F}_{cv} \circ \mathcal{F}_\mathcal{C}) \curlywedge \mathcal{F}_{id})(K_\oplus(m))(\Delta, \Delta') &\leq (\mathcal{F}_{cv} \circ \mathcal{F}_\mathcal{C})(K_\oplus(m))(\Delta, \Delta'') + (K_\oplus(m))(\Delta'', \Delta') \\
&\leq \tfrac{1}{2} \cdot (K_\oplus(m))(\delta(A), \delta(A')) + (K_\oplus(m))(\Delta'', \Delta') \\
&\leq \tfrac{1}{4} + \tfrac{1}{4}
\end{aligned}
$$

Note that the same premetric and the same proof can be applied when an arbitrary pCCS *process P is substituted to b in the definition of the constants A, A'.*

Finally, we give an example to illustrate how the generalized Kantorovich lifting captures differential privacy, and how the techniques developed in this paper can help to verify this property. Following [3], we model differential privacy in pCCS as a bound e^{ε} on the ratio between the probability that a process P produce a set of traces ψ, and the probability that an "adjacent" process P' produce the same set ψ, for any ψ. In [3] it is shown that in order to establish this property it is sufficient to show that $bm_{\otimes}(P, P') \leq \varepsilon$, where $bm_{\otimes}$ is defined based on the multiplicative Kantorovich $K_{\otimes}$ and $\top = +\infty$.

In the example, we consider a database D containing medical information relative to (at most) n patients. We assume that we are interested in obtaining statistical information about a certain disease, and that for this purpose we are allowed to ask queries like "how many patients are affected by the disease". Queries of this kind are called *counting queries* and it is well known that they can be sanitized, i.e. made ε-differentially private, by adding *geometric noise* to the real answer, namely a noise distribution $p_y(z) = c_z e^{|z-y|\varepsilon}$, where y is the real answer, z is the reported answer (ranging between 0 and n), and c_z is a normalization constant that depends only on z. Another database D' is *adjacent* to D if it differs from D for only one record (i.e., one patient). Clearly, the (sanitized) answers to the above query in two adjacent databases will differ by at most 1, and it is easy to see that the ratio between $p_{y+1}(z)$ and $p_y(z)$ is at most e^{ε}, which proves that ε-differential privacy is satisfied by the geometrical-noise method.

Example 2. *Consider the adjacent databases D, D' where y and $y+1$ patients are affected by the disease, respectively. We model D and D' in* pCCS *as*

$$D = q. \oplus_{z=0}^{n} [p_y(z)]\bar{v}_z.D \qquad\qquad D' = q. \oplus_{z=0}^{n} [p_{y+1}(z)]\bar{v}_z.D'$$

where the prefix q represents the acceptance of a query request, and the action $\bar{v}_z$ represents the delivery of the reported answer. Consider now a process Q that queries the database. This can be defined as $Q = \bar{q}. +_{z=0}^{n} v_z.\bar{w}_z$, where $+_{z=0}^{n} P_z$ denotes the nondeterministic choice $P_0 + P_2 + ... + P_n$. It is possible to prove that the processes $D \,|\, Q$ and $D' \,|\, Q$ satisfy ε-differential privacy, by proving that $bm_{\otimes}(D \,|\, Q, D' \,|\, Q) \leq \varepsilon$.

What we want to prove now is that the level of differential privacy decreases linearly with the number of queries (this is a well-known fact, the interest here is to show it using up-to techniques). Namely that if we define the processes P and P' as the parallel composition of i instances of Q and D and D' respectively, then $K_{\otimes}(P, P') \leq i\varepsilon$ We prove this for the case $i = 2$. Define the premetric m as $m(D \,|\, Q \,|\, Q, D' \,|\, Q \,|\, Q) = 2\varepsilon$, and as the discrete metric on all other pairs. The interesting case is when D (symmetrically: D') synchronizes with one of the queries. Suppose that $D \,|\, Q \,|\, Q \xrightarrow{\tau} \Delta$, with $\Delta = \sum_{z=0}^{n} p_y(z) \cdot \delta(\bar{v}_z.D \,|\, (+_{z=0}^{n} v_z.\bar{w}_z) \,|\, Q)$. Then $D' \,|\, Q \,|\, Q \xrightarrow{\tau} \Delta'$, with $\Delta' = \sum_{z=0}^{n} p_{y+1}(z) \cdot \delta(\bar{v}_z.D' \,|\, (+_{z=0}^{n} v_z.\bar{w}_z) \,|\, Q)$. We derive the result by exploiting the soundness of the composition of up-to-quasiconvexity, up-to-context and up-to-bm functions, chained with up-to-identity.

Let $\Delta'' = \sum_{z=0}^{n} p_y(z) \cdot \delta(\bar{v}_z.D' \mid (+_{z=0}^{n} v_z.\bar{w}_z) \mid Q)$*. We have:*

$$\begin{aligned}
&((\mathcal{F}_{qcv} \circ \mathcal{F}_{\mathcal{C}} \circ \mathcal{F}_{bm}) \curlywedge \mathcal{F}_{id})(K_{\otimes}(m))(\Delta, \Delta') \\
&\leq (\mathcal{F}_{qcv} \circ \mathcal{F}_{\mathcal{C}} \circ \mathcal{F}_{bm})(K_{\otimes}(m))(\Delta, \Delta'') + (K_{\otimes}(m))(\Delta'', \Delta') \\
&\leq (K_{\otimes}(bm))(\delta(D \mid Q), \delta(D' \mid Q)) + (K_{\otimes}(m))(\Delta'', \Delta') \\
&\leq \varepsilon + \varepsilon
\end{aligned}$$

6 Conclusion and future work

In this paper we studied techniques to increase the efficiency of the bisimulation proof method in the case of the (extended) Kantorovich metric. To this purpose, we have explored properties of the Kantorovich lifting, and we have generalized to the case of metrics the bisimulation up to $\mathcal{F}$ method by Sangiorgi. This allows us to reduce the size of the set of pairs for which we have to show the progress relation.

The theory of compatibility [11] for up-to techniques generalizes the respectfulness conditions on relations in a lattice-theoretic setting, where general properties of the progress relation and of the up-to functions (seen as functionals on the same lattice) can be proved and later instantiated to capture bisimulation relations on automata. A more recent approach [10] consists in directly focusing on the greatest compatible (or respectful) function. In this paper we considered probabilistic systems and metrics, where the domain and the target of the progress relation are not in the same lattice anymore, and the up-to functions are defined on the target domain. The generalization of the techniques presented in this paper to a lattice-theoretic setting provides an interesting line of research. In [2], up-to techniques are developed in an abstract fibrational setting, from which one could be able to obtain techniques for metrics. Studying whether the techniques of this paper can be obtained in this way is left as future work.

References

[1] Giorgio Bacci, Giovanni Bacci, Kim G. Larsen, and Radu Mardare. On-the-fly exact computation of bisimilarity distances. In *Proc. of TACAS*, volume 7795 of *LNCS*, pages 1–15. Springer, 2013.

[2] Filippo Bonchi, Daniela Petrişan, Damien Pous, and Jurriaan Rot. Coinduction up-to in a fibrational setting. In *Proc. of CSL-LICS*, pages 20:1–20:9. ACM, 2014.

[3] Konstantinos Chatzikokolakis, Daniel Gebler, Catuscia Palamidessi, and Lili Xu. Generalized bisimulation metrics. In *Proc. of CONCUR*, volume 8704 of *LNCS*, pages 32–46. Springer, 2014.

[4] Konstantinos Chatzikokolakis, Catuscia Palamidessi, and Valeria Vignudelli. Up-to techniques for generalized bisimulation metrics. Technical report, 2016. URL: https://hal.inria.fr/hal-01335234.

[5] Josee Desharnais, Vineet Gupta, Radha Jagadeesan, and Prakash Panangaden. Metrics for labeled markov systems. In *Proc. of CONCUR*, volume 1664 of *LNCS*, pages 258–273. Springer, 1999.

[6] Josee Desharnais, Radha Jagadeesan, Vineet Gupta, and Prakash Panangaden. The metric analogue of weak bisimulation for probabilistic processes. In *Proc. of LICS*, pages 413–422. IEEE, 2002.

[7] Josee Desharnais, Radha Jagadeesan, Vineet Gupta, and Prakash Panangaden. Metrics for labelled Markov processes. *Theor. Comp. Sci.*, 318(3):323–354, 2004.

[8] Cynthia Dwork. Differential privacy. In *Proc. of ICALP*, volume 4052 of *LNCS*, pages 1–12. Springer, 2006.

[9] R. Milner. *Communication and Concurrency.* Series in Comp. Sci. Prentice Hall, 1989.

[10] Damien Pous. Coinduction all the way up. To appear in *Proc. of LICS*, 2016.

[11] Damien Pous and Davide Sangiorgi. Enhancements of the bisimulation proof method. In *Advanced Topics in Bisimulation and Coinduction.* Cambridge University Press, 2012.

[12] Davide Sangiorgi. On the bisimulation proof method. *Mathematical Structures in Computer Science*, 8(5):447–479, 1998.

[13] Franck van Breugel and James Worrell. An algorithm for quantitative verification of probabilistic transition systems. In *Proc. of CONCUR*, volume 2154 of *LNCS*, pages 336–350. Springer, 2001.

[14] Franck van Breugel and James Worrell. Towards quantitative verification of probabilistic transition systems. In *Proc. of ICALP*, volume 2076 of *LNCS*, pages 421–432. Springer, 2001.

[15] Franck van Breugel and James Worrell. Approximating and computing behavioural distances in probabilistic transition systems. *Theor. Comp. Sci.*, 360(1–3):373–385, 2006.

[16] Franck van Breugel and James Worrell. The complexity of computing a bisimilarity pseudometric on probabilistic automata. In *Horizons of the Mind*, volume 8464 of *LNCS*, pages 191–213. Springer, 2014.

A Appendix

This appendix is not part of the original paper published by LIPIcs. It contains all proofs omitted from the main body of the paper due to space constraints, an expanded version of the proof of Theorem 4.2, and a short description of the Hausdorff lifting.

A.1 Proofs of Section 3

Proposition 1. *The following hold:*

1. *f is m_A, m_B-Lip iff $m_A \geq f^{\leftarrow}(m_B)$.*
2. *$f^{\leftarrow}$ is monotone.*
3. *$f^{\leftarrow}$ preserves metrics: $m_B \in \mathcal{M}_d(B)$ implies $f^{\leftarrow}(m_B) \in \mathcal{M}_d(A)$.*
4. *$f^{\leftarrow}$ preserves $\bigwedge, \bigvee$, that is: $f^{\leftarrow}(\bigwedge M) = \bigwedge f^{\leftarrow}(M)$ and $f^{\leftarrow}(\bigvee M) = \bigvee f^{\leftarrow}(M)$.*

Proof. The first three are immediate from the definition. The fourth is also straightforward:

$$\begin{aligned}
f^{\leftarrow}(\textstyle\bigwedge M)(a,a') &= (\textstyle\bigwedge M)(f(a), f(a')) \\
&= \inf_{m_A \in M} m_A(f(a), f(a')) \\
&= \inf_{m_B \in f^{\leftarrow}(M)} m_B(a,a') && \text{set } m_B = f^{\leftarrow}(m_A) \\
&= (\textstyle\bigwedge f^{\leftarrow}(M))(a,a')
\end{aligned}$$

and similarly for $\bigvee$. ☐

Proposition 2. *The following hold:*

1. *K is monotone.*
2. *$K(m) \in \mathcal{M}_d(S)$ (a proper metric) for all premetrics $m \in \mathcal{M}(S)$.*

Proof. Monotonicity comes from the definition of K and the fact that $m_1 \leq m_2$ implies that any m_1, d_V-Lip function is also m_2, d_V-Lip.

$K(m)(\Delta, \Delta) = 0$ comes directly from the fact that

$$d_V(\hat{f}(\Delta), \hat{f}(\Delta)) = 0$$

Similarly, symmetry comes from the fact that:

$$d_V(\hat{f}(\Delta), \hat{f}(\Theta)) = d_V(\hat{f}(\Theta), \hat{f}(\Delta))$$

Finally, the triangle-inequality of $K(m)$ comes from that of d_V:

$$\begin{aligned}
&K(m)(\Delta_1, \Delta_2) \\
&= \sup_f \{d_V(\hat{f}(\Delta_1), \hat{f}(\Delta_2)) \\
&\leq \sup_f \{d_V(\hat{f}(\Delta_1), \hat{f}(\Delta_3)) + d_V(\hat{f}(\Delta_3), \hat{f}(\Delta_2))\} && \text{triang. ineq. of } d_V \\
&\leq \sup_f \{d_V(\hat{f}(\Delta_1), \hat{f}(\Delta_3))\} + \sup_f \{d_V(\hat{f}(\Delta_3), \hat{f}(\Delta_2))\} \\
&= K(m)(\Delta_1, \Delta_3) + K(m)(\Delta_3, \Delta_2)
\end{aligned}$$

☐

Proposition 3. *Let $f : A \to B$ and $m_B \in \mathcal{M}(B)$. Then $(K \circ f^{\leftarrow})(m_B) \geq (f_*{}^{\leftarrow} \circ K)(m_B)$.*

Proof. By definition of K we have that:

$$(K \circ f^{\leftarrow})(m_B)(\Delta, \Theta) = \sup\{d_V(\hat{h}(\Delta), \hat{h}(\Theta)) \mid h : f^{\leftarrow}(m_B), d_V\text{-Lip}\}$$

On the other hand:

$$\begin{aligned}
&(f_*{}^{\leftarrow} \circ K)(m)(\Delta, \Theta) \\
&= K(m)(f_*(\Delta), f_*(\Theta)) \\
&= \sup\{d_V(\hat{g}(f_*(\Delta)), \hat{g}(f_*(\Theta))) \mid g : m, d_V\text{-Lip}\} \\
&= \sup\{d_V(g \mathbin{\hat{\circ}} f(\Delta), g \mathbin{\hat{\circ}} f(\Theta)) \mid g : m, d_V\text{-Lip}\}
\end{aligned}$$

Now for every m, d_V-Lip function $g : B \to V$, the function $h = g \circ f : A \to V$ is $f^{\leftarrow}(m_B), d_V$-Lip since

$$\begin{aligned}
d_V(h(a), h(a')) &= d_V(g(f(a)), g(f(a'))) && h = g \circ f \\
&\leq m(f(a), f(a')) && g : m, d_V\text{-Lip} \\
&= f^{\leftarrow}(m)(a, a') && \text{Def. of } f^{\leftarrow}
\end{aligned}$$

Hence the sup for $(K \circ f^{\leftarrow})(m_B)(\Delta, \Theta)$ ranges over a (possibly) larger set, from which the result follows. □

Proposition 4. *Let $M \subseteq \mathcal{M}$. Then $\bigwedge_d(M^{\triangledown}) = (\bigwedge M)^{\triangledown}$.*

Proof. For all $m \in M$ we have that $\bigwedge_d(M^{\triangledown}) \leq m^{\triangledown} \leq m$, in other words $\bigwedge_d(M^{\triangledown})$ is a lower bound of M. Hence we have that $\bigwedge_d(M^{\triangledown}) \leq \bigwedge M$. Since $(\bigwedge M)^{\triangledown}$ is the greatest metric below $\bigwedge M$ and $\bigwedge_d(M^{\triangledown})$ is a metric below $\bigwedge M$ we must have $\bigwedge_d(M^{\triangledown}) \leq (\bigwedge M)^{\triangledown}$.

Moreover for all $m \in M$ we have that $\bigwedge M \leq m$ and by the monotonicity of $^{\triangledown}$ we get $(\bigwedge M)^{\triangledown} \leq m^{\triangledown}$, so $(\bigwedge M)^{\triangledown}$ is a lower bound of $M^{\triangledown}$. Since $(\bigwedge M)^{\triangledown} \in \mathcal{M}_d$ is a lower bound of $M^{\triangledown} \subseteq \mathcal{M}_d$, it must be below its $\mathcal{M}_d$-infimum, that is $(\bigwedge M)^{\triangledown} \leq \bigwedge_d(M^{\triangledown})$. We conclude by anti-symmetry. □

Proposition 5. *The following hold:*

1. *$\curlywedge$ is associative and monotone on both arguments*
2. *$m_1 \wedge_d m_2 \leq m_1 \curlywedge m_2 \leq m_1 \wedge m_2$*
3. *$K(m_1 \curlywedge m_2) \leq K(m_1) \curlywedge K(m_2)$*

Proof. Associativity and monotonicity comes directly from the definition.

$m_1 \curlywedge m_2 \leq m_1 \wedge m_2$ is obtained by setting $t = s_1, s_2$ in the definition above. Setting $d = m_1 \wedge_d m_2$ (a metric) we have that $d(s_1, s_2) \leq d(s_1, t) + d(t, s_2) \leq m_1(s_1, t) + m_2(t, s_2)$ which shows $m_1 \wedge_d m_2 \leq m_1 \curlywedge m_2$.

Then $K(m_1 \curlywedge m_2)$ is a metric below both $K(m_1), K(m_2)$ (monotonicity of K), so it must hold that $K(m_1 \curlywedge m_2) \leq K(m_1) \wedge_d K(m_2) \leq K(m_1) \curlywedge K(m_2)$. □

Let $f : A \to B$ and assume it is m_A, m_B-Lip. Moreover, let $M_A = \{m_A^i\}_{i \in I}$ and $M_B = \{m_B^i\}_{i \in I}$ such that f is m_A^i, m_B^i-Lip for all $i \in I$. The following hold:

1. Inc/dec-reasing the source/target metric: f is m'_A, m'_B-Lip $\quad \forall m'_A \geq m_A, m'_B \leq m_B$

2. Infima and suprema: f is $\bigvee M_A, \bigvee M_B$-Lip and $\bigwedge M_A, \bigwedge M_B$-Lip

3. Metric closure: f is ${m_A}^{\triangledown}, {m_B}^{\triangledown}$-Lip

4. Kantorovich lifting: f_* is $K(m_A), K(m_B)$-Lip

Proof. 1) Directly from the definition of Lipschitz.

2) We know that $f^{\leftarrow}$ preserves infima and suprema and that f is m_A, m_B-Lip iff $m_A \geq f^{\leftarrow}(m_B)$ (Prop 1).

For all i we have that

$$m_A^i \geq f^{\leftarrow}(m_B^i) \geq \bigwedge f^{\leftarrow}(M_B) = f^{\leftarrow}(\bigwedge M_B)$$

Hence $f^{\leftarrow}(\bigwedge M_B)$ is a lower bound of M_A, so it must hold that $\bigwedge M_A \geq f^{\leftarrow}(\bigwedge M_B)$.

Similarly for $\bigvee$.

3) Assume that f is m_A, d_B-Lip, from Prop 1 this means that $m_A \geq f^{\leftarrow}(d_B)$. But $f^{\leftarrow}$ preserves metrics and ${m_A}^{\triangledown}$ is the largest metric below m_A, hence we must have ${m_A}^{\triangledown} \geq f^{\leftarrow}(d_B)$, which again from Prop 1 gives that f is ${m_A}^{\triangledown}, d_B$-Lip.

4) We know that f is m_A, m_B-Lip iff $m_A \geq f^{\leftarrow}(m_B)$ (Prop 1). Hence we have:

$$\begin{aligned} K(m_A) &\geq K(f^{\leftarrow}(m_B)) && \text{monotonicity of } K \\ &\geq {f_*}^{\leftarrow}(K(m_B)) && \text{Prop 3} \end{aligned}$$

which means that f_* is $K(m_A), K(m_B)$-Lip. □

A.2 Proofs of Section 4

The following proof is an expanded version of the proof presented in the main body of the paper, with the purpose of being easier to follow.

Any respectful function is sound.

Proof. Let $\mathcal{F}$ be respectful and let $\mathcal{G}$ be its corresponding source map from the definition of respectfulness. Assume $m \rightarrowtail \mathcal{F}(K(m))$, we need to show that $bm \leq m$. We construct a sequence of metrics $m_n, n \geq 0$ as:

$$\begin{aligned} m_0 &= m \\ m_{n+1} &= \mathcal{G}(m_n) \wedge m_n \end{aligned}$$

By construction we have that

$$m_n \geq m_{n+1} \qquad \forall n \geq 0$$

We now show that

$$m_n \rightarrowtail K(m_{n+1}) \qquad \forall n \geq 0$$

For the base case $n = 0$, from the respectfulness of $\mathcal{F}$ and the monotonicity of K we have that $\mathcal{F}(K(m)) \geq K(\mathcal{G}(m)) \geq K(\mathcal{G}(m) \wedge m)$. Hence $m \rightarrowtail \mathcal{F}(K(m))$ implies $m_0 = m \rightarrowtail K(\mathcal{G}(m) \wedge m) = K(m_1)$.

Assuming the property holds for $n \geq 0$, we want to show that $m_{n+1} \rightarrowtail K(m_{n+2})$, that is:

$$\mathcal{G}(m_n) \wedge m_n \rightarrowtail K(\mathcal{G}(m_{n+1}) \wedge m_{n+1})$$

We have that:

$$\begin{array}{rcr} & m_n \rightarrowtail K(m_{n+1}) & \text{(induction. hyp.)} \\ \Rightarrow & \mathcal{G}(m_n) \rightarrowtail K(\mathcal{G}(m_{n+1})) & \text{(respectf., } m_n \geq m_{n+1}) \\ \Rightarrow & \mathcal{G}(m_n) \wedge m_n \rightarrowtail K(\mathcal{G}(m_{n+1})) \wedge K(m_{n+1}) & \wedge \text{ preserves } \rightarrowtail \\ \Rightarrow & \mathcal{G}(m_n) \wedge m_n \rightarrowtail K(\mathcal{G}(m_{n+1}) \wedge m_{n+1}) & K(a \wedge b) \leq K(a) \wedge K(b) \end{array}$$

Finally, we have that:

$$\begin{array}{rcc} & \forall n \geq 0 : m_n \rightarrowtail K(m_{n+1}) & \\ \Rightarrow & \bigwedge_{n\geq 0} m_n \rightarrowtail \bigwedge_{n\geq 1} K(m_n) & \wedge \text{ preserves } \rightarrowtail \\ \Rightarrow & \bigwedge_{n\geq 0} m_n \rightarrowtail \bigwedge_{n\geq 0} K(m_n) & \text{smaller target} \\ \Rightarrow & \bigwedge_{n\geq 0} m_n \rightarrowtail K(\bigwedge_{n\geq 0} m_n) & K(\bigwedge A) \leq K(\bigwedge A) \end{array}$$

hence $\bigwedge_{n\geq 0} m_n$ is a bisimulation metric, and $m \geq \bigwedge_{n\geq 0} m_n$, which concludes the proof. □

The composition of respectful functions is respectful.

Proof. Let $\mathcal{F}_1, \mathcal{F}_2$ be respectful functions, and $\mathcal{G}_1, \mathcal{G}_2$ their corresponding source maps from the definition of respectfulness. Also let $\mathcal{F} = \mathcal{F}_1 \circ \mathcal{F}_2$ and $\mathcal{G} = \mathcal{G}_1 \circ \mathcal{G}_2$. We show that $\mathcal{F}, \mathcal{G}$ satisfy the requirements of respectfulness.

$\mathcal{F}$ is monotone as the composition of monotone functions. Moreover, from the respectfulness of $\mathcal{F}_1$ we have that $K(\mathcal{G}_1(\mathcal{G}_2(m))) \leq \mathcal{F}_1(K(\mathcal{G}_2(m))$. Then, from the respectfulness of $\mathcal{F}_2$ we have $K(\mathcal{G}_2(m)) \leq \mathcal{F}_2(K(m))$ which from the monotonicity of $\mathcal{F}_1$ implies $\mathcal{F}_1(K(\mathcal{G}_2(m)) \leq \mathcal{F}_1(\mathcal{F}_2(K(m)))$. Hence $(K \circ \mathcal{G}_1 \circ \mathcal{G}_2)(m) \leq (\mathcal{F}_1 \circ \mathcal{F}_2 \circ K)(m)$.

Finally, suppose $m \rightarrowtail K(m')$ and $m \geq m'$, since $\mathcal{F}_2$ is respectful we have that $\mathcal{G}_2(m) \rightarrowtail K(\mathcal{G}_2(m))$ and $\mathcal{G}_2(m) \geq \mathcal{G}_2(m')$. For the latter progression, since $\mathcal{F}_1$ is respectful we get that $\mathcal{G}_1(\mathcal{G}_2(m)) \rightarrowtail K(\mathcal{G}_1(\mathcal{G}_2(m')))$ and $\mathcal{G}_1(\mathcal{G}_2(m)) \geq \mathcal{G}_1(\mathcal{G}_2(m')))$ which concludes the proof. □

Define the chaining of two functions $\mathcal{F}_1, \mathcal{F}_2$ as $(\mathcal{F}_1 \curlywedge \mathcal{F}_2)(m) = \mathcal{F}_1(m) \curlywedge \mathcal{F}_2(m)$.

The chaining of respectful functions is respectful.

Proof. Let $\mathcal{F}_1, \mathcal{F}_2$ be respectful functions, and $\mathcal{G}_1, \mathcal{G}_2$ their corresponding source maps from the definition of respectfulness. Also let $\mathcal{F} = \mathcal{F}_1 \curlywedge \mathcal{F}_2$ and $\mathcal{G} = \mathcal{G}_1 \curlywedge \mathcal{G}_2$. We show that $\mathcal{F}, \mathcal{G}$ satisfy the requirements of respectfulness.

The monotonicity of $\mathcal{F}$ comes directly from the monotonicity of $\curlywedge$ (Prop 5). From the faithfulness of $\mathcal{F}_1, \mathcal{F}_2$ we get $K(\mathcal{G}_i(m)) \leq \mathcal{F}_i(K(m)), i \in \{1, 2\}$, which from the monotonicity of $\curlywedge$ implies that $K(\mathcal{G}_1(m)) \curlywedge K(\mathcal{G}_2(m)) \leq \mathcal{F}_1(K(m)) \curlywedge \mathcal{F}_2(K(m))$. From Prop 5 we have that

$$K(\mathcal{G}_1(m) \curlywedge \mathcal{G}_2(m)) \leq K(\mathcal{G}_1(m)) \curlywedge K(\mathcal{G}_2(m)) \leq \mathcal{F}_1(K(m)) \curlywedge \mathcal{F}_2(K(m))$$

which means that $(K \circ \mathcal{G})(m) \leq (\mathcal{F} \circ K)(m)$.

Now assume that $m \rightarrowtail K(m')$ and $m \geq m'$. From faithfulness we have that $\mathcal{G}_i(m) \rightarrowtail K(\mathcal{G}_i(m))$ and $\mathcal{G}_i(m) \geq \mathcal{G}_i(m')$, $i \in \{1,2\}$. Hence:

$$\begin{array}{rlr}
 & \mathcal{G}_1(m) \wedge \mathcal{G}_2(m) \rightarrowtail K(\mathcal{G}_1(m')) \wedge K(\mathcal{G}_2(m')) & \wedge \text{ preserves } \rightarrowtail \\
\Rightarrow & \mathcal{G}_1(m) \wedge \mathcal{G}_2(m) \rightarrowtail K(\mathcal{G}_1(m')) \curlywedge K(\mathcal{G}_2(m')) & (a \wedge b) \geq (a \curlywedge b) \\
\Rightarrow & \mathcal{G}_1(m) \wedge \mathcal{G}_2(m) \rightarrowtail K(\mathcal{G}_1(m') \curlywedge \mathcal{G}_2(m')) & K(a) \curlywedge K(b) \geq K(a \curlywedge b) \\
\Rightarrow & \mathcal{G}_1(m) \wedge_d \mathcal{G}_2(m) \rightarrowtail K(\mathcal{G}_1(m') \curlywedge \mathcal{G}_2(m')) & \text{metric closure, } K(a) \in \mathcal{M}_d \\
\Rightarrow & \mathcal{G}_1(m) \curlywedge \mathcal{G}_2(m) \rightarrowtail K(\mathcal{G}_1(m') \curlywedge \mathcal{G}_2(m')) & (a \wedge_d b) \leq (a \curlywedge b) \\
\Rightarrow & \mathcal{G}(m) \rightarrowtail K(\mathcal{G}(m')) &
\end{array}$$

Finally by the monotonicity of $\curlywedge$ we get that $\mathcal{G}_1(m) \curlywedge \mathcal{G}_2(m) \geq \mathcal{G}_1(m') \curlywedge \mathcal{G}_2(m')$, hence $\mathcal{G}(m) \geq \mathcal{G}(m')$, which means that $\mathcal{F}, \mathcal{G}$ satisfy all the requirements of respectfulness. □

The infimum of a set of respectful functions is respectful.

Proof. Given a countable set $\{\mathcal{F}_i\}$ of respectful functions with corresponding source maps $\{\mathcal{G}_i\}$, we prove that the function $\mathcal{F} = \bigwedge\{\mathcal{F}_i\}$ is respectful by using the source map $\mathcal{G} = \bigwedge\{\mathcal{G}_i\}$. $\mathcal{F}$ is monotone as the infimum of monotone functions. Moreover, from the respectfulness of $\mathcal{F}_i$ we have that $K(\mathcal{G}_i(m)) \leq \mathcal{F}_i(K(m))$. Hence, $\bigwedge\{K(\mathcal{G}_i(m))\} \leq \bigwedge\{\mathcal{F}_i(K(m))\}$. By the monotonicity of K we have $K(\bigwedge\{\mathcal{G}_i(m)\}) \leq \bigwedge\{K(\mathcal{G}_i(m))\}$, and we can conclude that $K(\bigwedge\{\mathcal{G}_i(m)\}) \leq \bigwedge\{\mathcal{F}_i(K(m))\}$.

Finally, suppose $m \rightarrowtail K(m')$ and $m \geq m'$. Since $\mathcal{F}_i$ is respectful we have that $\mathcal{G}_i(m) \rightarrowtail K(\mathcal{G}_i(m))$ and $\mathcal{G}_i(m) \geq \mathcal{G}_i(m')$. By the properties of progressions (Section 4.1) we have $\bigwedge\{\mathcal{G}_i(m)\} \rightarrowtail \bigwedge\{K(\mathcal{G}_i(m))\}$, and by $\bigwedge\{K(\mathcal{G}_i(m))\} \geq K(\bigwedge\{\mathcal{G}_i(m)\})$ (monotonicity of K) we derive $\bigwedge\{\mathcal{G}_i(m)\} \rightarrowtail K(\bigwedge\{\mathcal{G}_i(m)\})$. By definition, $\mathcal{G}_i(m) \geq \mathcal{G}_i(m')$ for all i implies $\bigwedge\{\mathcal{G}_i(m)\} \leq \bigwedge\{\mathcal{G}_i(m')\}$, which concludes the proof. □

If K is quasiconvex (resp. convex) then $\mathcal{F}_{qcv}$ (resp. $\mathcal{F}_{cv}$) is respectful.

Proof. We prove the result using the identity function on states $\mathcal{G}_{id}$. Both functions are monotonic, and if K is quasiconvex (respectively: convex) then by definition $K(\mathcal{G}_{id}(m))(\Delta, \Theta) \leq \sup_i K(m)(\Delta_i, \Theta_i)$ (respectively: $K(\mathcal{G}_{id}(m))(\Delta, \Theta) \leq \sum_i p_i K(m)(\Delta_i, \Theta_i)$) for all $\Delta = \sum_i p_i \Delta_i$ and $\Theta = \sum_i p_i \Theta_i$, from which we derive $K(\mathcal{G}_{id}(m)) \leq \mathcal{F}_{qcv}(m)$ (respectively: $K(\mathcal{G}_{id}(m)) \leq \mathcal{F}_{cv}(m)$). Finally, the identity satisfies the second condition of the definition of respectfulness. □

The functions $\mathcal{C}$ and $\mathcal{C}_*$ can be alternatively characterized as follows:

1. $\mathcal{C}(m)(s,t) = \inf\{m(s',t') \mid (s,t) \in (s',t')^{\mathcal{C}}\}$
2. $\mathcal{C}_*(l)(\Delta, \Theta) = \inf\{l(\Delta', \Theta') \mid (\Delta, \Theta) \in (\Delta', \Theta')^{\mathcal{C}_*}\}$

Proof. The left-to-right inequalities are a direct consequence of $\mathcal{C}$ and $\mathcal{C}_*$ being m, m-Lipschitz. For the opposite inequalities, we first prove that $\inf\{m(s',t') | (s,t) \in (s',t')^{\mathcal{C}}\} \leq \mathcal{C}(m)(s,t)$ by showing that:

- $\inf\{m(s',t') | (s,t) \in (s',t')^{\mathcal{C}}\} \leq m(s,t)$
- for all $C \in \mathcal{C}$, $\inf\{m(s',t') | (C[s], C[t]) \in (s',t')^{\mathcal{C}}\} \leq \inf\{m(s',t') | (s,t) \in (s',t')^{\mathcal{C}}\}$

The inequality on distributions, i.e., $\inf\{l(\Delta', \Delta') | (\Delta, \Theta) \in (\Delta', \Theta')^{\mathcal{C}}\} \leq \mathcal{C}_*(l)(\Delta, \Theta)$, follows analogously.

$\square$

Lemmas used in Theorem 4.3:

- K preserves closedness under $\mathcal{C}$. That is, if m is closed under $\mathcal{C}$, then $K(m)$ is closed under $\mathcal{C}_*$.

  Proof. Direct consequence of the fact that K on both metrics preserves the Lip property (Theorem 3.4). $\square$

- Lemma $\star$: for any faithful context C, $C[s] \xrightarrow{a} \Delta$ implies that, for all t, if $m(s,t) < \top$ then $C[t] \to \Theta$ with $K(\mathcal{G}(m'))(\Delta, \Theta) \leq m(s,t)$

  Proof. By the definition of faithfulness, we have two cases:

  1. $\Delta = \sum_i p_i C_i[\delta(s)]$ where $C_i \in \mathcal{C}$, and for all $t : C[t] \xrightarrow{a} \Theta = \sum_i p_i C_i[\delta(t)]$.
     We have that:

     $$\begin{aligned} & K(\mathcal{G}(m'))(\Delta, \Theta) \\ & = K(\mathcal{G}(m'))(\textstyle\sum_i p_i C_i[\delta(s)], \sum_i p_i C_i[\delta(t)]) \\ & \leq \max_i K(\mathcal{G}(m'))(C_i[\delta(s)], C_i[\delta(t)]) && \text{quasiconv. of } K(\mathcal{G}(m')) \\ & \leq K(\mathcal{G}(m'))(\delta(s), \delta(t)) && \text{closedness under } \mathcal{C} \\ & \leq K(m)(\delta(s), \delta(t)) && \mathcal{G}(m') \leq m' \leq m \\ & \leq m(s,t) \end{aligned}$$

  2. $\Delta = \sum_i p_i C_i[\Delta']$ where $C_i \in \mathcal{C}$, $s \xrightarrow{a'} \Delta'$ and for all t: if $t \xrightarrow{a'} \Theta'$ then $C[t] \xrightarrow{a} \Theta = \sum_i p_i C_i[\Theta']$.

     We derive from $m \rightarrowtail K(m')$, $m(s,t) < \top$ and $s \xrightarrow{a'} \Delta'$ that $t \xrightarrow{a'} \Theta'$ with $K(m')(\Delta', \Theta') \leq m(s,t)$. Hence, $C[t] \xrightarrow{a} \Theta = \sum_i p_i C_i[\Theta']$ and we have:

     $$\begin{aligned} & K(\mathcal{G}(m'))(\Delta, \Theta) \\ & = K(\mathcal{G}(m'))(\textstyle\sum_i p_i C_i[\Delta'], \sum_i p_i C_i[\Theta']) \\ & \leq \max_i K(\mathcal{G}(m'))(C_i[\Delta'], C_i[\Theta']) && \text{quasiconv. of } K(\mathcal{G}(m')) \\ & \leq K(\mathcal{G}(m'))(\Delta', \Theta') && \text{closedness under } \mathcal{C} \\ & \leq K(m')(\Delta', \Theta') && \mathcal{G}(m') \leq m' \\ & \leq m(s,t) \end{aligned}$$

  $\square$

A.3 Proofs of Section 5

The (unary) contexts of pCCS are faithful.

Proof. The proof is by induction on the structure of contexts C. The case $C = \mathbf{0}$ is trivial.

- Case $C = \alpha. \oplus_i [p_i]C_i$.
  The only transition allowed is $C[s] \xrightarrow{\alpha} \sum_i p_i \cdot \delta(C_i[s])$, and indeed for all t we have that $C[t] \xrightarrow{\alpha} \sum_i p_i \cdot \delta(C_i[t])$.

- Case $C = C_1 + C_2$. We can assume without loss of generality that $C[s] = C_1[s] + s_2$. If s_2 moves, the result immediately follows from the fact that for any t, the component s_2 of $C_1[t] + s_2$ can do the same transition and identical distributions are reached (and identical distributions satisfy both cases of the definition of faithfulness, by just considering contexts without holes). If $C_1[s]$ moves, i.e., $C_1[s] + s_2 \xrightarrow{\alpha} \Delta$ with $C_1[s] \xrightarrow{\alpha} \Delta$, then the result follows from the inductive hypothesis on C_1.

- Case $C = C_1 \,|\, C_2$. We can assume without loss of generality that $C[s] = C_1[s] \,|\, s_2$. and $C[t] = C_1[t] \,|\, s_2$ for some s_2. We have three cases:
  - If $C[s] \xrightarrow{\alpha} \Delta$ with $s_2 \xrightarrow{\alpha} \Delta' = \sum_i p_i \cdot s'_i$ and $\Delta = \sum_i p_i \cdot \delta(C[s] \,|\, s_{i'})$ then $C[t] \xrightarrow{\alpha} \Theta = \sum_i p_i \cdot \delta(C[t] \,|\, s_{i'})$ and the result follows by considering as contexts $C'_i = C[\cdot] \,|\, s'_i$, since $\sum_i p_i \cdot K(m)(\delta(s), \delta(t)) \leq K(m)(\delta(s), \delta(t)) \leq m(s,t)$.
  - If $C[s] \xrightarrow{\alpha} \Delta$ with $C_1[s] \xrightarrow{\alpha} \Delta_1$ and $\Delta = \Delta_1 \,|\, s_2$ then by the inductive hypothesis on C_1 we have two cases:
    1. $\Delta_1 = \sum_i p_i C_i[\delta(s)]$ and $\forall t$: $C_1[t] \xrightarrow{\alpha} \sum_i p_i C_i[\delta(t)]$.
       Hence, $\Delta = \sum_i p_i C_i[\delta(s)] \,|\, s_2$ and for all t, $C[t] \xrightarrow{\alpha} \Theta = \sum_i p_i C_i[\delta(t)] \,|\, s_2$
    2. $\Delta_1 = \sum_i p_i C_i[\Delta']$ with $s \xrightarrow{\alpha'} \Delta'$ and $\forall t$: if $t \xrightarrow{\alpha'} \Theta'$ then $C_1[t] \xrightarrow{\alpha} \sum_i p_i C_i[\Theta']$.
       The result follows as in the previous case.
  - $C[s] \xrightarrow{\tau} \Delta$ resulting form the synchronization $C_1[s] \xrightarrow{\alpha} \Delta_1$ and $s_2 \xrightarrow{\bar{\alpha}} \Delta_2 = \sum_j q_j \cdot u_j$. By the inductive hypothesis we have two cases:
    1. $\Delta_1 = \sum_i p_i C_i[\delta(s)]$ and $\forall t$: $C_1[t] \xrightarrow{\alpha} \sum_i p_i C_i[\delta(t)]$.
       Hence, $\Delta = \sum_{i,j} p_i q_j C_i[\delta(s)] \,|\, u_j$ and for all t, $C[t] \xrightarrow{\alpha} \Theta = \sum_{i,j} p_i q_j C_i[\delta(t)] \,|\, u_j$
    2. $\Delta_1 = \sum_i p_i C_i[\Delta']$ with $s \xrightarrow{\alpha'} \Delta'$ and $\forall t$: if $t \xrightarrow{a'} \Theta'$ then $C_1[t] \xrightarrow{\alpha} \sum_i p_i C_i[\Theta']$.
       The result follows as in the previous case.

- Case $C = (\nu a) C_1$.
  $(\nu a) C_1[s] \xrightarrow{\alpha} \Delta$ iff $C_1[s] \xrightarrow{\alpha} \Delta_1$ and $\alpha \neq a, \bar{a}$. Hence, the result directly follows from the inductive hypothesis.

- Case $C = A$.
  Since C has no empty holes, the definition is trivially satisfied.

$\square$

A.4 Hausdorff lifting

The Hausdorff lifting $H : \mathcal{M}(A) \to \mathcal{M}(2^A)$, which can be used to define the progression relation (Definition 1), lifts a metric on A to a metric on sets over A. It is defined as:

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

Note that the inf of the empty set is $\top$ ($+\infty$ or whatever the top element of our range of distances is). Hence for all $X \in 2^A$ we have that $H(d)(\emptyset, X) = \top$, and we only need to set $H(d)(\emptyset, \emptyset) = 0$ as a special case to make it a metric.