Divergences on Monads for Relational Program Logics

Tetsuya Sato   Shin-ya Katsumata

June 14, 2022

Abstract

Several relational program logics have been introduced for integrating reasoning about relational properties of programs and measurement of quantitative difference between computational effects. Towards a general framework for such logics, in this paper, we formalize quantitative difference between computational effects as *divergence on monad*, then develop a relational program logic acRL that supports generic computational effects and divergences on them. To give a categorical semantics of acRL supporting divergences, we give a method to obtain *graded strong relational liftings* from divergences on monads. We derive two instantiations of acRL for the verification of 1) various differential privacy of higher-order functional probabilistic programs and 2) difference of distribution of costs between higher-order functional programs with probabilistic choice and cost counting operations.

1 Introduction

Comparing behavior of programs is one of the fundamental activities in the verification and analysis of programs, and many concepts, techniques and formal systems have been proposed for this purpose, such as *product program construction* ([11]), *relational Hoare logic* ([14]), *higher-order relational refinement types* ([9]) and so on.

Several recent relational program logics integrate compositional reasoning about relational properties of programs and over-approximation of *quantitative difference* between computational effects of programs; the latter is done in the style of *effect system* ([36]). One successful logic of this kind is Barthe et al.'s *approximate probabilistic relational Hoare logic* (apRHL for short) designed for verifying *differential privacy* of probabilistic programs ([12]). A judgment of apRHL is of the form $c \sim_{\epsilon,\delta} c' : \Phi \Rightarrow \Psi$, and its intuitive meaning is that for any state pair (ρ, ρ') related by Φ, the ϵ-distance between two probability distributions of final states $[\![c]\!]\rho$ and $[\![c']\!]\rho'$ is below δ, and final states satisfy Ψ. Another relational program logic that measures the difference between computational effects of programs is Çiçek et al.'s $\mathsf{RelCost}$ ([18]). The target of the reasoning is a higher-order programming language equipped with cost counting effect. When we derive a judgment $\Delta; \Psi; \Gamma \vdash M_1 \ominus M_2 \precsim n : \Phi$ in $\mathsf{RelCost}$, the sound semantics ensures that the difference of cost counts by M_1 and M_2 is bound by n.

A high-level view on these relational program logics is that they integrate the feature of measuring quantitative difference between computational effects into relational program logic. We are interested in extracting mathematical essence of this design and making relational program logics versatile. Towards this goal, we contribute the following development.

- We introduce a structure called *divergence on monad* for measuring quantitative difference between computational effects (Section 4, 5). This generalizes various statistical divergences, such as Kullback-Leibler divergence and total variation distance on probability distributions. After exploring examples of divergence on monads, we introduce a method to transfer divergences on a monad to those on another monad through monad opfunctors.

- The key structure to integrate divergences on monads and relational program logics is something called *graded strong relational lifting* of monads that extends given divergences. We present a general construction of such liftings from divergences on monads in Section 7. This generalization shows that the development of relational program logics with quantitative measurement on computational effects can be done with various combinations of monads and divergences on them.
- We introduce a generic relational program logic (called acRL) over Moggi's computational metalanguage (the simply-typed lambda calculus with monadic types) in Section 8. Inside acRL, we can use graded strong relational liftings constructed from divergences on a monad, and reason about relational properties of programs together with quantitative difference of computational effects. To illustrate how the reasoning works in acRL, we instantiate it with the computational metalanguage having effectful operations for continuous random sampling (Section 9) and cost counting operation (Section 10).

2 Preliminaries

We assume basic knowledge about category theory ([37]) and Moggi's model of computational effects ([42]). The definition of monad [37, Chapter VI] and Kleisli category [37, Section VI.5] are omitted.

In this paper, a Cartesian category (CC for short) is specified by a category $\mathbb{C}$ with a designated final object 1 and a binary product functor $(\times)\colon \mathbb{C}^2 \to \mathbb{C}$. The associated pairing operation and projection morphisms are denoted by $\langle -, - \rangle, \pi_1, \pi_2$, respectively. The unique morphism to the terminal object is denoted by $!_I : I \to 1$. A Cartesian closed category (CCC for short) is a CC $(\mathbb{C}, 1, (\times))$ with a specified exponential functor $(\Rightarrow)\colon \mathbb{C}^{\mathrm{op}} \times \mathbb{C} \to \mathbb{C}$. The associated evaluation morphism and currying operation is denoted by $\mathrm{ev}, \lambda(-)$ respectively.

Let $(\mathbb{C}, 1, (\times))$ be a CC. A *global element* of $I \in \mathbb{C}$ is a morphism of type $1 \to I$. For a category $\mathbb{C}$, we define the functor $U^{\mathbb{C}} : \mathbb{C} \to \mathbf{Set}$ by $U^{\mathbb{C}} = \mathbb{C}(1, -)$. When $\mathbb{C}$ is obvious, $U^{\mathbb{C}}$ is denoted by $\mathbb{C}$. Morphisms in $\mathbb{C}$ act on global elements by the composition. To emphasize this action, we introduce a dedicated notation $(\bullet)$ whose type is $\mathbb{C}(I, J) \times UI \to UJ$. Of course, $f \bullet x \triangleq f \circ x = (Uf)(x)$. We also define the partial application of a binary morphism $f\colon I \times J \to K$ to a global element $i \in UI$ by $f_i \triangleq f \circ \langle i \circ !_J, \mathrm{id}_J \rangle : J \to K$. When $\mathbb{C}$ is a CCC, there is an evident isomorphism $\lfloor - \rfloor\colon U(I \Rightarrow J) \cong \mathbb{C}(I, J)$. We write $\lceil - \rceil$ for its inverse.

A monad (T, η, μ) on a category $\mathbb{C}$ determines the operation $(-)^\sharp : \mathbb{C}(I, TJ) \to \mathbb{C}(TI, TJ)$ called *Kleisli extension*. It is defined by $f^\sharp \triangleq \mu_J \circ Tf$. A monad may be given as a *Kleisli triple* [42, Definition 1.2]. A *strong monad* on a CC $(\mathbb{C}, 1, (\times))$ is a pair of a monad (T, η, μ) and a natural transformation $\theta_{I,J}\colon I \times TJ \to T(I \times J)$ called *strength*. It should satisfy four axioms; see [42, Definition 3.2] for detail.

In a CC-SM $(\mathbb{C}, 1, (\times), T, \eta, \mu, \theta)$, the application of the strength to a global element can be expressed by the unit and the Kleisli extension of T [42, Proof of Proposition 3.4]:

$$\theta_{I,J} \bullet \langle i, c \rangle = ((\eta_{I \times J})_i)^\sharp \bullet c \quad (i \in UI, c \in U(TJ)). \tag{1}$$

We will use this fact in Proposition 7 and Proposition 1.

There are plenty of examples of C(C)Cs. For the models of probabilistic computation, we will later use CC **Meas** of measurable spaces and CCC **QBS** of quasi-Borel spaces ([28]). Their definitions are deferred to Section 13.

2.1 Category of Binary Relations

We next introduce the category $\mathbf{BRel}(\mathbb{C})$ of binary relations over $\mathbb{C}$-objects. This category is equivalent to *subscones* of $\mathbb{C}^2$ ([41]). It offers an underlying category for relational reasoning about programs interpreted in $\mathbb{C}$.

- An object in $\mathbf{BRel}(\mathbb{C})$ is a triple (I_1, I_2, R) where $R \subseteq UI \times UJ$.
- A morphism from (I_1, I_2, R) to (J_1, J_2, S) in $\mathbf{BRel}(\mathbb{C})$ is a pair of $\mathbb{C}$-morphisms $f_1 \colon I_1 \to J_1$ and $f_2 \colon I_2 \to J_2$ such that for any $(i_1, i_2) \in R$, we have $(f_1 \bullet i_1, f_2 \bullet i_2) \in S$.

When X is a name of a $\mathbf{BRel}(\mathbb{C})$-object, by X_1, X_2 we mean its first and second component, and by R_X we mean its third component; so $X = (X_1, X_2, R_X)$. By $(x_1, x_2) \in X$ we mean $(x_1, x_2) \in R_X$. For objects $X, Y \in \mathbf{BRel}(\mathbb{C})$ and a morphism $(f_1, f_2) \colon (X_1, X_2) \to (Y_1, Y_2)$ in $\mathbb{C}^2$, by

$$(f_1, f_2) \colon X \dot{\to} Y$$

we mean that $(f_1, f_2) \in \mathbf{BRel}(\mathbb{C})(X, Y)$, that is, for any $(x_1, x_2) \in X$, we have $(f_1 \bullet x_1, f_2 \bullet x_2) \in Y$. We say that $X \in \mathbf{BRel}(\mathbb{C})$ is an *endorelation* (over I) if $X_1 = X_2 (= I)$.

We next define the forgetful functor $p_{\mathbb{C}} : \mathbf{BRel}(\mathbb{C}) \to \mathbb{C}^2$ by

$$p_{\mathbb{C}} X \triangleq (X_1, X_2), \quad p_{\mathbb{C}}(f_1, f_2) \triangleq (f_1, f_2).$$

For $(I_1, I_2) \in \mathbb{C}^2$, by $\mathbf{BRel}(\mathbb{C})_{(I_1, I_2)}$ we mean the complete boolean algebra $\{X \in \mathbf{BRel}(\mathbb{C}) \mid X_1 = I_1 \wedge X_2 = I_2\}$ with the order given by $X \le Y \iff R_X \subseteq R_Y$.

When $\mathbb{C}$ is a C(C)C, so is $\mathbf{BRel}(\mathbb{C})$ [41, Proposition 4.3]. We specify a final object, a binary product functor and an exponential functor (in case $\mathbb{C}$ is a CCC) on $\mathbf{BRel}(\mathbb{C})$ by:

$$\begin{aligned}
\dot{1} &\triangleq (1, 1, \{(\mathrm{id}_1, \mathrm{id}_1)\}) \\
X \mathbin{\dot{\times}} Y &\triangleq (X_1 \times Y_1, X_2 \times Y_2, \{(\langle x_1, y_1 \rangle, \langle x_2, y_2 \rangle) \mid (x_1, x_2) \in X, (y_1, y_2) \in Y\}) \\
X \mathbin{\dot{\Rightarrow}} Y &\triangleq (X_1 \Rightarrow Y_1, X_2 \Rightarrow Y_2, \{(f_1, f_2) \mid \forall (x_1, x_2) \in X \ . \ (\mathrm{ev} \circ \langle f_1, x_1 \rangle, \mathrm{ev} \circ \langle f_2, x_2 \rangle) \in Y\}).
\end{aligned}$$

3 Divergences on Objects

We introduce the concept of *divergence* on objects in a CC $\mathbb{C}$. Major differences between divergence and metric are threefold: 1) it is defined over objects in $\mathbb{C}$, 2) no axioms is imposed on it, and 3) it takes values in a partially ordered monoid called *divergence domain*, which we define below.

Definition 1. A *divergence domain* $\mathcal{Q} = (Q, \le, 0, (+))$ is a partially ordered commutative monoid whose poset part is a complete lattice.

The monoid addition $(+)$ is only required to be monotone; no interaction with the sup / inf is required. We reserve the letter $\mathcal{Q}$ to denote a general divergence domain. Examples of divergence domains are:

$$\begin{aligned}
\mathcal{N} &= (\mathbb{N} \cup \{\infty\}, \le, 0, (+)), & \mathcal{R}^+ &= ([0, \infty], \le, 0, (+)), \\
\mathcal{R}^\times &= ([0, \infty], \le, 1, (\times)), & \mathcal{R}^+_1 &= ([0, \infty], \le, 0, \lambda(p, q) \ . \ p + q + pq), \\
\mathcal{Z} &= (\mathbb{Z} \cup \{\infty, -\infty\}, \le, 0, (\bar{+})), & \mathcal{R} &= ([-\infty, \infty], \le, 0, (\bar{+}))
\end{aligned}$$

Here, $\bar{+}$ is an extension of the addition by $r \mathbin{\bar{+}} (-\infty) = (-\infty) \mathbin{\bar{+}} r = -\infty$.

Definition 2. Let $\mathbb{C}$ be a CC. A *$\mathcal{Q}$-divergence* on an object $I \in \mathbb{C}$ is a function $d\colon (UI)^2 \to \mathcal{Q}$.

A suitable notion of morphism between $\mathbb{C}$-objects with divergences is *nonexpansive morphism*.

Definition 3. Let $\mathbb{C}$ be a CC. We define the category $\mathbf{Div}_{\mathcal{Q}}(\mathbb{C})$ of $\mathcal{Q}$-divergences on $\mathbb{C}$-objects and nonexpansive morphisms between them by the following data.

- An object is a pair (I, d) of an object $I \in \mathbb{C}$ and a $\mathcal{Q}$-divergence d on I.
- A morphism from (I, d) to (J, e) is a $\mathbb{C}$-morphism $f\colon I \to J$ such that for any $x_1, x_2 \in UI$, $e(f \bullet x_1, f \bullet x_2) \le d(x_1, x_2)$ holds.

For an object $X \in \mathbf{Div}_{\mathcal{Q}}(\mathbb{C})$, by d_X we mean its $\mathcal{Q}$-divergence part. We also define the forgetful functor $V_{\mathcal{Q},\mathbb{C}} : \mathbf{Div}_{\mathcal{Q}}(\mathbb{C}) \to \mathbb{C}$ by $V_{\mathcal{Q},\mathbb{C}}(I, d) \triangleq I$ and $V_{\mathcal{Q},\mathbb{C}}(f) \triangleq f$.

We remark that the forgetful functor $V_{\mathcal{Q},\mathbf{Set}} : \mathbf{Div}_{\mathcal{Q}}(\mathbf{Set}) \to \mathbf{Set}$ is a (Grothendieck) *fibration*, and the functor $\overline{U}\colon \mathbf{Div}_{\mathcal{Q}}(\mathbb{C}) \to \mathbf{Div}_{\mathcal{Q}}(\mathbf{Set})$ defined by $\overline{U}(I, d) \triangleq (UI, d)$ and $\overline{U}(f) \triangleq f$ makes the following commutative square a pullback in **CAT** (the large category of categories and functors between them):

$$\begin{tikzcd}
\mathbf{Div}_{\mathcal{Q}}(\mathbb{C}) \arrow[r, "\overline{U}"] \arrow[d, "V_{\mathcal{Q},\mathbb{C}}"'] \arrow[dr, phantom, "\lrcorner", very near start] & \mathbf{Div}_{\mathcal{Q}}(\mathbf{Set}) \arrow[d, "V_{\mathcal{Q},\mathbf{Set}}"] \\
\mathbb{C} \arrow[r, "U"'] & \mathbf{Set}
\end{tikzcd}$$

Therefore this pullback diagram asserts that $V_{\mathcal{Q},\mathbb{C}} : \mathbf{Div}_{\mathcal{Q}}(\mathbb{C}) \to \mathbb{C}$ arises from the *change-of-base* of the fibration $V_{\mathcal{Q},\mathbf{Set}}$ along the global section functor $U : \mathbb{C} \to \mathbf{Set}$ ([29]).

4 Divergences on Monads

We introduce the concept of *divergence on monad* as a quantitative measure of difference between computational effects. This is hinted from Barthe and Olmedo's composable divergences on probability distributions ([13]). Divergences on monads are defined upon two extra data called *grading monoid* and *basic endorelation*.

Definition 4. A *grading monoid* is a partially ordered monoid $(M, \le, 1, (\cdot))$.

Definition 5. A *basic endorelation* is a functor $E\colon \mathbb{C} \to \mathbf{BRel}(\mathbb{C})$ such that EI is an endorelation on I.

Grading monoids will be used when formulating (ε, δ)-differential privacy as a divergence on a monad. Basic endorelations specify which global elements are regarded as identical. Any CC $\mathbb{C}$ has at least two basic endorelations of *equality relations* and *total relations*:

$$\mathrm{Eq}I \triangleq (I, I, \{(i, i) \mid i \in UI\}) \qquad\qquad \mathrm{Top}I \triangleq (I, I, UI \times UI).$$

Other examples of basic endorelations can be found in concrete categories.

- The category $\mathbf{Div}_{\mathcal{Q}}(\mathbb{C})$ of $\mathcal{Q}$-divergences on $\mathbb{C}$-objects has a basic relation E_δ parameterized by $\delta \in \mathcal{Q}$. It collects all pairs of global elements whose divergence is bound by d. That is, $E_\delta(I, d) \triangleq (I, I, \{(x_1, x_2) \mid d(x_1, x_2) \le \delta\})$.
- The category of preorders and monotone functions has the basic endorelation E_{eq} collecting equivalent global elements: $E_{eq}(I, \le) \triangleq (I, I, \{(x, y) \mid x \le y \wedge y \le x\})$.

Definition 6. Let $(\mathbb{C}, 1, (\times), T, \eta, \mu, \theta)$ be a CC-SM, $\mathcal{Q}$ be a divergence domain, $(M, \leq, 1, (\cdot))$ be a grading monoid and $E\colon \mathbb{C} \to \mathbf{BRel}(\mathbb{C})$ be a basic endorelation. An *E-relative M-graded $\mathcal{Q}$-divergence* (when $M = 1$, we drop "M-graded") on the monad T is a doubly-indexed family of $\mathcal{Q}$-divergences $\Delta = \{\Delta_I^m\colon (U(TI))^2 \to \mathcal{Q}\}_{m\in M, I\in\mathbb{C}}$ satisfying the following conditions:

Monotonicity For any $m \leq m'$ in M, $I \in \mathbb{C}$ and $c_1, c_2 \in U(TI)$,

$$\Delta_I^m(c_1, c_2) \geq \Delta_I^{m'}(c_1, c_2).$$

E-unit Reflexivity For any $I \in \mathbb{C}$,

$$\sup_{(x_1,x_2)\in EI} \Delta_I^1(\eta_I \bullet x_1, \eta_I \bullet x_2) \leq 0.$$

E-composability For any $m_1, m_2 \in M$, $I, J \in \mathbb{C}$, $c_1, c_2 \in U(TI)$ and $f_1, f_2\colon I \to TJ$,

$$\Delta_J^{m_1\cdot m_2}(f_1^\sharp \bullet c_1, f_2^\sharp \bullet c_2) \leq \Delta_I^{m_1}(c_1, c_2) + \sup_{(x_1,x_2)\in EI} \Delta_J^{m_2}(f_1 \bullet x_1, f_2 \bullet x_2).$$

We write $\mathbf{Div}(T, E, M, \mathcal{Q})$ for the collection of E-relative M-graded $\mathcal{Q}$-divergences on T. We introduce a partial order $\preceq$ on $\mathbf{Div}(T, E, M, \mathcal{Q})$ by:

$$\Delta_1 \preceq \Delta_2 \iff \forall m \in M, I \in \mathbb{C}, c_1, c_2 \in U(TI) \;.\; (\Delta_1)_I^m(c_1, c_2) \geq (\Delta_2)_I^m(c_1, c_2).$$

The E-composability condition is a generalization of the composability of differential privacy stated as [13, Theorem 1]. What is new in this paper is that 1) we introduce a condition on the monad unit (E-unit reflexivity), and that 2) the sup computed in E-unit reflexivity and E-composability scans global elements related by E, while [13] only considers the case where $E = \mathrm{Eq}$. We will later show that both E-unit reflexivity and E-composability play an important role when connecting divergences, relational liftings of T, and the monad structure of T - these conditions are necessary and sufficient to construct *strong graded relational liftings* of T satisfying *fundamental property* with respect to divergences (Proposition 2).

5 Examples of Divergences on Monads

5.1 Cost Difference for Deterministic Computations

To aid in understanding the E-unit reflexivity and E-composability conditions, we illustrate a few divergences on an elementary monad: the *cost count monad* $T = \mathbb{N} \times -$ on **Set**. Its unit and Kleisli extension are defined by

$$\eta_I(x) \triangleq (0, x) \qquad f^\sharp(i, x) \triangleq (i + \pi_1(f(x)), \pi_2(f(x))) \qquad (x \in I, i \in \mathbb{N}, f\colon I \to TJ).$$

The monad T can be used to record the cost incurred by deterministic computations. For instance, consider the quick sort algorithm qsort and the insertion sort algorithm isort, both of which are modified so that they tick a count whenever they compare two elements to be sorted. These two modified sort programs are interpreted as functions $[\![\mathsf{qsort}]\!], [\![\mathsf{isort}]\!]\colon \mathbb{N}^* \to T(\mathbb{N}^*)$, so that the first component of $[\![\mathsf{qsort}]\!](x)$ and that of $[\![\mathsf{isort}]\!](x)$ report the number of comparisons performed during sorting x.

We first define an $\mathcal{N}$-divergence C_I on TI, for each $I \in \mathbf{Set}$, by

$$\mathsf{C}_I((i, x), (j, y)) \triangleq |i - j|.$$

This divergence C_I computes the difference of costs between two computations $(i,x),(j,y) \in TI$, ignoring their return values. The family $\mathsf{C} = \{\mathsf{C}_I\}_{I \in \mathbf{Set}}$ *forms* a Top-relative $\mathcal{N}$-divergence on T. The Top-unit reflexivity of C means that the difference of costs between pure computations is zero:

$$\mathsf{C}_I(\eta_I(x), \eta_I(y)) = \mathsf{C}_I((0,x),(0,y)) = 0.$$

The Top-composability of C says that we can limit the cost difference of two runs of programs $f^\sharp(i,x)$ and $g^\sharp(j,y)$ by the sum of cost difference of the preceding computations $(i,x),(j,y)$ and that of two programs $f,g \colon I \to TJ$. The latter is measured by taking the sup of cost difference of $f(x)$ and $g(y)$, where (x,y) range over the basic endorelation $\mathrm{Top}I$.

$$\begin{aligned}
\mathsf{C}_I(f^\sharp(i,x), g^\sharp(j,y)) &= \mathsf{C}_I(i + \pi_1(f(x)), \pi_2(f(x)), j + \pi_1(g(y)), \pi_2(g(y))) \\
&\leq |i-j| + \sup_{x,y \in I} |\pi_1(f(x)) - \pi_1(g(y))| \\
&= \mathsf{C}_I((i,x),(j,y)) + \sup_{(x,y) \in \mathrm{Top}I} \mathsf{C}_J(f(x), g(y)).
\end{aligned}$$

We remark that C is *not* an Eq-relative $\mathcal{N}$-divergence on T because the Eq-composability fails: when $f(x) = (0,w)$, $f(y) = (1,w)$ and $f(z) = (0,v)$ (for $z \neq x,y$) we have $\mathsf{C}_I((0,x),(0,y)) = 0$ and $\sup_{(x,y) \in \mathrm{Eq}I} \mathsf{C}_J(f(x), f(y)) = 0$, but we have $\mathsf{C}_J(f^\sharp(0,x), f^\sharp(0,y)) = \mathsf{C}_J((0,w),(1,w)) = 1$.

Alternatively, we may consider the following $\mathcal{N}$-divergence C'_I on TI for each $I \in \mathbf{Set}$:

$$\mathsf{C}'_I((i,x),(j,y)) \triangleq \begin{cases} |i-j| & x = y \\ \infty & x \neq y \end{cases}.$$

This divergence is *sensitive* on return values of computations. When return values of two computations agree, C' measures the cost difference as done in C, but when they do not agree, the cost difference is judged as ∞. This divergence is an Eq-relative $\mathcal{N}$-divergence on T.

5.2 Cost Difference for Nondeterministic Computations

Deterministic and nondeterministic computations with cost counting can be respectively modeled by the monads $(\mathbb{N} \times -)$ and $P(\mathbb{N} \times -)$ on $\mathbf{Set}$.

We define the divergences for cost difference as in Table 1. These divergences extract the upper bound of cost difference between two computations. The divergences C and NC measure the usual distance of costs for deterministic and nondeterministic computations respectively. The divergence NCI measures the *subtraction* of costs of two nondeterministic computations. For results of two nondeterministic computations $A, B \in P(\mathbb{N} \times I)$, the divergence $\mathsf{NCI}_I(A,B)$ is an upper bound of $i - j$ for all possible choices of $(i,x) \in A$ and $(j,y) \in B$, where a lower bound of $i - j$ is also given by $-\mathsf{NCI}_I(B,A)$. The same idea to measure the difference of costs between two programs by subtraction also appears in ([18, 47]). If either A or B is empty, we fail to get an information of costs. We then have $\mathsf{NCI}_I(A,B) = -\infty$. On the other hand, if both A and B are not empty, their *cost intervals* are defined by

$$[l_A, h_A] \triangleq [\inf_{(i,x) \in A} i, \sup_{(i,x) \in A} i], \qquad [l_B, h_B] \triangleq [\inf_{(j,y) \in B} j, \sup_{(j,y) \in B} j].$$

We then have $\mathsf{NCI}_I(A,B) = h_A - l_B$ and $-\mathsf{NCI}_I(B,A) = l_A - h_B$.

Table 1: (1-graded) Top-relative $\mathcal{Q}$-divergences for cost counting monads

$\Delta \in \mathbf{Div}(T, \mathrm{Top}, 1, \mathcal{Q})$	T	$\mathcal{Q}$	Definition of $\Delta_I(c_1, c_2)$
C	$\mathbb{N} \times -$	$\mathcal{N}$	$\mathsf{C}_I((i,x),(j,y)) = \lvert i-j \rvert$
NC	$P(\mathbb{N} \times -)$	$\mathcal{N}$	$\mathsf{NC}_I(A,B) = \sup_{(i,x)\in A,(j,y)\in B} \lvert i-j \rvert$
NCI	$P(\mathbb{N} \times -)$	$\mathcal{Z}$	$\mathsf{NCI}_I(A,B) = \sup_{(i,x)\in A,(j,y)\in B} i-j$

5.3 Divergences for Differential Privacy

Differential privacy (DP for short) is a quantitative definition of privacy of randomized queries in databases. DP is based on the idea of noise-adding anonymization against background-knowledge attacks. In the study of DP, a query is modeled by a measurable function $c \colon I \to GJ$, where I and J are measurable spaces of inputs and outputs respectively, and GJ is the measurable space of all probability measures over J; here G itself refers to the *Giry monad* ([26]; see also Section 13).

Definition 7 (Differential Privacy, ([21]))**.** Let $c \colon I \to GJ$ be a morphism in **Meas**, representing a randomized query. The query c satisfies (ε, δ)-differential privacy ($\epsilon, \delta \geq 0$ are reals) if for any adjacent datasets $(d_1, d_2) \in R_{\mathrm{adj}}$[1], the following holds:

$$\forall S \subseteq_{\mathrm{measurable}} J.\ \Pr[c(d_1) \in S] \leq \exp(\varepsilon) \Pr[c(d_2) \in S] + \delta.$$

To express this definition in terms of divergence on monad, we introduce a doubly-indexed family of $\mathcal{R}^+$-divergence $\mathsf{DP} = \{\mathsf{DP}^{\varepsilon}_J\}_{\varepsilon \in [0,\infty], J \in \mathbf{Meas}}$ on GJ by

$$\mathsf{DP}^{\varepsilon}_J(\mu_1, \mu_2) \triangleq \sup_{S \in \Sigma_J} (\mu_1(S) - \exp(\varepsilon)\mu_2(S)) \quad (\mu_1, \mu_2 \in GJ).$$

Then the query $c \colon I \to GJ$ satisfies (ε, δ)-DP if and only if

$$\forall (d_1, d_2) \in R_{\mathrm{adj}}\ .\ \mathsf{DP}^{\varepsilon}_J(c(d_1), c(d_2)) \leq \delta.$$

The pair (ε, δ) indicates the difference between output probability distributions $c(d_1)$ and $c(d_2)$ of the query c for given datasets d_1 and d_2. Intuitively, the parameter ε is an upper bound of the ratio $\Pr[c(d_1) = s] / \Pr[c(d_2) = s]$ of probabilities which indicates the leakage of privacy. If ε is large, attackers can distinguish the datasets d_1 and d_2 from the outputs of the query c. The parameter δ is the probability of failure of privacy protection.

The family DP forms an Eq-relative $\mathcal{R}^+$-graded $\mathcal{R}^+$-divergence on the Giry monad G [52, Lemma 6]. This is proved by extending the composability of the divergence for DP on discrete probability distributions shown as [12, Lemmas 3 and 6] and [13, Proposition 5], based on the *composition theorem* of DP [22, Section 3.5].

The conditions in Definition 6 on DP corresponds to the following basic properties of DP:

(monotonicity) The monotonicity of DP corresponds to weakening the differential privacy of queries: if c satisfies (ε, δ)-DP and $\varepsilon \leq \varepsilon'$ and $\delta \leq \delta'$ holds, then c satisfies (ε', δ')-DP.

(Eq-unit reflexivity) The Eq-unit reflexivity of DP implies $\mathsf{DP}^0_J(\eta_J \circ h(x), \eta_J \circ h(x)) = 0$ for any measurable function $h \colon I \to J$ and $x \in I$. This, together with the composability

[1]Strictly speaking, differential privacy depends on the definition of adjacency of datasets. The adjacency relation R_{adj} is usually defined as $\{(d_1, d_2) | \rho(d_1, d_2) \leq 1\}$ with a metric ρ over I.

below, ensures the *robustness* of DP of a query $c\colon I \to GJ$ with respect to deterministic postprocessing:

$$\forall h : J \to K \;.\; c \text{ is } (\epsilon,\delta)\text{-DP} \implies Gh \circ c \text{ is } (\epsilon,\delta)\text{-DP}. \tag{2}$$

In fact, the divergence DP is reflexive: we have $\mathsf{DP}^0_J(\mu,\mu) = 0$ for every $\mu \in GJ$. Therefore $h\colon J \to K$ and Gh in (2) can be replaced by $h\colon J \to GK$ and $h^\sharp$; the replaced condition states the *robustness* of DP of a query with respect to probabilistic postprocessing.

(Eq-composability) The Eq-composability of DP corresponds to the known property of DP called the *sequential composition theorem* ([22]). If $c_1\colon I \to GJ'$ and $c_2\colon J' \to GJ$ are (ε_1,δ_1)-DP and (ε_2,δ_2)-DP respectively, then the sequential composition $c_2^\sharp \circ c_1\colon I \to GJ$ of the queries c_1 and c_2 is $(\varepsilon_1+\varepsilon_2,\delta_1+\delta_2)$-DP.

A Non-Example: Pointwise Differential Privacy. We stated above that a parameter (ε,δ) of DP intuitively gives an upper bound of the probability ratio $\Pr[c(d_1)=s]/\Pr[c(d_2)=s]$ and the probability of failure of privacy protection. However, strictly speaking, there is a gap between the definition of (ε,δ)-DP and this intuition of ε and δ. *Pointwise differential privacy* ([46, Definition 3.2] and [27, Proposition 1.2.3]) is a finer definition of DP that is faithful to the intuition.

Definition 8. A measurable function $c\colon I \to GJ$ (regarded as a query) is *pointwise* (ε,δ)-*differentially private* if whenever d_1 and d_2 are adjacent, for some $A \in \Sigma_J$ with $\Pr[c(d_1) \notin A] \le \delta$, we have

$$\forall s \in A.\; \Pr[c(d_1)=s] \le \exp(\varepsilon)\Pr[c(d_2)=s],$$

which is equivalent to [2]

$$\forall S \subseteq_{\text{measurable}} A.\; \Pr[c(d_1)\in S] \le \exp(\varepsilon)\Pr[c(d_2)\in S].$$

To express this definition in terms of divergence on monad, we introduce a doubly-indexed family of $\mathcal{R}^+$-divergences $\mathsf{pwDP} = \{\mathsf{pwDP}^\varepsilon_J\}_{\varepsilon\in\mathcal{R}^+, J\in\mathbf{Meas}}$ called *pointwise indistinguishability*:

$$\mathsf{pwDP}^\varepsilon_J(\mu_1,\mu_2) \triangleq \inf\{\mu_1(J\setminus A) \mid A \in \Sigma_X \wedge (\forall S\in\Sigma_J. S\subseteq A \implies \mu_1(S) \le \exp(\varepsilon)\mu_2(S))\}.$$

Then $c\colon I \to GJ$ is pointwise (ε,δ)-differentially private if and only if

$$\forall (d_1,d_2) \in R_{\text{adj}} \;.\; \mathsf{pwDP}^\varepsilon_J(c(d_1),c(d_2)) \le \delta.$$

The family pwDP is obviously reflexive: $\mathsf{pwDP}^\varepsilon_J(\mu,\mu)=0$ holds for any $\mu\in GJ$ and $\varepsilon \ge 0$. Hence it is Eq-unit reflexive too. However, it is not Eq-composable. We let $3=\{0,1,2\}$ and $2=\{0,1\}$ be discrete spaces, and let $\alpha=\exp(\varepsilon)$. We define two probability distributions $\mu_1,\mu_2\in G3$ by

$$\mu_1 \triangleq \frac{1}{10}\mathbf{d}_0 + \frac{9}{10}\mathbf{d}_1, \qquad \mu_2 \triangleq \frac{9}{10\alpha}\mathbf{d}_1 + (1-\frac{9}{10\alpha})\mathbf{d}_2.$$

We then have $\mathsf{pwDP}^\varepsilon_3(\mu_1,\mu_2)=\frac{1}{10}$ with $A=\{1,2\}$ since $\frac{1}{10} > \exp(\varepsilon)\cdot 0$, $\frac{9}{10} \le \exp(\varepsilon)\cdot\frac{9}{10\alpha}$, and $0 \le \exp(\varepsilon)\cdot(1-\frac{9}{10\alpha})$. Next, we define $f\colon 3 \to G2$ by

$$f(0) \triangleq \frac{1}{10}\mathbf{d}_0 + \frac{9}{10}\mathbf{d}_1, \quad f(1) \triangleq \frac{9}{10}\mathbf{d}_0 + \frac{1}{10}\mathbf{d}_1, \quad f(2) \triangleq \mathbf{d}_1.$$

[2]Remark that $\Pr[c(d_1)=s]$ and $\Pr[c(d_2)=s]$ are Radon-Nikodym derivatives of $c(d_1)$ and $c(d_2)$ with respect to a measure ν such that $c(d_1), c(d_2) \ll \nu$. [$\Longrightarrow$] Obvious. [$\Longleftarrow$] By Radon-Nikodym theorem we can take the Radon-Nikodym derivatives $\Pr[c(d_1)=s]$ and $\Pr[c(d_2)=s]$ with respect to $\nu = c(d_1)+c(d_2)$. The inequality does not depend on the choice of ν.

Table 2: Eq-relative M-graded $\mathcal{Q}$- ($\mathcal{Q}_s$-)divergences on G (G_s)

Δ	M	$\mathcal{Q}$	$\mathcal{Q}_s$	Definition of $\Delta_I^m(\mu_1,\mu_2)$	Composability proof
DP	$\mathcal{R}^+$	$\mathcal{R}^+$	$\mathcal{R}^+$	$\sup_{S\in\Sigma_I}(\mu_1(S)-\exp(\varepsilon)\mu_2(S))$	([13])
${}^\alpha\mathsf{Re}$	1	$\mathcal{R}^+$	$\mathcal{R}$	$\frac{1}{\alpha-1}\log\int_I\left(\frac{\mu_1(x)}{\mu_2(x)}\right)^\alpha\mu_2(x)\ dx.$	([40])
zCDP	$\mathcal{R}^+$	$\mathcal{R}^+$	$\mathcal{R}$	$\sup_{1<\alpha}\frac{1}{\alpha}({}^\alpha\mathsf{Re}_I(\mu_1,\mu_2)-m)$	([17])
${}^w\mathsf{tCDP}$	1	$\mathcal{R}^+$	$\mathcal{R}$	$\sup_{1<\alpha<w}\frac{1}{\alpha}({}^\alpha\mathsf{Re}_I(\mu_1,\mu_2))$	([16])

Table 3: Statistical divergences that are Eq-relative $\mathcal{Q}$- (resp. $\mathcal{Q}_s$-) divergences on G (resp. G_s)

Name	Δ	$\mathcal{Q}$	$\mathcal{Q}_s$	Definition of $\Delta_I^m(\mu_1,\mu_2)$
Total variation distance	TV	$\mathcal{R}^+$	$\mathcal{R}^+$	$\frac{1}{2}\int_I|\mu_1(x)-\mu_2(x)|\ dx$
Kullback-Leibler divergence	KL	$\mathcal{R}^+$	?	$\int_I\mu_1(x)\log\left(\frac{\mu_1(x)}{\mu_2(x)}\right)\ dx$
Hellinger distance	HD	$\mathcal{R}^+$	?	$\frac{1}{2}\int_I\left(\sqrt{\mu_1(x)}-\sqrt{\mu_2(x)}\right)^2\ dx$
χ^2-divergence	Chi	$\mathcal{R}_1^+$	?	$\int_I\frac{(\mu_1(x)-\mu_2(x))^2}{\mu_2(x)}\ dx$

We then calculate

$$f^\sharp(\mu_1)=\frac{82}{100}\mathbf{d}_0+\frac{18}{100}\mathbf{d}_1,\quad f^\sharp(\mu_2)=\frac{81}{100\alpha}\mathbf{d}_0+(\frac{100\alpha-90+9}{100\alpha})\mathbf{d}_1.$$

Then, we obtain $\mathsf{pwDP}_2^\varepsilon(f^\sharp(\mu_1),f^\sharp(\mu_2))=\frac{82}{100}$ with $A=\{0\}$ since $\frac{82}{100}>\exp(\varepsilon)\frac{81}{100\alpha}$. Hence $\mathsf{pwDP}_2^\varepsilon(f^\sharp(\mu_1),f^\sharp(\mu_2))=\frac{82}{100}>\frac{1}{10}=\mathsf{pwDP}_3^\varepsilon(\mu_1,\mu_2)$. Thus pwDP is not Eq-composable, because by the reflexivity of pwDP, we have $\sup_{(x,y)\in\mathrm{Eq}_3}\mathsf{pwDP}^0(f(x),f(y))=0$.

Various Relaxations of Differential Privacy Since the seminal work on DP by [21], various relaxations of differential privacy have been proposed: *Rényi DP* ([40]), *zero-concentrated DP* ([17]) and *truncated zero-concentrated DP* ([16]). They give tighter bounds of differential privacy. These relaxations of differential privacy can be expressed by suitable divergences on the Giry monad G and sub-Giry monad G_s; see Table 2 for their definitions. There, $\alpha,w\in(1,\infty)$ are non-grading parameters for Re and tCDP. Each row of the table represents that Δ is an Eq-relative $\mathcal{Q}$- (resp. $\mathcal{Q}_s$-) divergences on G (resp. G_s), and the definition of $\Delta_I(\mu_1,\mu_2)$ follows.

5.4 Statistical Divergences and Composablity of f-Divergences

Apart from differential privacy, various distances between (sub-)probability distributions are introduced in probability theory. They are called *statistical divergences.* Examples include: *total variation distance* TV, *Hellinger distance* HD, *Kullback-Leibler divergence* KL, and χ^2*-divergence* Chi; they are defined in Table 3. These statistical divergences are Eq-relative divergences on the Giry monad G (and G_s for TV); see the same table for their divergence domains. Question marks in the column of $\mathcal{Q}_s$ means that we do not know with which monoid structure the Eq-composability holds. We remark that these divergences are also reflexive, that is, $\Delta(c,c)=0$. Eq-composability of these divergences in discrete form are proved in ([13, 45]). Later, [52] extends their results to the composability of divergences in continuous form.

Each of four divergences in Table 3 can be expressed as an *f-divergence* ${}^f\mathsf{Div}$ ([19, 20, 43]):

$${}^f\mathsf{Div}_I(\mu_1,\mu_2)\triangleq\int_I\mu_2(x)f\left(\frac{\mu_1(x)}{\mu_2(x)}\right)dx.$$

Table 4: Parameters for Proposition 1

${}^{f}\mathsf{Div}$	Weight function f	γ	α	β	β'
TV	$f(t) = \lvert t-1\rvert/2$	0	0	1	0
KL	$f(t) = f\log(t) - t + 1$	0	-1	1	1
HD	$f(t) = (\sqrt{t}-1)^2/2$	0	$-1/4$	$1/2$	$1/2$
Chi	$f(t) = (t-1)^2/2$	1	-2	2	2

Here, f is a parameter called *weight function*, and has to be a convex function $f\colon [0,\infty) \to \mathbb{R}$, continuous at 0 and satisfying $\lim_{x\to+0} xf(x) = 0$. Weight functions for four divergences $\mathsf{TV}, \mathsf{KL}, \mathsf{HD}, \mathsf{Chi}$ are in Table 4. In fact, $\mathsf{DP}^{\varepsilon}$ is also an f-divergence with weight function $f(t) = \max(0, t - \exp(\varepsilon))$; see [13, Proposition 2]. We also remark that Rényi divergence ${}^{\alpha}\mathsf{Re}$ of order α is the logarithm of the f-divergence with weight function $f(t) = t^{\alpha}$.

f-divergences have several nice properties such as reflexivity, postprocessing inequality, joint-convexity, duality and continuity ([20, 35]). However, the Eq-composability of f-divergences is not guaranteed in general. Here we provide a sufficient condition for the Eq-composability of ${}^{f}\mathsf{Div}$ over a specific form of divergence domain.

Proposition 1. *Let $\gamma \geq 0$ be a nonnegative real number, $\mathcal{R}^{+}_{\gamma} = ([0,\infty], \leq, 0, \lambda(p,q)\ .\ p+q+\gamma pq)$ be the divergence domain, and f be a weight function such that $f \geq 0$ and $f(1) = 0$. If there exists $\alpha, \beta, \beta' \in \mathbb{R}$ such that, for all $x, y, z, w \in [0,1]$, the following hold (suppose $0f(0/0) = 0$):*

$$
\begin{aligned}
0 &\leq (\beta' z + (1-\beta')x) + \gamma x f\,(z/x) \\
xyf\,(zw/xy) &\leq (\beta w + (1-\beta)y)xf\,(z/x) + (\beta' z + (1-\beta')x)yf\,(w/y) \\
&\quad + \gamma xyf\,(z/x)\,f\,(w/y) + \alpha(x-z)(w-y),
\end{aligned}
$$

then ${}^{f}\mathsf{Div}$ is an Eq-relative $\mathcal{R}^{+}_{\gamma}$-divergence on the Giry monad G. When $\alpha = 0$ and $\beta, \beta' \in [0,1]$, G can be replaced with the sub-Giry monad G_s.

The proof of this proposition generalizes and integrates the proofs given in [45, Section 5.A.2]. This proposition is applicable to prove the composability of divergences in Table 3 by choosing suitable parameters; see Table 4.

5.5 Divergences on the Probability Monad on QBS via Monad Opfunctors.

We have seen various divergences on the Giry monad G. It would be nice if they are transferred to the probability monad P on **QBS** (Section 13). For this, we first develop a generic method for transferring divergences on monads.

Let $(\mathbb{C}, S)$ and $(\mathbb{D}, T)$ be two CC-SMs. A *monad opfunctor* [53, Section 4] is a functor $p\colon \mathbb{C} \to \mathbb{D}$ together with a natural transformation $\lambda\colon p\circ S \to T\circ p$ making the following diagrams commute:

$$
\begin{tikzcd}
p \arrow[d, "p\circ\eta^S"'] \arrow[dr, "\eta^T\circ p"] & \\
p\circ S \arrow[r, "\lambda"'] & T\circ p
\end{tikzcd}
\qquad
\begin{tikzcd}
p\circ S\circ S \arrow[r, "\lambda\circ S"] \arrow[d, "p\circ\mu^S"'] & T\circ p\circ S \arrow[r, "T\circ\lambda"] & T\circ T\circ p \arrow[d, "\mu^T\circ p"] \\
p\circ S \arrow[rr, "\lambda"'] & & T\circ p
\end{tikzcd}
$$

Proposition 2. *Let* $(\mathbb{C}, S), (\mathbb{D}, T)$ *be two CC-SMs,* $(p\colon \mathbb{C} \to \mathbb{D}, \lambda\colon p \circ S \to T \circ p)$ *be a monad opfunctor, and assume that* $U^{\mathbb{D}} \circ p = U^{\mathbb{C}}$ *holds, and basic endorelations* $F\colon \mathbb{C} \to \mathbf{BRel}(\mathbb{C})$ *and* $E\colon \mathbb{D} \to \mathbf{BRel}(\mathbb{D})$ *satisfy* $R_{FpI} = R_{EI}$ *for all* $I \in \mathbb{C}$ *(we here use* $U^{\mathbb{D}} \circ p = U^{\mathbb{C}}$*). Then for any* $\Delta \in \mathbf{Div}(T, E, M, \mathcal{Q})$*, the following doubly-indexed family of* $\mathcal{Q}$*-divergences* $\langle p, \lambda\rangle^* \Delta = \{(\langle p, \lambda\rangle^* \Delta)^m_I\}_{m \in M, I \in \mathbb{C}}$ *on* SI *is an* F*-relative* M*-graded* $\mathcal{Q}$*-divergence on* S*:*

$$(\langle p, \lambda\rangle^* \Delta)^m_I(\nu_1, \nu_2) \triangleq \Delta^m_{pI}(\lambda_I \bullet \nu_1, \lambda_I \bullet \nu_2) = \Delta^m_{pI}((U^{\mathbb{D}}\lambda_I)(\nu_1), (U^{\mathbb{D}}\lambda_I)(\nu_2)).$$

The left adjoint $L\colon \mathbf{QBS} \to \mathbf{Meas}$ of the adjunction $L \dashv K\colon \mathbf{Meas} \to \mathbf{QBS}$ and the natural transformation $l\colon LP \Rightarrow GL$ defined by $l_X([\alpha, \mu]_{\sim_X}) = \mu(\alpha^{-1}(-))$ forms a monad opfunctor from the probability monad P on $\mathbf{QBS}$ to the Giry monad G on $\mathbf{Meas}$ [28, Prop. 22 (3)]. Through this monad opfunctor (L, l), we can convert Eq-divergences on G to those on P. This conversion can be applied to all the statistical divergences in Table 2 and 3.

In addition, for any standard Borel space, we can view such converted divergences $\langle L, l\rangle^* \Delta$ as the same thing as the original Δ. When $\Omega \in \mathbf{Meas}$ is standard Borel, we have an equality $LK\Omega = \Omega$, and $l_{K\Omega}$ is an isomorphism. Therefore we obtain an isomorphism $l_{K\Omega}\colon LPK\Omega \cong GLK\Omega = G\Omega$ [28, Prop. 22 (4)]. A concrete description of its inverse is $l^{-1}_{K\Omega} \bullet \mu = [\gamma', \mu(\gamma^{-1}(-))]_{\sim K\Omega}$, where $\gamma'\colon \mathbb{R} \to \Omega$ and $\gamma\colon \Omega \to \mathbb{R}$ are a section-retraction pair (i.e. $\gamma' \circ \gamma = \mathrm{id}_\Omega$) that exists for any standard Borel Ω.

Theorem 1. *For any* $\Delta \in \mathbf{Div}(G, \mathrm{Eq}, M, \mathcal{Q})$ *and standard Borel* $\Omega \in \mathbf{Meas}$*,*

$$(\langle L, l\rangle^* \Delta)^m_{K\Omega}(l^{-1}_{K\Omega} \bullet \mu_1, l^{-1}_{K\Omega} \bullet \mu_2) = \Delta^m_\Omega(\mu_1, \mu_2) \quad (\mu_1, \mu_2 \in U(G\Omega)).$$

5.6 Divergences on State Monads

The state monad $T_S \triangleq S \Rightarrow (- \times S)$ with a state space S is used to represent programs that update the state. We construct divergences on T_S using divergences d_S on the state space S in several ways.

5.6.1 Lipschitz Constant on States

We first consider the state monad T_S on $\mathbf{Set}$. We also consider a function $d_S\colon S^2 \to [0, \infty]$ satisfying $d_S(s, s) = 0$. The following $\mathcal{R}^\times$-divergence $\Delta^{\mathsf{lip}, d_S}_I(f_1, f_2)$ on $T_S I$ measures how much the function pair $(\pi_2 \circ f_1, \pi_2 \circ f_2)$ extends the distance between two states before updated. In short, $\Delta^{\mathsf{lip}, d_S}$ measures the Lipschitz constant on state transformers.

Proposition 3. *The family* $\Delta^{\mathsf{lip}, d_S} = \{\Delta^{\mathsf{lip}, d_S}_I\}_{I \in \mathbf{Set}}$ *of* $\mathcal{R}^\times$*-divergences on* $T_S I$ *defined by*

$$\Delta^{\mathsf{lip}, d_S}_I(f_1, f_2) \triangleq \sup_{s_1, s_2 \in S} \frac{d_S(\pi_2(f_1(s_1)), \pi_2(f_2(s_2)))}{d_S(s_1, s_2)} \quad (f_1, f_2 \in T_S I,\ \textit{we suppose } 0/0 = 1)$$

is a Top*-relative* $\mathcal{R}^\times$*-divergence on* T_S*.*

For state transformers $f_1, f_2 \in T_S I$, their state-updating part is given as functions $\pi_2 \circ f_1, \pi_2 \circ f_2 \in S \Rightarrow S$. When $f_1 = f_2 = g$, $\Delta^{\mathsf{lip}, d_S}_I(g, g)$ is exactly the Lipschitz constant of $\pi_2 \circ g$.

5.6.2 Distance between State Transformers with the Same Inputs

Suppose that the function d_S also satisfies the triangle inequality. The following $\mathcal{R}^+$-divergence $\Delta^{\mathsf{met}, d_S}_I(f_1, f_2)$ on $T_S I$ estimates the distance between updated states after the state transformers f_1 and f_2 are applied to the same input.

Proposition 4. *Suppose that the function d_S also satisfy the triangle-inequality. The family $\Delta^{\mathsf{met},d_S} = \{\Delta_I^{\mathsf{met},d_S}\}_{I\in\mathbf{Set}}$ of $\mathcal{R}^+$-divergences on $T_S I$ defined by:*

$$\Delta_I^{\mathsf{met},d_S}(f_1,f_2) \triangleq \begin{cases} \sup_{s\in S} d_S(\pi_2(f_1(s)),\pi_2(f_2(s))) & \pi_1\circ f_1 = \pi_1\circ f_2 \text{ and} \\ & \pi_2\circ f_1, \pi_2\circ f_2\colon \text{nonexpansive} \\ \infty & \text{otherwise} \end{cases}$$

is an Eq*-relative $\mathcal{R}^+$-divergence on T_S.*

5.6.3 Sup-Metric on the State Monad on the Category of Generalized Ultrametric Spaces

The category **Gum** of generalized ([0, 1]-valued) ultrametric spaces[3] and nonexpansive functions is Cartesian closed [49, Section 2.2]. We consider the state monad $T_S = S \Rightarrow (-\times S)$ on **Gum** for a fixed space $(S, d_S) \in \mathbf{Gum}$. From the definition of exponential objects in **Gum**, $T_S(I, d_I)$ consists of the set of nonexpansive state transformers with the sup metric between them. In fact, the metric part of all $T_S(I, d_I)$ forms a divergence on T_S.

Proposition 5. *The family $\{d_{T_S I}\colon (T_S(I,d_I))^2 \to [0,1]\}_{(I,d_I)\in\mathbf{Gum}}$ consisting of the metric part of the spaces $T_S(I, d_I)$, given by*

$$d_{T_S I}(f_1,f_2) \triangleq \sup_{s\in S} \max\left(d_I(\pi_1(f_1(s)),\pi_1(f_2(s))), d_S(\pi_2(f_1(s)),\pi_2(f_2(s)))\right)$$

forms an Eq*-relative $([0,1],\le,\max,0)$-divergence on T_S.*

In the category **Gum**, instead of Eq, there is another basic endorelation Dist_0:

$$\mathsf{Dist}_0(I,d_I) \triangleq \{(x_1,x_2) \mid d_I(x_1,x_2)=0\}.$$

By modifying the divergence $d_{T_S(-)}$, we obtain a Dist_0-relative $([0,1],\le,\max,0)$-divergence as below:

Proposition 6. *The following forms a Dist_0-relative $([0,1],\le,\max,0)$-divergence on T_S.*

$$\Delta^{\mathsf{Dist}_0}_{(I,d_I)}(f_1,f_2) \triangleq \sup_{d_S(s_1,s_2)=0} \max(d_S(\pi_1(f_1(s_1)),\pi_1(f_2(s_2))), d_I(\pi_2(f_1(s_1)),\pi_2(f_2(s_2)))).$$

5.7 Combining Divergence with Cost

In Section 5.2, we have introduced a divergence on the monad $P(\mathbb{N}\times -)$ modeling nondeterministic choice and cost counting. In this section we construct a divergence on the combination of a general computational effect and cost counting.

Let $(\mathbb{C}, T)$ be a CC-SM and $\Delta \in \mathbf{Div}(T, \mathrm{Eq}, 1, \mathcal{Q})$ be a divergence and $(N, 1_N\colon 1\to N, (\star)\colon N\times N\to N)$ be a monoid object in $\mathbb{C}$ (for cost counting). Then the composite $T(N\times -)$ of the monad T and the monoid action monad $N\times(-)$ again carries a monad structure. We now define a family $\mathsf{C}(\Delta,N) = \{\mathsf{C}(\Delta,N)_I\colon (U(T(N\times I)))^2 \to \mathcal{Q}\}_{I\in\mathbb{C}}$ of $\mathcal{Q}$-divergences by

$$\mathsf{C}(\Delta,N)_I(c_1,c_2) \triangleq \begin{cases} \Delta_N(T\pi_1\bullet c_1, T\pi_1\bullet c_2) & \Delta_{N\times I}(c_1,c_2)\le\Delta_N(T\pi_1\bullet c_1,T\pi_1\bullet c_2) \\ \top_{\mathcal{Q}} & \text{otherwise} \end{cases}.$$

[3]Recall that an ultrametric space (I, d_I) is a set I together with a function $d_I : I^2 \to [0,1]$ such that $d_I(x,x) = 0$ and $d_I(x,z) \le \max(d_I(x,y), d_I(y,z))$.

Proposition 7. *The family* $\mathsf{C}(\Delta, N)$ *is an* Eq*-relative* $\mathcal{Q}$*-divergence on* $T(N \times -)$.

For example, the divergence $\mathsf{C}(\mathsf{KL}, \mathbb{R})$ on the composite monad $G(\mathbb{R} \times -)$ on **Meas** describes Kullback-Leibler divergence between distributions of costs in the probabilistic computations with real-valued costs. Intuitively, the side condition $\mathsf{KL}_{\mathbb{R}\times I}(\mu_1, \mu_2) \leq \mathsf{KL}_{\mathbb{R}}(G\pi_1 \bullet \mu_1, G\pi_1 \bullet \mu_2)$ in the definition of $\mathsf{C}(\mathsf{KL}, \mathbb{R})$ means that the difference between μ_1 and μ_2 lies only in the costs.

5.8 Preorders on Monads

To explore the generality of our framework, we look at the case where the divergence domain is $\mathcal{B} = (\{0 \geq 1\}, 1, \times)$; here $\times$ is the numerical multiplication. We identify an indexed family $\Delta = \{\Delta_I \colon (U(TI))^2 \to \mathcal{B}\}_{I \in \mathbb{C}}$ of $\mathcal{B}$-divergences and a family of adjacency relations $\tilde{\Delta}(1)I \triangleq \{(c_1, c_2) \mid \Delta_I(c_1, c_2) \leq 1\}$.

We point out a connection between Eq-relative $\mathcal{B}$-divergences and *preorders on monads* studied in ([32, 50]). A preorder on a monad T on **Set** assigns a preorder $\sqsubseteq_I$ on TI for each $I \in$ **Set**, and this assignment satisfies:

Substitutivity For any function $f \colon I \to TJ$ and $c_1, c_2 \in TI$, $c_1 \sqsubseteq_I c_2$ implies $f^\sharp(c_1) \sqsubseteq_J f^\sharp(c_2)$.

Congruence For any function $f_1, f_2 \colon I \to TJ$, if $f_1(x) \sqsubseteq_J f_2(x)$ holds for any $x \in I$, then $f_1^\sharp(c) \sqsubseteq_J f_2^\sharp(c)$ holds for any $c \in TI$.

Proposition 8. *A preorder on a monad* T *on* **Set** *bijectively corresponds to an* Eq*-relative* $\mathcal{B}$*-divergence* Δ *on* T *such that each* $\tilde{\Delta}(1)I$ *is a preorder.*

For a preorder $\sqsubseteq$ on a monad T on **Set**, by $\Delta^{\sqsubseteq}$ we mean the divergence corresponding to $\sqsubseteq$ by Proposition 8 (in fact, we have $\widetilde{\Delta^{\sqsubseteq}}(1)I = \sqsubseteq_I$ for all set I).

6 Properties of Divergences on Monads

6.1 Divergences on Monads as Structures in $\mathbf{Div}_{\mathcal{Q}}(\mathbb{C})$

In this section we examine divergences on monads from the view point of monoidal structure of $\mathbf{Div}_{\mathcal{Q}}(\mathbb{C})$. For any CC $\mathbb{C}$, the category $\mathbf{Div}_{\mathcal{Q}}(\mathbb{C})$ has a symmetric monoidal structure, whose unit and tensor product are given by

$$\mathbf{I} \triangleq (1, \lambda(x_1, x_2) \ . \ 0),$$
$$(I, d) \otimes (J, e) \triangleq (I \times J, \lambda(\langle x_1, y_1 \rangle, \langle x_2, y_2 \rangle) \ . \ d(x_1, x_2) + e(y_1, y_2)).$$

The coherence isomorphisms of this symmetric monoidal structure are inherited from the Cartesian monoidal structure on $\mathbb{C}$. Moreover, $V_{\mathcal{Q},\mathbb{C}} : \mathbf{Div}_{\mathcal{Q}}(\mathbb{C}) \to \mathbb{C}$ becomes a *symmetric strict monoidal functor* of type $(\mathbf{Div}_{\mathcal{Q}}(\mathbb{C}), \mathbf{I}, \otimes) \to (\mathbb{C}, 1, (\times))$.

6.1.1 Enrichments of Kleisli Categories Induced by Divergences

Let $(\mathbb{C}, T)$ be a CC-SM. We first show that a (non-graded) divergence on a monad T *attaches* a $\mathbf{Div}_{\mathcal{Q}}(\mathbf{Set})$*-enrichment* on the Kleisli category $\mathbb{C}_T$ of T. What we mean by attaching an enrichment to an ordinary category is formulated as follows.

Definition 9. A $\mathbf{Div}_{\mathcal{Q}}(\mathbf{Set})$-*enrichment* of a category $\mathbb{D}$ is a family $\{d_{I,J} : \mathbb{D}(I,J)^2 \to \mathcal{Q}\}_{I,J\in\mathbb{D}}$ of $\mathcal{Q}$-divergences on the homset $\mathbb{D}(I,J)$ such that the following inequalities hold:

$$d_{I,I}(\mathrm{id}_I, \mathrm{id}_I) \le 0, \tag{3}$$
$$d_{I,K}(g_1 \circ f_1, g_2 \circ f_2) \le d_{J,K}(g_1, g_2) + d_{I,J}(f_1, f_2). \tag{4}$$

Such an enrichment determines a $\mathbf{Div}_{\mathcal{Q}}(\mathbf{Set})$-enriched category $\mathbb{D}^d$, whose object collection and homobjects are given by

$$\mathbf{Obj}(\mathbb{D}^d) \triangleq \mathbf{Obj}(\mathbb{D}), \quad \mathbb{D}^d(I,J) \triangleq (\mathbb{D}(I,J), d_{I,J}).$$

The identity and composition morphisms of $\mathbb{D}^d$:

$$j_I : \mathbf{I} \to \mathbb{D}^d(I,I), \quad m_{I,J,K} : \mathbb{D}^d(J,K) \otimes \mathbb{D}^d(I,J) \to \mathbb{D}^d(I,K)$$

are inherited from $\mathbb{D}$; they are guaranteed to be nonexpansive by the conditions (3) and (4). The change of base of enrichment of $\mathbb{D}^d$ by the symmetric strict monoidal functor $V_{\mathcal{Q},\mathbb{D}} : \mathbf{Div}_{\mathcal{Q}}(\mathbb{D}) \to \mathbb{D}$ coincides with $\mathbb{D}$. [4]

We relate conditions (3) and (4) with the unit reflexivity and composability conditions in the definition of divergence on monad (Definition 6).

Theorem 2. *Let $(\mathbb{C}, T)$ be a CC-SM, $E : \mathbb{C} \to \mathbf{BRel}(\mathbb{C})$ be a basic endorelation such that $R_{E1} \ne \emptyset$ [5] , $\mathcal{Q}$ be a divergence domain and $\Delta = \{\Delta_I : (U(TI))^2 \to \mathcal{Q}\}_{I\in\mathbb{C}}$ be a family of $\mathcal{Q}$-divergences on TI. Define a family $d = \{d_{I,J}\}_{I,J\in\mathbb{C}}$ of $\mathcal{Q}$-divergences on the homset $\mathbb{C}_T(I,J)$ of the Kleisli category $\mathbb{C}_T$ by*

$$d_{I,J}(f_1, f_2) \triangleq \sup_{(x_1,x_2)\in EI} \Delta_J(f_1 \bullet x_1, f_2 \bullet x_2). \tag{5}$$

Then d is a $\mathbf{Div}_{\mathcal{Q}}(\mathbf{Set})$-enrichment of $\mathbb{C}_T$ if and only if Δ is an E-relative $\mathcal{Q}$-divergence on T.

6.1.2 Internalizing Divergences as Structures in $\mathbf{Div}_{\mathcal{Q}}(\mathbb{C})$

One might wonder how the $\mathcal{Q}$-divergence (5) given to each homset of $\mathbb{C}_T$ arises. Under a strengthened assumption, we derive it from the *closed structure* with respect to the monoidal product of $\mathbf{Div}_{\mathcal{Q}}(\mathbb{C})$. This allows us to *internalize* divergences on monads as structures in $\mathbf{Div}_{\mathcal{Q}}(\mathbb{C})$.

Let $(\mathbb{C}, T)$ be a CCC-SM and $\mathcal{Q}$ be a divergence domain whose monoid operation $(+)$ preserves the largest element $\top \in \mathcal{Q}$, that is, $x + \top = \top$. A consequence of this strengthened assumption is the following:

Lemma 1. *Let $(I, d) \in \mathbf{Div}_{\mathcal{Q}}(\mathbb{C})$ be an object such that $d(x_1, x_2)$ takes only values in $\{0, \top\} \subseteq \mathcal{Q}$. Then the functor $(-) \otimes (I, d) : \mathbf{Div}_{\mathcal{Q}}(\mathbb{C}) \to \mathbf{Div}_{\mathcal{Q}}(\mathbb{C})$ has a right adjoint, which we denote by $(I, d) \multimap (-)$. Moreover, $V_{\mathcal{Q},\mathbb{C}} : \mathbf{Div}_{\mathcal{Q}}(\mathbb{C}) \to \mathbb{C}$ is a map of adjunction of type:*

$$V_{\mathcal{Q},\mathbb{C}} : ((-) \otimes (I,d) \dashv (I,d) \multimap (-)) \to ((-) \times I \dashv I \Rightarrow (-)).$$

The proof of this lemma exhibits that the $\mathcal{Q}$-divergence h associated to the internal hom object $(I, d) \multimap (J, e)$ measures the divergence between $f_1, f_2 \in U(I \Rightarrow J)$ by

$$h(f_1, f_2) = \sup_{x_1,x_2\in UI, d(x_1,x_2)=0} e(\lfloor f_1 \rfloor \bullet x_1, \lfloor f_2 \rfloor \bullet x_2),$$

[4] The underlying category of $\mathbb{D}^d$ [34, Section 1.3] does *not* coincide with $\mathbb{D}$.

[5] $R_{E1} = \emptyset$ happens if and only if $R_{EI} = \emptyset$ for any $I \in \mathbb{C}$. Therefore nontrivial basic endorelations always satisfy $R_{E1} \ne \emptyset$.

which almost coincides with the sup part of (5); here $\lfloor - \rfloor : U(I \Rightarrow J) \to \mathbb{C}(I, J)$ is the bijection given in Section 2. We use this coincidence to characterize the unit-reflexivity and composability conditions in the definition of divergence on monad (Definition 6). First, we define the *internal Kleisli extension morphism* $kl_{I,J} : TI \times (I \Rightarrow TJ) \to TJ$ by

$$kl_{I,J} \triangleq\ TI \times (I \Rightarrow TJ) \xrightarrow{\langle \pi_2, \pi_1 \rangle} (I \Rightarrow TJ) \times TI \xrightarrow{\theta_{I \Rightarrow TJ, I}} T((I \Rightarrow TJ) \times I) \xrightarrow{\mathrm{ev}^\#} TJ\ . \qquad (6)$$

Next, for a basic endorelation $E : \mathbb{C} \to \mathbf{BRel}(\mathbb{C})$, we define the functor $E' : \mathbb{C} \to \mathbf{Div}_{\mathcal{Q}}(\mathbb{C})$ by

$$E'I \triangleq (I, d_{E'I}), \quad E'f \triangleq f, \quad \text{where} \quad d_{E'I}(x_1, x_2) \triangleq \begin{cases} 0 & (x_1, x_2) \in E \\ \infty & (x_1, x_2) \notin E. \end{cases}$$

Theorem 3. *Let $(\mathbb{C}, T)$ be a CCC-SM, $(M, \leq, 1, (\cdot))$ be a grading monoid, $\mathcal{Q}$ be a divergence domain whose monoid operation $(+)$ satisfies $x + \top = \top$, and $E : \mathbb{C} \to \mathbf{BRel}(\mathbb{C})$ be a basic endorelation. Let $\Delta = \{\Delta^m_I\}_{m \in M, I \in \mathbb{C}}$ be a doubly-indexed family of $\mathcal{Q}$-divergences on TI, regarded as $\mathbf{Div}_{\mathcal{Q}}(\mathbb{C})$-objects. Then*

1. *Δ satisfies the E-unit reflexivity condition if and only if for any $I \in \mathbb{C}$, the following nonexpansivity holds on the global element $\ulcorner \eta_I \urcorner : 1 \to I \Rightarrow TI$ corresponding to the monad unit:*
$$\ulcorner \eta_I \urcorner \in \mathbf{Div}_{\mathcal{Q}}(\mathbb{C})(\mathbf{I}, E'I \multimap \Delta^1_I).$$

2. *Δ satisfies the E-composablity condition if and only if for any $I, J \in \mathbb{C}$ and $m, n \in M$, the following nonexpansivity holds on the internal Kleisli extension morphism $kl_{I,J} : TI \times (I \Rightarrow TJ) \to TJ$:*
$$kl_{I,J} \in \mathbf{Div}_{\mathcal{Q}}(\mathbb{C})(\Delta^m_I \otimes (E'I \multimap \Delta^n_J), \Delta^{m \cdot n}_J).$$

[5] formalized families of composable divergences as *parameterized assignment* in *weakly closed monoidal refinement*. Roughly speaking, they adopted the equivalence (2) of Theorem 3 as the definition of parameterized assignment. However, divergence on monads and parameterized assignments are built on slightly different categorical foundations, and their generalities are incomparable. Notable differences from parameterized assignment are: 1) divergences on monads are defined in relative to basic endorelations, and 2) the underlying category of divergences on monads is any CCs, while parameterized assignments requires closed structure on their underlying category. In this sense divergences on monads are a mild generalization of parameterized assignments.

6.1.3 Divergences on Monads and Divergence Liftings of Monads

We next relate graded divergences on monads and monad-like structures on the category $\mathbf{Div}_{\mathcal{Q}}(\mathbb{C})$ of $\mathcal{Q}$-divergences on $\mathbb{C}$-objects. What we mean by monad-like structures is *graded divergence liftings* of monads on $\mathbb{C}$, which we introduce below. It is a graded monad on $\mathbf{Div}_{\mathcal{Q}}(\mathbb{C})$ ([31]) whose unit and multiplication are inherited from a monad on $\mathbb{C}$.

Definition 10. Let $(\mathbb{C}, T)$ be a CC-SM, M be a grading monoid and $\mathcal{Q}$ be a divergence domain. An M-graded $\mathcal{Q}$-divergence lifting of T is an mapping $\dot{T} : M \times \mathbf{Obj}(\mathbf{Div}_{\mathcal{Q}}(\mathbb{C})) \to \mathbf{Obj}(\mathbf{Div}_{\mathcal{Q}}(\mathbb{C}))$ such that (below V stands for the forgetful functor $V_{\mathcal{Q},\mathbb{C}} : \mathbf{Div}_{\mathcal{Q}}(\mathbb{C}) \to \mathbb{C}$)

1. $V(\dot{T}mX) = T(VX)$
2. $m \leq n$ implies $\dot{T}mX \leq \dot{T}nX$

3. $\eta_{VX} \in \mathbf{Div}_{\mathcal{Q}}(\mathbb{C})(X, \dot{T}1X)$

4. $\mu_{VX} : \mathbf{Div}_{\mathcal{Q}}(\mathbb{C})(\dot{T}m(\dot{T}nX), \dot{T}(m \cdot n)X)$.

Let $E : \mathbb{C} \to \mathbf{BRel}(\mathbb{C})$ be a basic endorelation. We say that an M-graded $\mathcal{Q}$-divergence lifting $\dot{T}$ of T is *E-strong* if the strength θ of T satisfies

$$\theta_{VX,J} \in \mathbf{Div}_{\mathcal{Q}}(\mathbb{C})(X \otimes \dot{T}m(E'J), \dot{T}m(X \otimes E'J)).$$

We write $\mathbf{SGDLift}(T, E, M, \mathcal{Q})$ for the collection of E-strong M-graded $\mathcal{Q}$-divergence liftings of T. We introduce a partial order $\preceq$ on $\mathbf{SGDLift}(T, E, M, \mathcal{Q})$ by

$$\dot{T} \preceq \dot{S} \iff \forall m \in M, X \in \mathbf{Div}_{\mathcal{Q}}(\mathbb{C}), c_1, c_2 \in U(T(VX)) \ . \ d_{\dot{T}mX}(c_1, c_2) \geq d_{\dot{S}mX}(c_1, c_2).$$

We will later see a similar concept of *strong graded relational lifting* of monad in Definition 15. Divergence lifting and relational lifting are actually instances of a common general definition of *strong graded lifting of monad* ([31]), but in this paper we omit this general definition.

The following theorem relates that every divergence can be expressed as the composite of a graded divergence lifting and the divergence corresponding to a basic endorelation.

Theorem 4. *Let $(\mathbb{C}, T)$ be a CC-SM, M be a grading monoid, $\mathcal{Q}$ be a divergence domain and $E : \mathbb{C} \to \mathbf{BRel}(\mathbb{C})$ be a basic endorelation. For any $\Delta \in \mathbf{Div}(T, E, M, \mathcal{Q})$, define a mapping $[\Delta] : M \times \mathbf{Obj}(\mathbf{Div}_{\mathcal{Q}}(\mathbb{C})) \to \mathbf{Obj}(\mathbf{Div}_{\mathcal{Q}}(\mathbb{C}))$ by, for $X = (I, d)$, $[\Delta]mX \triangleq (TI, d_{[\Delta]mX})$ where*

$$d_{[\Delta]mX}(c_1, c_2) \triangleq \sup_{J \in \mathbb{C}, n \in M, f \in \mathbf{Div}_{\mathcal{Q}}(\mathbb{C})(X, \Delta^n_J)} \Delta^{m \cdot n}_J(f^\sharp \bullet c_1, f^\sharp \bullet c_2).$$

Then $[\Delta]$ is an M-graded $\mathcal{Q}$-divergence lifting $\dot{T}$ such that $\Delta^m_I = [\Delta]m(E'I)$.

When $M = 1$, Theorem 4 implies that the assignment $I \mapsto \Delta_I$ extends to the *E'-relative monad* $[\Delta] \circ E' : \mathbb{C} \to \mathbf{Div}_{\mathcal{Q}}(\mathbb{C})$ in the sense of [3].

When we strengthen the assumptions on $(\mathbb{C}, T)$ and $\mathcal{Q}$ as done in Section 6.1.2, we obtain a sharper correspondence between divergences on monads and strong graded divergence liftings of monads.

Theorem 5. *Let $(\mathbb{C}, T)$ be a CCC-SM, M be a grading monoid, $\mathcal{Q}$ be a divergence domain such that $(+)$ satisfies $x + \top = \top$ and $E : \mathbb{C} \to \mathbf{BRel}(\mathbb{C})$ be a basic endorelation. Then there exists an adjunction between partial orders:*

$$\begin{tikzcd} (\mathbf{SGDLift}(T, E, M, \mathcal{Q}), \preceq) \arrow[r, shift left, "\langle - \rangle"] \arrow[r, phantom, "\perp"] & (\mathbf{Div}(T, E, M, \mathcal{Q}), \preceq) \arrow[l, shift left, "{[-]}"] \end{tikzcd}$$

where $\langle \dot{T} \rangle mI \triangleq \dot{T}m(E'I)$

6.2 Generation of Divergences

It has been shown that DP can be interpreted as hypothesis testing ([54, 30]). Given a query $c \colon I \to GJ$ and adjacent datasets $(d_1, d_2) \in R_{\mathrm{adj}} \subseteq I^2$, we consider the following hypothesis testing with the null and alternative hypotheses:

H_0: The output y comes from the dataset d_1,

H_1: The output y comes from the dataset d_2.

For any rejection region $S \in \Sigma_J$, the Type I and Type II errors are then represented by $\Pr[c(d_1) \in S]$ and $\Pr[c(d_2) \notin S]$, respectively. [30] showed that c is (ε, δ)-DP *if and only if* for any adjacent datasets $(d_1, d_2) \in R_{\mathrm{adj}} \subseteq I^2$, the pair of Type I error and Type II error lands in the *privacy region* $R(\varepsilon, \delta)$:

$$\forall S \in \Sigma_J \ . \ (\Pr[c(d_1) \in S], \Pr[c(d_2) \notin S]) \in \underbrace{\{(x, y) \in [0,1]^2 | (1 - x) \leq \exp(\varepsilon) y + \delta\}}_{\triangleq R(\varepsilon,\delta)} .$$

They also showed that this is equivalent to the testing using probabilistic decision rules [30, Corollary 2.3]:

$$\forall k \colon J \to G\{\mathsf{Acc}, \mathsf{Rej}\} \ . \ (\Pr[k^\sharp c(d_1) = \mathsf{Acc}], \Pr[k^\sharp c(d_2) = \mathsf{Rej}]) \in R(\varepsilon, \delta).$$

Later [7] generalized this probabilistic variant of hypothesis testing to general statistical divergences, and arrived at a notion of *k-generatedness* of statistical divergences ($k \in \mathbb{N} \cup \{\infty\}$). Following their generalization, we introduce the concept of *Ω-generatedness* of divergences on monads.

Definition 11. Let $\Omega \in \mathbb{C}$. A divergence $\Delta \in \mathbf{Div}(T, E, M, \mathcal{Q})$ is *Ω-generated* if for any $m \in M$, $I \in \mathbb{C}$ and $c_1, c_2 \in U(TI)$,

$$\Delta^m_I(c_1, c_2) = \sup_{k \colon I \to T\Omega} \Delta^m_\Omega(k^\sharp \bullet c_1, k^\sharp \bullet c_2).$$

An equivalent definition of $\Delta \in \mathbf{Div}(T, E, M, \mathcal{Q})$ being Ω-generated is: the following holds for any $m \in M, I \in \mathbb{C}, c_1, c_2 \in U(TI), v \in \mathcal{Q}$:

$$\Delta^m_I(c_1, c_2) \leq v \iff \forall k \colon I \to T\Omega \ . \ (k^\sharp \bullet c_1, k^\sharp \bullet c_2) \in \tilde{\Delta}(m, v)\Omega.$$

Here $\tilde{\Delta}(m, v)\Omega$ is the binary relation $\{(c_1, c_2) \mid \Delta^m_\Omega(c_1, c_2) \leq v\}$; see also (9). For an Ω-generated divergence Δ, its component Δ^m_Ω at Ω is an essential part that determines all components Δ^m_I of Δ. When a divergence is shown to be Ω-generated, the calculation of the codensity lifting $T^{[\Delta]}$ given in Section 7 will be simplified (Section 7.1).

We illustrate Ω-generatedness of various divergences. First, we show the Ω-generatedness of divergences on the Giry monad G in Tables 2 and 3.

- Divergence DP is generated over the two-point discrete space 2 [7, Section B.7]. The binary relation $(\widetilde{\mathsf{DP}}(\varepsilon, \delta)2)$ coincides with the privacy region $R(\varepsilon, \delta)$.
- Divergence TV is also generated over 2 [7, Section C.1].
- Divergences Re^α, Chi, HD and KL are generated over the countably infinite discrete space $\mathbb{N}$. In contrast, they are not N-generated for every finite discrete space N [7, Sections B.5 and B.9].

On the sub-Giry monad G_s, the divergence DP is 1-generated, and the total variation distance TV is 2-generated.

Proposition 9. *The divergence* $\mathsf{DP} \in \mathbf{Div}(G_s, \mathrm{Eq}, \mathcal{R}^+, \mathcal{R}^+)$ *is* 1*-generated.*

Proposition 10. *The divergence* $\mathsf{TV} \in \mathbf{Div}(G_s, \mathrm{Eq}, 1, \mathcal{R}^+)$ *is not* 1*-generated but* 2*-generated.*

Ω-Generatedness of Preorders on Monads We relate Ω-generatedness of divergences and preorders on monads studied in ([32]). Let T be a monad on **Set** and Ω be a set. [32] introduced the concept of *congruent* and *substitutive* preorders on $T\Omega$ as those satisfying:

Substitutivity For any function $f\colon \Omega \to T\Omega$ and $c_1, c_2 \in T\Omega$, $c_1 \le c_2$ implies $f^\sharp(c_1) \le f^\sharp(c_2)$.

Congruence For any function $f_1, f_2\colon J \to T\Omega$, if $f_1(x) \le f_2(x)$ holds for any $x \in J$, then $f_1^\sharp(c) \le f_2^\sharp(c)$ holds for any $c \in T\Omega$.

For instance, any component of a preorder on T at Ω forms a congruent and substitutive preorder on $T\Omega$. We write $\mathbf{CSPre}(T, \Omega)$ for the set of all congruent and substitutive preorders on $T\Omega$, and $\mathbf{Pre}(T)$ for the collection of all preorders on T. [32] gave a construction $[-]^\Omega\colon \mathbf{CSPre}(T, \Omega) \to \mathbf{Pre}(T)$ of preorders on T from congruent and substitutive preorders on $T\Omega$:

$$c_1 [\le]^\Omega_J c_2 \iff \forall g\colon J \to T\Omega \,.\, g^\sharp(c_1) \le g^\sharp(c_2)$$

The constructed preorders on T are Ω-generated in the following sense:

Proposition 11. *For any $\le \in \mathbf{CSPre}(T, \Omega)$, the $\mathcal{B}$-divergence $\Delta^{[\le]^\Omega}$ corresponding to the preorder $[\le]^\Omega$ on T is Ω-generated (see Proposition 8 for the correspondence).*

Applying this proposition, we can determine Ω-generatedness of preorders on monads:

- If the monad T has a rank α, the construction $[-]^\alpha$ is bijective [32, Theorem 7]. Hence for such a monad, each preorder on T corresponds to an α-generated $\mathcal{B}$-divergence.
- For the subprobability distribution monad D_s on **Set**, [50] identified all preorders on D_s: there are 41 preorders on D_s. Among them, 25 preorders are 1-generated, while 16 preorders are 2-generated [50, Proposition 6.3].

6.3 An Adjunction between Quantitative Equational Theories and Divergences

[39] introduced a concept of *quantitative equational theory* as an algebraic presentation of monads on the category of (pseudo-)metric spaces. A quantitative equational theory is an equational theory with indexed equations $t =_\varepsilon u$ having the axioms of pseudometric spaces, plus suitable axioms reflecting properties of quantitative algebras. A quantitative equational theory determines a pseudometric on the set of Ω-terms.

Consider a set Ω of function symbols of finite arity. If n is the arity of a function $f \in \Omega$, we write $f\colon n \in \Omega$. Let X be a set of variables, and let $T_\Omega X$ be the Ω-term algebra over X. For $f\colon n \in \Omega$ and $t_1, \dots, t_n \in T_\Omega X$, we write $f(t_1, \dots, t_n)$ for the term obtained by applying f to $t_1, \dots, t_n$. The construction $X \mapsto T_\Omega X$ forms a (strong) monad on **Set** whose unit sends variables to terms, that is, $\eta_X(x) = x$, and Kleisli extension $h^\sharp\colon T_\Omega I \to T_\Omega X$ of function $h\colon I \to T_\Omega X$ is defined inductively by

$$h^\sharp(x) \triangleq h(x), \quad h^\sharp(f(t_1, \dots, t_n)) \triangleq f(h^\sharp(t_1), \dots, h^\sharp(t_n)).$$

A substitution of Ω-terms over X is a function $\sigma\colon X \to T_\Omega X$. For $t \in T_\Omega X$, we call $\sigma^\sharp(t)$ the substitution of σ to t. We define the set of *indexed equations* of terms by

$$\mathbb{V}(T_\Omega X) \triangleq \{t =_\varepsilon u \mid t, u \in T_\Omega X, \varepsilon \in \mathbb{Q}^+\}.$$

$$\emptyset \vdash t =_0 t \in U \tag{Ref}$$
$$\{t =_\varepsilon u\} \vdash u =_\varepsilon t \in U \tag{Sym}$$
$$\{t =_\varepsilon u, u =_{\varepsilon'} v\} \vdash t =_{\varepsilon+\varepsilon'} v \in U \tag{Tri}$$
$$\forall \varepsilon' \in \mathbb{Q}^+ \ . \ \{t =_\varepsilon u\} \vdash t =_{\varepsilon+\varepsilon'} u \in U \tag{Max}$$
$$\forall \varepsilon \in \mathbb{Q}^+ \ . \ \{t =_{\varepsilon'} u | \varepsilon < \varepsilon'\} \vdash t =_\varepsilon u \in U \tag{Arch}$$
$$\forall f \colon n \in \Omega \ . \ \{t_i =_\varepsilon u_i | 1 \leq i \leq n\} \vdash f(t_1, \dots, t_n) =_\varepsilon f(u_1, \dots, u_n) \tag{Nonexp}$$
$$\forall \sigma \colon X \to T_\Omega X \ . \ \Gamma \vdash t =_\varepsilon u \in U \implies \sigma(\Gamma) \vdash \sigma^\sharp(t) =_\varepsilon \sigma^\sharp(u) \in U \tag{Subst}$$
$$\Gamma' \vdash t =_\varepsilon u \in U \wedge \forall \psi \in \Gamma' \ . \ \Gamma \vdash \psi \in U \implies \Gamma \vdash t =_\varepsilon u \in U \tag{Cut}$$
$$t =_\varepsilon u \in \Gamma \implies \Gamma \vdash t =_\varepsilon u \in U \tag{Assumpt}$$

Figure 1: Quantitative Equational Theory Rules

Here the index ε runs over non-negative rational numbers. A *conditional quantitative equation* is a judgment of the following form

$$\{t_i =_{\varepsilon_i} u_i \mid i \in I\} \vdash t =_\varepsilon u \qquad (I \colon \text{countable}, t_i =_{\varepsilon_i} u_i, t =_\varepsilon u \in \mathbb{V}(T_\Omega X));$$

the left hand side of turnstile ($\vdash$) is called hypothesis and the right hand side conclusion. We denote by $\mathbb{E}(T_\Omega X)$ the set of conditional quantitative equations. For any countable subset Γ of $\mathbb{V}(T_\Omega X)$ and any substitution $\sigma \colon X \to T_\Omega X$, we define $\sigma(\Gamma) \triangleq \{\sigma^\sharp(t_i) =_{\varepsilon_i} \sigma^\sharp(u_i) \mid t_i =_{\varepsilon_i} u_i \in \Gamma\}$.

Definition 12 (Quantitative Equational Theory [39, Definition 2.1])**.** A quantitative equational theory (QET for short) of type Ω over X is a set $U \subseteq \mathbb{E}(T_\Omega X)$ closed under the rules summarized as Figure 1. We write $\mathbf{QET}(\Omega, X)$ for the set of QETs of type Ω over X. We regard it as a poset $(\mathbf{QET}(\Omega, X), \subseteq)$ by the set inclusion order. Given a set U_0 of conditional quantitative equations of type Ω over X, by $\overline{U_0}^{\mathrm{QET}(\Omega,X)}$ we mean the least QET containing U_0.

We state an adjunction between quantitative equational theories and divergences on free-algebra monads on **Set**. More specifically, we construct the following adjunction and isomorphism between posets:

$$\begin{tikzcd} (\mathbf{QET}(\Omega, X), \subseteq) \arrow[r, shift left, "d[-]"] \arrow[r, phantom, "\perp"] & (\mathbf{CSEPMet}(T_\Omega, X), \preceq) \arrow[l, shift left, "U[-]"] \arrow[r, shift left, "\mathrm{Gen}"] \arrow[r, phantom, "\cong"] & (\mathbf{DivEPMet}(T_\Omega, X), \preceq) \arrow[l, shift left, "(-)_X"] \end{tikzcd} \ . \tag{7}$$

By combining these, a QET of type Ω over X determines an X-generated Eq-relative $\mathcal{R}^+$-divergence on T_Ω and vice versa. The poset in the middle is that of *congruent* and *substitutive pseudometrics*, which are a quantitative analogue of congruent and substitutive preorders.

Definition 13. Let T be a monad on **Set** and $X \in$ **Set**. A congruent and substitutive pseudometric (CS-EPMet for short) on TX is an extended pseudometric[6] $d \colon (TX)^2 \to \mathcal{R}^+$ on TX satisfying

Substitutivity For all function $fX \to TX$ and $c_1, c_2 \in TX$, $d(f^\sharp(c_1), f^\sharp(c_2)) \leq d(c_1, c_2)$.

[6]A function $d \colon A^2 \to \mathcal{R}^+$ is called an extended pseudometric on A if $d(a, a) = 0$ (reflexivity), $d(b, a) = d(a, b)$ (symmetry) and $d(a, c) \leq d(a, b) + d(b, c)$ (triangle-inequality) hold for all $a, b, c \in A$.

Congruence For all set I, function $f_1, f_2 \colon I \to TX$ and $c \in TI$, $d(f_1^\sharp(c), f_2^\sharp(c)) \leq \sup_{i \in I} d(f_1(i), f_2(i))$.

We denote by $\mathbf{CSEPMet}(T, X)$ the set of CS-EPMets on TX. We then make it into a poset $(\mathbf{CSEPMet}(T, X), \preceq)$ by the following pointwise opposite order:

$$d \preceq d' \iff \forall c_1, c_2 \in TX \ . \ d(c_1, c_2) \geq d'(c_1, c_2).$$

Definition 14. Let T be a monad on $\mathbf{Set}$ and $X \in \mathbf{Set}$. We denote by $\mathbf{DivEPMet}(T, X)$ the collection of X-generated Eq-relative $\mathcal{R}^+$-divergences Δ on T such that each component Δ_I is an extended pseudometric. We restrict the partial order $\preceq$ on $\mathbf{Div}(T, \mathrm{Eq}, 1, \mathcal{R}^+)$ to $\mathbf{DivEPMet}(T, X)$.

We next introduce various monotone functions appearing in (7).

$$d[U](t, u) \triangleq \inf \left\{ \varepsilon \in \mathbb{Q}^+ \;\middle|\; \emptyset \vdash t =_\varepsilon u \in U \right\} \qquad \mathrm{Gen}(d)_I(c_1, c_2) \triangleq \sup_{k \colon I \to TX} d(k^\sharp(c_1), k^\sharp(c_2))$$

$$U[d] \triangleq \overline{\{\emptyset \vdash t =_\varepsilon u \mid \varepsilon \in \mathbb{Q}^+, d(t, u) \leq \varepsilon\}}^{\mathrm{QET}(\Omega, X)} \qquad (\Delta)_X \triangleq \Delta_X$$

Proposition 12. *The functions $d[-], U[-], \mathrm{Gen}, (-)_X$ defined above are all well-defined monotone functions having types given in* (7).

That $d[U]$ is an extended pseudometric is shown in the beginning of [39, Section 5]. Here we additionally show that it enjoys congruence and substitutivity of Definition 13. The function Gen is taken from the right hand side of the definition of Ω-generatedness (Definition 11). The function $(-)_X$ simply extracts the X-th component of a given divergence.

Theorem 6. *For any set Ω of function symbols with finite arity and set X, the following holds for the monotone functions in* (7)*:*

1. Gen *is the inverse of* $(-)_X$.
2. *We have an adjunction satisfying* $d[U[-]] = \mathrm{id}$*:*

$$(\mathbf{QET}(\Omega, X), \subseteq) \underset{U[-]}{\overset{d[-]}{\underset{\longleftarrow}{\overset{\longrightarrow}{\perp}}}} (\mathbf{CSEPMet}(T_\Omega, X), \preceq) \tag{8}$$

In the proof of this theorem, we used the definition of models of QET ([6]). Intuitively, the right adjoint $d[-]$ extracts the pseudometric on $T_\Omega X$ from a given QET. The left adjoint $U[-]$ constructs the least QET containing all information of a given pseudometric on $T_\Omega X$. The adjunction (8) also implies that we can construct monads on the category of extended metric spaces from CS-EPMets by Mardare et al.'s metric term monad construction ([39]). Overall adjunction (7) says that X-generated divergences can be axiomatized with QETs whose variable set is X.

The range of $U[-]$ is a subset of $\mathbf{UQET}(\Omega, X)$ of *unconditional QETs* defined below (See also [38, Section 3]):

$$\mathbf{UQET}(\Omega, X) \triangleq \left\{ V \in \mathbf{QET}(\Omega, X) \;\middle|\; \exists S \subseteq \{\emptyset \vdash t =_\varepsilon u \mid t, u \in T_\Omega X, \varepsilon \in \mathbb{Q}^+\}.\ V = \overline{S}^{\mathrm{QET}(\Omega, X)} \right\}.$$

Unconditional QETs of type Ω over X are *equivalent to* X-generated divergence on T_Ω: restricting QETs to unconditional QETs, the adjunction (8) becomes a pair of isomorphisms.

Theorem 7. $(\mathbf{UQET}(\Omega, X), \subseteq) \cong (\mathbf{CSEPMet}(T_\Omega, X), \preceq) \cong (\mathbf{DivEPMet}(T_\Omega, X), \preceq)$.

7 Graded Strong Relational Liftings for Divergences

We have introduced the concept of divergence on monad for measuring quantitative difference between two computational effects. To integrate this concept with relational program logic, we employ a semantic structure called *graded strong relational lifting* of monad. It is introduced for the semantics of approximate probabilistic relational Hoare logic for the verification of differential privacy ([12]), then later used in various program logics ([13, 8, 9, 51, 52]). Independently, it is also introduced as a semantic structure for effect system ([31]). Liftings introduced in the study of differential privacy are designed to satisfy a special property called *fundamental property* [12, Theorem 1]: when we supply the equivalence relation to the lifting, it returns the adjacency relation of the divergence. This special property is the key to express the differential privacy of probabilistic programs in relational program logics.

In this paper, we present a *general construction* of graded strong relational liftings from divergences on monads. First, we recall its definition ([31, 24]).

Definition 15. Let $(\mathbb{C}, T)$ be a CC-SM and $(M, \leq, 1, (\cdot))$ be a grading monoid. An *M-graded strong relational lifting* $\dot{T}$ of T is a mapping $\dot{T} : M \times \mathbf{Obj}(\mathbf{BRel}(\mathbb{C})) \to \mathbf{Obj}(\mathbf{BRel}(\mathbb{C}))$ satisfying the following conditions:

1. $p_{\mathbb{C}}(\dot{T}mX) = (TX_1, TX_2)$, and $m \leq m'$ implies $\dot{T}mX \leq \dot{T}m'X$.
2. $(\eta_{X_1}, \eta_{X_2}) : X \dot{\to} \dot{T}1(X)$.
3. $(f_1, f_2) : X \dot{\to} \dot{T}m(Y)$ implies $(f_1^\sharp, f_2^\sharp) : \dot{T}m'X \dot{\to} \dot{T}(m \cdot m')Y$.
4. $(\theta_{X_1,Y_1}, \theta_{X_2,Y_2}) : X \dot{\times} \dot{T}mY \dot{\to} \dot{T}m(X \dot{\times} Y)$.

Our interest is in the graded strong relational lifting that carries the information of a given divergence $\Delta \in \mathbf{Div}(T, E, M, \mathcal{Q})$. We identify such liftings by the following *fundamental property.* First define the *adjacency relation* of Δ by

$$\tilde{\Delta}(m, v)I \triangleq (TI, TI, \{(c_1, c_2) \mid \Delta_I^m(c_1, c_2) \leq v\}) \quad (m \in M, v \in \mathcal{Q}, I \in \mathbb{C}). \tag{9}$$

Note that $\tilde{\Delta}$ is monotone on m and v.

Definition 16. We say that an $M \times \mathcal{Q}$-graded strong relational lifting $\dot{T}$ of T satisfies the *fundamental property* with respect to $\Delta \in \mathbf{Div}(T, E, M, \mathcal{Q})$ if the following holds:

$$\dot{T}(m, v)(EI) = \tilde{\Delta}(m, v)I \quad (m \in M, v \in \mathcal{Q}, I \in \mathbb{C}).$$

Theorem 8. *Let $(\mathbb{C}, T)$ be a CC-SM, $(M, \leq, 1, (\cdot))$ be a grading monoid, $\mathcal{Q}$ be a divergence domain and $\Delta = \{\Delta_I^m \colon (U(TI))^2 \to \mathcal{Q}\}_{m \in M, I \in \mathbb{C}}$ be a doubly-indexed family of $\mathcal{Q}$-divergences satisfying monotonicity on m (Definition 6). Define the following mapping $T^{[\Delta]} : (M \times \mathcal{Q}) \times \mathbf{Obj}(\mathbf{BRel}(\mathbb{C})) \to \mathbf{Obj}(\mathbf{BRel}(\mathbb{C}))$:*

$$\begin{aligned} T^{[\Delta]}(m, v)X \triangleq (TX_1, TX_2, \{(c_1, c_2) \mid \forall I \in \mathbb{C}, n \in M, w \in \mathcal{Q}, (k_1, k_2) : X \dot{\to} \tilde{\Delta}(n, w)I\ . \\ (k_1^\sharp \bullet c_1, k_2^\sharp \bullet c_2) \in \tilde{\Delta}(m \cdot n, v + w)I\}) \end{aligned}$$

1. *The mapping $T^{[\Delta]}$ is an $M \times \mathcal{Q}$-graded strong relational lifting of T.*
2. *Let $E : \mathbb{C} \to \mathbf{BRel}(\mathbb{C})$ be a basic endorelation. Then*

$$\Delta \textit{ is } E\textit{-unit-reflexive} \iff \forall I \in \mathbb{C}, (m, v) \in M \times \mathcal{Q}\ .\ T^{[\Delta]}(m, v)(EI) \leq \tilde{\Delta}(m, v)I \tag{S}$$

$$\Delta \textit{ is } E\textit{-composable} \iff \forall I \in \mathbb{C}, (m, v) \in M \times \mathcal{Q}\ .\ T^{[\Delta]}(m, v)(EI) \geq \tilde{\Delta}(m, v)I. \tag{C}$$

The construction of $T^{[\Delta]}$ is a graded extension of the *codensity lifting* ([51, 33]). The remainder of this section is the proof of Theorem 8.

Proof. (Proof of (1)) Proving conditions 1-3 of graded strong relational lifting (Definition 15) are routine generalization of [33] and [31, Section 5]; thus omitted here (see Lemma 4 in appendix).

However, condition 4 of Definition 15 needs a special attention because in general codensity lifting does not automatically lift strength. The current setting works because of our particular choice of the category of binary relations over $\mathbb{C}$. We prove condition 4 as follows. Since $f_i \bullet j = f \bullet \langle i, j\rangle$ for any $j \in UJ$ holds, we have the equivalence

$$\begin{aligned}(f,g) : X \dot{\times} Y \dot{\rightarrow} Z &\iff \forall (x,x') \in X, (y,y') \in Y. (f \bullet \langle x,y\rangle, g \bullet \langle x', y'\rangle) \in Z \\ &\iff \forall (x,x') \in X, (y,y') \in Y. ((f_x) \bullet y, (g_{x'}) \bullet y') \in Z \\ &\iff \forall (x,x') \in X. (f_x, g_{x'}) \colon Y \dot{\rightarrow} Z.\end{aligned}$$

From this, condition 3 (law of graded Kleisli extension), and the equation (1) on the strength of a CC-SM, we prove condition 4 from condition 2 (unit law): for all $m \in M$ and $v \in \mathcal{Q}$, we have

$$\begin{aligned}&(\eta_{X_1 \times Y_1}, \eta_{X_2 \times Y_2}) \colon X \dot{\times} Y \dot{\rightarrow} T^{[\Delta]}(1,0)(X \dot{\times} Y) \\ \iff& \forall (x,x') \in X \ . \ ((\eta_{X_1\times Y_1})_x, (\eta_{X_2 \times Y_2})_{x'}) \colon Y \dot{\rightarrow} T^{[\Delta]}(1,0)(X \dot{\times} Y) \\ \implies& \forall (x,x') \in X \ . \ (((\eta_{X_1\times Y_1})_x)^\sharp, ((\eta_{X_2 \times Y_2})_{x'})^\sharp) : T^{[\Delta]}(m,v)Y \dot{\rightarrow} T^{[\Delta]}(m,v)(X \dot{\times} Y) \\ \iff& \begin{pmatrix}\forall (x,x') \in X, (c_1, c_2) \in T^{[\Delta]}(m,v)Y \ . \\ \qquad (((\eta_{X_1\times Y_1})_x)^\sharp \bullet c_1, ((\eta_{X_2 \times Y_2})_{x'})^\sharp \bullet c_2) \in T^{[\Delta]}(m,v)(X \dot{\times} Y)\end{pmatrix} \\ \iff& \begin{pmatrix}\forall (x,x') \in X, (c_1, c_2) \in T^{[\Delta]}(m,v)Y \ . \\ \qquad (\theta_{X_1,Y_1} \bullet \langle x, c_1\rangle, \theta_{X_2,Y_2} \bullet \langle x', c_2\rangle) \in T^{[\Delta]}(m,v)(X \dot{\times} Y)\end{pmatrix} \\ \iff& \forall (x,x') \in X \ . \ ((\theta_{X_1,Y_1})_x, (\theta_{X_2,Y_2})_{x'}) \colon T^{[\Delta]}(m,v)Y \dot{\rightarrow} T^{[\Delta]}(m,v)(X \dot{\times} Y) \\ \iff& (\theta_{X_1,Y_1}, \theta_{X_2,Y_2}) \colon X \dot{\times} T^{[\Delta]}(m,v)Y \dot{\rightarrow} T^{[\Delta]}(m,v)(X \dot{\times} Y).\end{aligned}$$

(Proof of (2)-(S)) We show the equivalence of Δ being E-unit-reflexive and the implication

$$\begin{aligned}&\forall I \in \mathbb{C}, m \in M, v \in \mathcal{Q}, c, c' \in U(TI) \ . \\ &\quad (\forall J \in \mathbb{C}, m' \in M, v' \in \mathcal{Q}, (k,l) : EI \dot{\rightarrow} \tilde{\Delta}(m',v')J \ . \ \Delta_J^{m \cdot m'}(k^\sharp \bullet c, l^\sharp \bullet c') \le v + v') \\ &\qquad \implies \Delta_I^m(c,c') \le v.\end{aligned} \tag{10}$$

We suppose that the above implication holds. We fix $I \in \mathbb{C}$. Let $(i,j) \in EI$. By instantiating the whole implication with $m = 1, v = 0, c = \eta_I \bullet i, c' = \eta_I \bullet j$, the middle part of (10) becomes

$$\forall J \in \mathbb{C}, m' \in M, v' \in \mathcal{Q}, (k,l) : EI \dot{\rightarrow} \tilde{\Delta}(m',v')J \ . \ \Delta_J^{m'}(k \bullet i, l \bullet j) \le v',$$

which is trivially true. Therefore we conclude $\Delta_I^m(\eta_I \bullet i, \eta_I \bullet j) \le 0$ for any $(i,j) \in EI$, that is, E-unit reflexivity holds.

Conversely, we suppose that Δ satisfies the unit-reflexivity. We take I, m, v, c, c' of appropriate type and assume the middle part of (10). By instantiating it with $J = I, m' = 1, v' = 0, k = l = \eta_I$, we conclude $\Delta_I^m(c,c') \le v$.

(Proof of (2)-(C)) We show the equivalence of Δ being E-composable and the implication $\forall I \in \mathbb{C}, m \in M, v \in \mathcal{Q} \ . \ \tilde{\Delta}I(m,v) \le T^{[\Delta]}I(m,v)(EI)$ as follows:

$$\forall I \in \mathbb{C}, m \in M, v \in \mathcal{Q} \ . \ \tilde{\Delta}I(m,v) \le T^{[\Delta]}I(m,v)(EI)$$

$$
\begin{aligned}
&\iff \begin{pmatrix}
\forall I \in \mathbb{C}, m \in M, v \in \mathcal{Q}, c, c' \in U(TI) \ . \\
\quad \Delta^m_I(c,c') \le v \implies \\
\quad\quad \forall J \in \mathbb{C}, m' \in M, v' \in \mathcal{Q}, (k,l)\colon EI \dot{\to} \tilde{\Delta}(m',v')J \ . \\
\quad\quad\quad (k^\sharp \bullet c, l^\sharp \bullet c') \in \tilde{\Delta}(m \cdot m', v + v')J
\end{pmatrix} \\
&\iff \begin{pmatrix}
\forall I, J \in \mathbb{C}, m \in M, v \in \mathcal{Q}, c, c' \in U(TI), m' \in M, v' \in \mathcal{Q}, k, l \in \mathbb{C}(I, TJ) \ . \\
\quad \Delta^m_I(c,c') \le v \implies \\
\quad\quad (\forall (i,j) \in EI \ . \ (k \bullet i, l \bullet j) \in \tilde{\Delta}(m',v')I) \implies \Delta^{m\cdot m'}_I(k^\sharp \bullet c, l^\sharp \bullet c') \le v + v'
\end{pmatrix} \\
&\iff \begin{pmatrix}
\forall I, J \in \mathbb{C}, m \in M, v \in \mathcal{Q}, c, c' \in U(TI), m' \in M, v' \in \mathcal{Q}, k, l \in \mathbb{C}(I, TJ) \ . \\
\quad \Delta^m_I(c,c') \le v \implies \\
\quad\quad \sup_{(i,j)\in EI} \Delta^{m'}_J(k \bullet i, l \bullet j) \le v' \implies \Delta^{m\cdot m'}_J(k^\sharp \bullet c, l^\sharp \bullet c') \le v + v'
\end{pmatrix} \\
&\iff \begin{pmatrix}
\forall I, J \in \mathbb{C}, m \in M, c, c' \in U(TI), m' \in M, k, l \in \mathbb{C}(I, TJ) \ . \\
\quad \Delta^{m\cdot m'}_I(k^\sharp \bullet c, l^\sharp \bullet c') \le \Delta^m_I(c,c') + \sup_{(i,j)\in EI} \Delta^{m'}_I(k \bullet i, l \bullet j).
\end{pmatrix}.
\end{aligned}
$$

The first two equivalences are obtained by expanding the definitions of $\mathbf{BRel}(\mathbb{C})$, $T^{[\Delta]}$ and $\tilde{\Delta}$, the last two equivalences hold because $\mathcal{Q}$ is a divergence domain. $\square$

Combining the fundamental property and the strength of $T^{[\Delta]}$, we recover a strength law of divergences.

Proposition 13. *Let* $(\mathbb{C}, T)$ *be a CC-SM,* $E : \mathbb{C} \to \mathbf{BRel}(\mathbb{C})$ *be a basic endorelation,* $(M, \le, 1, (\cdot))$ *be a grading monoid and* $\mathcal{Q}$ *be a divergence domain. Suppose also that* $EI \dot{\times} EJ \subseteq E(I \times J)$ *holds for all* $I, J \in \mathbb{C}$*. Then each divergence* $\Delta \in \mathbf{Div}(T, E, M, \mathcal{Q})$ *satisfies: for all* $(x_1, x_2) \in EI$ *and* $c_1, c_2 \in U(TI)$,

$$\Delta^m_{I\times J}(\theta_{I,J} \bullet \langle x_1, c_1\rangle, \theta_{I,J} \bullet \langle x_2, c_2\rangle) \le \Delta^m_J(c_1, c_2).$$

7.1 Simplifying Codensity Liftings by Ω-Generatedness of Divergences

We here show that for an Ω-generated divergence Δ, the calculation of the codensity lifting $T^{[\Delta]}$ can be simplified. For an object $I \in \mathbb{C}$, we define $T^{[\Delta],I}$ by

$$
\begin{aligned}
&(c_1, c_2) \in T^{[\Delta],I}(m,v)X \\
&\quad \iff \forall n, w, (k_1, k_2)\colon X \dot{\to} \tilde{\Delta}(n,w)I \ . \ (k_1^\sharp \bullet c_1, k_2^\sharp \bullet c_2) \in \tilde{\Delta}(m \cdot n, v + w)I.
\end{aligned}
$$

The original calculation of $T^{[\Delta]}$ is a large intersection $T^{[\Delta]} = \bigwedge_{I \in \mathbb{C}} T^{[\Delta],I}$ where I runs over all $\mathbb{C}$-objects, but if Δ is Ω-generated, the parameter I can be fixed at Ω.

Proposition 14. *For any* Ω*-generated divergence* $\Delta \in \mathbf{Div}(T, E, M, \mathcal{Q})$*, we have* $T^{[\Delta]} = T^{[\Delta],\Omega}$.

Proof. We show the equivalence $T^{[\Delta]}X = T^{[\Delta],\Omega}X$ for each $X \in \mathbf{BRel}(\mathbb{C})$.

($\supseteq$) Immediate from $T^{[\Delta]} = \bigwedge_{I \in \mathbb{C}} T^{[\Delta],I}$.

($\subseteq$) By the Ω-generatedness of $\tilde{\Delta}$, we have for all $I \in \mathbb{C}$ and $c'_1, c'_2 \in U(TI)$,

$$(c'_1, c'_2) \in \tilde{\Delta}(m', v')I \iff \forall k\colon I \to T\Omega \ . \ (k^\sharp \bullet c'_1, k^\sharp \bullet c'_2) \in \tilde{\Delta}(m', v')\Omega$$

Therefore, for any $(c_2, c_2) \in U(TX_1) \times U(TX_2)$, we have

$$(c_1, c_2) \in T^{[\Delta],\Omega}X$$

$$
\begin{aligned}
&\iff \forall n \in M, w \in \mathcal{Q}, (k_1, k_2) \colon X \dot{\to} \tilde{\Delta}(n, w)\Omega \;.\; (k_1^\sharp \bullet c_1, k_2^\sharp \bullet c_2) \in \tilde{\Delta}(m \cdot n, v + w)\Omega \\
&\implies \begin{pmatrix} \forall I \in \mathbb{C}, n \in M, w \in \mathcal{Q}, (l_1, l_2) \colon X \dot{\to} \tilde{\Delta}(n, w)I, k \colon I \to T\Omega \;. \\ \qquad\qquad (k^\sharp \circ l_1^\sharp \bullet c_1, k^\sharp \circ l_2^\sharp \bullet c_2) \in \tilde{\Delta}(m \cdot n, v + w)\Omega \end{pmatrix} \\
&\iff \forall I \in \mathbb{C}, n \in M, w \in \mathcal{Q}, (l_1, l_2) \colon X \dot{\to} \tilde{\Delta}(n, w)I \;.\; (l_1^\sharp \bullet c_1, l_2^\sharp \bullet c_2) \in \tilde{\Delta}(m \cdot n, v + w)I \\
&\iff (c_1, c_2) \in T^{[\Delta]}X.
\end{aligned}
$$

This completes the proof. □

For example, the generatedness of DP shown in Section 6.2 implies that $G^{[\mathsf{DP}]} = G^{[\mathsf{DP}],2}$ and $G_s^{[\mathsf{DP}]} = G_s^{[\mathsf{DP}],1}$. In fact, the simplification $G_s^{[\mathsf{DP}],1}$ is equal to the $(\mathcal{R}^+)^2$-graded relational lifting $G_s^{\top\top}$ for DP given in [51, Section 2.2], which is defined by, for each $(X_1, X_2, R_X) \in \mathbf{BRel}(\mathbf{Meas})$,

$$
\begin{aligned}
&G_s^{\top\top}(\varepsilon, \delta)(X_1, X_2, R_X) \\
&\triangleq (G_s(X_1), G_s(X_2), \{(\nu_1, \nu_2) \mid \forall A \in \Sigma_{X_1}, B \in \Sigma_{X_2} \;.\; R_X(A) \subseteq B \implies \nu_1(A) \le \exp(\varepsilon)\nu_2(B) + \delta\}).
\end{aligned}
$$

For detail, see the proof of equalities (†) and (‡) in the proof of [51, Theorem 2.2(iv)].

7.2 Two Lifting Approaches: Codensity and Coupling

We briefly compare two lifting approaches: graded codensity lifting and coupling-based lifting employed in ([12, 13, 8, 9, 52]).

We compare the role of the unit-reflexivity and composability in the codensity graded lifting and the coupling-based graded lifting. Consider the CCC-SM $(\mathbf{Set}, D)$, where D is the probability distribution monad. Given an Eq-relative M-graded $\mathcal{Q}$-divergence Δ on D, the coupling-based graded lifting is defined by

$$
\dot{D}^{\Delta}(m, v)X \triangleq \{(Dp_1 \bullet \mu_1, Dp_2 \bullet \mu_2) \mid (\mu_1, \mu_2) \in (DR_X)^2, \Delta^m_{R_X}(\mu_1, \mu_2) \le v\} \tag{11}
$$

where $p_i \colon R_X \to X_i$ is the projection $(i = 1, 2)$ from the binary relation. The pair (μ_1, μ_2) of probability distributions collected in the right hand side of (11) is called a *coupling*.

The fundamental property $\dot{D}^{\Delta}(\mathrm{Eq}I) = \tilde{\Delta}(m, v)I$ immediately follows from the definition of $\dot{D}^{\Delta}$, while the composability and unit-reflexivity of Δ are used to make $\dot{D}^{\Delta}$ a strong $M \times \mathcal{Q}$-graded lifting [13, Proposition 9]. On the other hand, the codensity graded lifting $D^{[\Delta]}$ is always an $M \times \mathcal{Q}$-graded lifting; this does not rely on the unit-reflexivity and composability of Δ (Proposition 1). These properties are used to show that $D^{[\Delta]}$ satisfies the fundamental property (Proposition 2).

The coupling-based lifting (11) can be naturally generalized to any **Set**-monad T. However, at this moment we do not know how to generalize the coupling technique to any CC-SM $(\mathbb{C}, T)$. As the prior study by [52] pointed out, there is already a difficulty in extending it to the CC-SM $(\mathbf{Meas}, G)$.

We illustrate how the problem arises. Let $X \in \mathbf{BRel}(\mathbf{Meas})$. We would like to pick two probability measures over R_X as couplings, but R_X is merely a set. We therefore equip it with the subspace σ-algebra of $X_1 \times X_2$, and let H_X be the derived measurable space (hence $|H_X| = R_X$). We write $p_i : H_X \to X_i$ for measurable projections $(i = 1, 2)$. We then define a candidate $M \times \mathcal{Q}$-graded lifting of G by

$$
\dot{G}(m, v)X = \{(Gp_1 \bullet \mu_1, Gp_2 \bullet \mu_2) \mid (\mu_1, \mu_2) \in (UGH_X)^2, \Delta^m_{H_X}(\mu_1, \mu_2) \le v\}.
$$

We now verify that $\dot{G}$ also lifts the Kleisli extension of G, that is,

$$
(f, g) \colon Y \dot{\to} \dot{G}(m', v')X \implies (f^\sharp, g^\sharp) \colon \dot{G}(m, v)Y \to \dot{G}(mm', v + v')X.
$$

Let $(f,g)\colon Y \dot{\to} \dot{G}(m',v')X$ be pair of measurable functions. Then for each $(x,y) \in R_Y$, we have $(f \bullet x, g \bullet y) \in R_{\dot{G}(m,v)X}$. Therefore there exists $(\mu_1^{(x,y)}, \mu_2^{(x,y)}) \in (UGH_X)^2$ such that $G\pi_1 \bullet \mu_1^{(x,y)} = f \bullet x$ and $G\pi_2 \bullet \mu_2^{(x,y)} = g \bullet y$. Using *the axiom of choice*, we turn this relationship into *functions* $\mu_1, \mu_2 \colon R_Y \to UGH_X$. If they *were* measurable functions of type $H_Y \to GH_X$, then from the composability of Δ, we would have $\Delta_{H_X}^{mm'}(\mu_1^\sharp \bullet w_1, \mu_2^\sharp \bullet w_2) \le v + v'$ for $w_1, w_2 \in UGH_Y$ such that $\Delta_{H_Y}^{m'}(w_1, w_2) \le v'$. This gives $(f^\sharp, g^\sharp)\colon \dot{G}(m,v)Y \dot{\to} \dot{G}(mm', v+v')X$. However, in general, ensuring the measurability of μ_1, μ_2 is not possible, especially because they are picked up by the axiom of choice. A solution given in [52] is to use the category $\mathbf{Span}(\mathbf{Meas})$ of spans, that guarantees the existence of good measurable functions $h_1, h_2 \colon H_Y \to GH_X$.

8 Approximate Computational Relational Logic

We introduce a program logic called *approximate computational relational logic* (acRL for short). It is a combination of Moggi's computational metalanguage and a relational refinement type system ([9]). The strong graded relational lifting of a monad constructed from a divergence will be used to relationally interpret monadic types, and gradings give upper bounds of divergences between computational effects caused by two programs. acRL is similar to the relational refinement type system HOARe2 ([9]), which is designed for verifying differential privacy of probabilistic programs. Compared to HOARe2, acRL supports general monads and divergences, while it does not support dependent products nor non-termination.

The relational logic acRL adopts the *extensional approach* (cf. [44, Chapter 9.2]):

- Relational assertions between contexts Γ and Δ are defined as binary relations between $U[\![\Gamma]\!]$ and $U[\![\Delta]\!]$, or equivalently $\mathbf{BRel}(\mathbb{C})$-objects ϕ such that $p_{\mathbb{C}}(\phi) = ([\![\Gamma]\!], [\![\Delta]\!])$. Logical connectives and quantifications are defined as operations on such $\mathbf{BRel}(\mathbb{C})$-objects. This is in contrast to the standard design of logic where assertions are defined by a BNF.
- Let $\Gamma \vdash M : \tau$ and $\Delta \vdash N : \sigma$ be well-typed terms, ϕ be a relational assertion between Γ, Δ, and ψ be an assertion between τ, σ. The main concern of acRL is the statement "$\forall(\gamma,\delta) \in \phi.([\![M]\!] \bullet \gamma, [\![N]\!] \bullet \delta) \in \psi$" (equivalently $([\![M]\!], [\![N]\!]) : \phi \dot{\to} \psi$). In this section we denote this statement by $\phi \vdash (M, M') : \psi$.
- Inference rules of the logic consists of the *facts* about the statement $\phi \vdash (M, M') : \psi$. We remark that in the standard logic, proving these facts corresponds to the soundness of inference rules.

8.1 Moggi's Computational Metalanguage

8.1.1 Syntax of the Computational Metalanguage

For the higher-order programming language, we adopt Moggi's *computational metalanguage* ([42]). It is an extension of the simply typed lambda calculus with monadic types. For a set B, we define the set $\mathbf{Typ}(B)$ of types over B by the first BNF in Figure 2. We then define the set $\mathbf{Typ_1}(B)$ of first-order types to be the subset of $\mathbf{Typ}(B)$ consisting only of $b, 1, \times, +$.

We next introduce *computational signatures* for specifying constants in the computational metalanguage. A computational signature is a tuple (B, Σ_v, Σ_e) where B is a set of base types, and Σ_v and Σ_e are functions whose range is $\mathbf{Typ_1}(B)^2$. The domains of Σ_v, Σ_e are sets of *value operation* symbols and *effectful operation* symbols, and are denoted by O_v, O_e, respectively. These functions assign input and output types to these operations.

Figure 2: Syntax of Types and Raw Terms of the Computational Metalanguage

$$\mathbf{Typ}(B) \ni \tau ::= b \mid 1 \mid \tau \times \tau \mid 0 \mid \tau + \tau \mid \tau \Rightarrow \tau \mid \mathtt{T}\tau \quad (b \in B)$$

$$\begin{aligned} M ::= {} & x \mid o(M) \mid c(M) \mid () \mid (M, M) \mid \pi_1(M) \mid \pi_2(M) \quad (o \in O_v, c \in O_e) \\ & \mid \iota_1(M) \mid \iota_2(M) \mid M \,\mathtt{with}\, \iota_1(x : \tau).M \shortmid \iota_2(x : \tau).M \\ & \mid (\lambda x : \tau \,.\, M) \mid (MM) \mid \mathtt{ret}(M) \mid \mathtt{let}\, x : \tau = M \,\mathtt{in}\, M \end{aligned}$$

Figure 3: Data for the Categorical Semantics of Metalanguage

1. $(\mathbb{C}, T)$ is a CCC-SM and $\mathbb{C}$ has finite coproducts.
2. $[\![b]\!] \in \mathbb{C}$ for each $b \in B$
3. $[\![o]\!] : [\![b]\!] \to [\![b']\!]$ for each $o \in O_v$ such that $\Sigma_v(o) = (b, b')$
4. $[\![c]\!] : [\![b]\!] \to T[\![b']\!]$ for each $c \in O_e$ such that $\Sigma_e(c) = (b, b')$

Fix a countably infinite set V of variables. A context is a function from a finite subset of V to $\mathbf{Typ}(B)$; contexts are often denoted by capital Greek letters Γ, Δ. For contexts Γ, Δ such that $\mathrm{dom}(\Gamma) \cap \mathrm{dom}(\Delta) = \emptyset$, by Γ, Δ we mean the join of Γ and Δ.

The set of raw terms is defined by the second BNF in Figure 2. The type system of the computational metalanguage has judgments of the form $\Gamma \vdash M : \tau$ where Γ is a context, M a raw term and τ a type. It adopts the standard rules for products, coproducts, implications and monadic types; see e.g. [42]. The typing rules for value operations and effectful operations are given by

$$\frac{o \in O_v \quad \Sigma_v(o) = (b, b') \quad \Gamma \vdash M : b}{\Gamma \vdash o(M) : b'} \qquad \frac{o \in O_e \quad \Sigma_e(c) = (b, b') \quad \Gamma \vdash M : b}{\Gamma \vdash c(M) : \mathtt{T}b'}$$

A *simultaneous substitution* from Γ to Γ' is a function θ from the set $\mathrm{dom}(\Gamma')$ of variables to raw terms such that the well-typedness $\Gamma \vdash \theta(x) : \Gamma'(x)$ holds for each $x \in \mathrm{dom}(\Gamma')$. The application of θ to a term $\Gamma' \vdash M : \tau$ is denoted by $M\theta$, which has a typing $\Gamma \vdash M\theta : \tau$. For disjoint contexts Γ_i $(i = 1, 2)$, we define the projection substitutions $\Gamma_1, \Gamma_2 \vdash \pi_i^{\Gamma_1, \Gamma_2} : \Gamma_i$ by $\pi_i^{\Gamma_1, \Gamma_2}(x) = x$.

8.1.2 Categorical Semantics of the Computational Metalanguage

The interpretation of the computational metalanguage over a computational signature (B, Σ_v, Σ_e) is given by the data specified by Figure 3.

We first inductively extend the interpretation of base types to all types using the bi-Cartesian closed structure and the monad. Next, for each context Γ, we fix a product diagram $([\![\Gamma]\!], \{\pi_x : [\![\Gamma]\!] \to [\![\Gamma(x)]\!]\}_{x \in \mathrm{dom}(\Gamma)})$; when $\mathrm{dom}(\Gamma) = \{x\}$, we assume that $[\![\Gamma]\!] = [\![\Gamma(x)]\!]$ with $\pi_x = \mathrm{id}$. Lastly we interpret a typing derivation of $\Gamma \vdash M : \tau$ as a morphism $[\![M]\!] : [\![\Gamma]\!] \to [\![\tau]\!]$ in the standard way, using the interpretations of operations given in Figure 3. We further extend this to the interpretation of each simultaneous substitution $\Gamma \vdash \theta : \Gamma'$ as a morphisms $[\![\theta]\!] : [\![\Gamma]\!] \to [\![\Gamma']\!]$.

8.2 Approximate Relational Computational Logic

8.2.1 Relational Logic in External Form

A *relational assertion* ϕ between disjoint contexts Γ and Δ is a binary relation between $U[\![\Gamma]\!]$ and $U[\![\Delta]\!]$. We denote such a relational assertion by ${}^{\Gamma}_{\Delta} \vdash \phi$, and identify it as a $\mathbf{BRel}(\mathbb{C})$-object ϕ such that $p_{\mathbb{C}}(\phi) = ([\![\Gamma]\!], [\![\Delta]\!])$. Similarly, a relational assertion between types τ and σ is defined to be a relational assertion ${}^{u:\tau}_{d:\sigma} \vdash \phi$; here u, d are reserved and fixed variables.

Relational assertions between contexts Γ and Δ carry a *boolean algebra structure* $\wedge, \vee, \neg$ given by the set-intersection, set-union and set-complement (see the boolean algebra $\mathbf{BRel}(\mathbb{C})_{([\![\Gamma]\!],[\![\Delta]\!])}$ in Section 2.1). The pseudo-complement $\phi \Rightarrow \psi$ is defined to be $\neg\phi \vee \psi$. For ${}^{\Gamma,x:\tau}_{\Delta,y:\sigma} \vdash \phi$, by ${}^{\Gamma}_{\Delta} \vdash \forall^{x}_{y} \,.\, \phi$ and ${}^{\Gamma}_{\Delta} \vdash \exists^{x}_{y} \,.\, \phi$ we mean the relational assertions defined by the following equivalence:

$$
\begin{aligned}
(\gamma, \delta) \in \forall^{x}_{y} \,.\, \phi \iff{}& \forall \gamma' \in U[\![\Gamma, x : \tau]\!], \delta' \in U[\![\Delta, y : \sigma]\!] \,. \\
&\qquad ([\![\pi_1^{\Gamma,x:\tau}]\!] \bullet \gamma' = \gamma) \wedge ([\![\pi_1^{\Delta,y:\sigma}]\!] \bullet \delta' = \delta) \Rightarrow (\gamma', \delta') \in \phi \\
(\gamma, \delta) \in \exists^{x}_{y} \,.\, \phi \iff{}& \exists \gamma' \in U[\![\Gamma, x : \tau]\!], \delta' \in U[\![\Delta, y : \sigma]\!] \,. \\
&\qquad ([\![\pi_1^{\Gamma,x:\tau}]\!] \bullet \gamma' = \gamma) \wedge ([\![\pi_1^{\Delta,y:\sigma}]\!] \bullet \delta' = \delta) \wedge (\gamma', \delta') \in \phi
\end{aligned}
$$

The boolean algebra structure and the above quantifier operations allow us to interpret first-order logical formulas as relational assertions; we omit its detail here. In addition to these standard logical connectives, we will use graded strong relational lifting $T^{[\Delta]}$ to form relational assertions. That is, for any basic endorelation $E : \mathbb{C} \to \mathbf{BRel}(\mathbb{C})$, grading monoid M, divergence domain $\mathcal{Q}$ and divergence $\Delta \in \mathbf{Div}(T, E, M, \mathcal{Q})$, we obtain a relational assertion ${}^{u:\mathtt{T}\tau}_{d:\mathtt{T}\sigma} \vdash T^{[\Delta]}(m, v)\phi$ from any ${}^{u:\tau}_{d:\sigma} \vdash \phi$, $m \in M$ and $v \in \mathcal{Q}$.

For substitutions $\Gamma \vdash \theta : \Gamma', \Delta \vdash \theta' : \Delta'$ and an assertion ${}^{\Gamma}_{\Delta} \vdash \phi$, by ${}^{\Gamma'}_{\Delta'} \vdash \phi[\theta; \theta']$ we mean the relational assertion $\{(\gamma, \delta) \mid ([\![\theta]\!] \bullet \gamma, [\![\theta']\!] \bullet \delta) \in \phi\}$. For disjoint context pairs Γ, Γ' and Δ, Δ' and relational assertions ${}^{\Gamma}_{\Delta} \vdash \phi$ and ${}^{\Gamma'}_{\Delta'} \vdash \psi$, by the juxtaposition ${}^{\Gamma,\Gamma'}_{\Delta,\Delta'} \vdash \phi, \psi$ we mean the relational assertion ${}^{\Gamma,\Gamma'}_{\Delta,\Delta'} \vdash \phi[\pi_1^{\Gamma,\Gamma'}; \pi_1^{\Delta,\Delta'}] \wedge \psi[\pi_2^{\Gamma,\Gamma'}; \pi_2^{\Delta,\Delta'}]$.

8.2.2 Inference Rules for acRL

For well-typed computational metalanguage terms $\Gamma \vdash M : \tau$ and $\Delta \vdash N : \sigma$, and relational assertions ${}^{\Gamma}_{\Delta} \vdash \phi$ and ${}^{u:\tau}_{d:\sigma} \vdash \psi$, by the judgment

$$\phi \vdash (M, N) \colon \psi$$

we mean the inclusion $\phi \subseteq \psi[[M/u]; [N/d]]$ of binary relations. This is equivalent to $([\![M]\!], [\![N]\!]) : \phi \dot{\to} \psi$. We show basic facts about judgments $\phi \vdash (M, N) \colon \psi$.

Proposition 15. *1.* $\phi \vdash (M, N) \colon \psi$ *and* $[\![M]\!] = [\![M']\!]$ *and* $[\![N]\!] = [\![N']\!]$ *implies* $\phi \vdash (M', N') \colon \psi$.

2. $\phi \vdash (M, N) \colon \psi$ *and* $\phi' \subseteq \phi$ *and* $\psi \subseteq \psi'$ *implies* $\phi' \vdash (M, N) \colon \psi'$.

3. $\phi \vdash (M, N) \colon T^{[\Delta]}(m, v)\psi$ *and* $m \leq n$ *and* $v \leq w$ *and* $\psi \leq \psi'$ *implies* $\phi \vdash (M, N) \colon T^{[\Delta]}(n, w)\psi'$.

4. $\phi \vdash (M, N) \colon \psi$ *implies* $\phi \vdash (\mathtt{ret}(M), \mathtt{ret}(N)) \colon T^{[\Delta]}(1, 0)\psi$.

5. $\phi \vdash (M, N) \colon T^{[\Delta]}(m, v)\psi$ *and* $\phi, \psi[[x/u]; [x'/d]] \vdash (M', N') \colon T^{[\Delta]}(n, w)\rho$ *implies* $\phi \vdash (\mathtt{let}\, x = M \,\mathtt{in}\, M', \mathtt{let}\, x' = N \,\mathtt{in}\, N') \colon T^{[\Delta]}(m \cdot n, v \cdot w)\rho$.

We next establish relational judgments on effectful operations. We present a convenient way to establish such judgments using the fundamental property of the graded relational lifting $T^{[\Delta]}$.

Proposition 16. *For any* $c \in O_e$ *such that* $\Sigma_e(c) = (b, b')$*, relational assertion* ${}^{u:b}_{d:b} \vdash \phi$ *and* $m \in M$*, putting* $v = \sup\{\Delta^m_{[\![b']\!]}([\![c]\!]\bullet x, [\![c]\!]\bullet y) \mid (x, y) \in \phi\}$*, we have* $\phi \vdash (c(u), c(d)) \colon T^{[\Delta]}(m, v)(E[\![b']\!])$.

Proof. Take an arbitrary pair $(x, y) \in \phi$. We have $\Delta^m_{[\![b']\!]}([\![c]\!] \bullet x, [\![c]\!] \bullet y) \le v$ by definition of v. Thanks to the fundamental property of $T^{[\Delta]}$ (Theorem 8), it is equivalent to $([\![c]\!] \bullet x, [\![c]\!] \bullet y) \in T^{[\Delta]}(m, v)(E[\![b']\!])$. ☐

9 Case Study I: Higher-Order Probabilistic Programs

We represent a higher-order probabilistic programming language with sampling commands from continuous distributions as a computational metalanguage. For now we assume that the language supports sampling from Gaussian distribution and Laplace distribution. This computational metalanguage is specified by the computational signature:

$$\mathcal{C} = (\{\mathtt{R}\}, \Sigma_v, \{\mathtt{norm} : (\mathtt{R} \times \mathtt{R}, \mathtt{R}),\ \mathtt{lap} : (\mathtt{R} \times \mathtt{R}, \mathtt{R})\}),$$

where Σ_v is some chosen signature for value operations over reals. We interpret this computational metalanguage by filling Figure 3 as follows:

1. for the CCC-SM, we take $(\mathbb{C}, T) = (\mathbf{QBS}, P)$ (see Section 13),
2. for the interpretation $[\![\mathtt{R}]\!]$ of $\mathtt{R}$, we take the quasi-Borel space $K\mathbb{R}$ associated with the standard Borel space $\mathbb{R}$,
3. the interpretation of value operations is given as expected (we omit it here); for example when Σ_v contains the real number addition operator $+$ as type $(\mathtt{R} \times \mathtt{R}, \mathtt{R})$, its interpretation is the QBS morphism $[\![+]\!](x, y) = x + y : [\![\mathtt{R} \times \mathtt{R}]\!] \to [\![\mathtt{R}]\!]$,
4. for the interpretation of effectful operations, we put

$$[\![\mathtt{norm}]\!](x, \sigma) = [\mathrm{id}, \mathcal{N}(x, \sigma^2)]_{\sim_{K\mathbb{R}}}, \qquad [\![\mathtt{lap}]\!](x, \lambda) = [\mathrm{id}, \mathrm{Lap}(x, \lambda)]_{\sim_{K\mathbb{R}}}.$$

Here, $\mathcal{N}(x, \sigma^2) \in G\mathbb{R}$ is the Gaussian distribution with mean x and variance σ^2. $\mathrm{Lap}(x, \lambda) \in G\mathbb{R}$ is the Laplacian distribution with mean x and variance $2\lambda^2$ [7]. Every probability (Borel-)measure $\mu \in G\mathbb{R}$ on $\mathbb{R}$ can be converted to the probability measure $[\mathrm{id}, \mu]_{\sim_{K\mathbb{R}}} \in PK\mathbb{R}$ on the quasi-Borel space $K\mathbb{R}$ (see Section 5.5).

9.1 A Relational Logic Verifying Differential Privacy

To formulate differential privacy and its relaxations in the quasi-Borel setting, we convert statistical divergences Δ on the Giry monad G in Table 2 to Eq-relative divergences $\langle L, l\rangle^* \Delta$ on the probability monad P on $\mathbf{QBS}$ by the construction in Section 5.5. Then, we construct the graded relational lifting $P^{[\langle L, l\rangle^* \Delta]}$ by Theorem 8. Using this, as an instantiation of acRL, we build a

[7] If $\sigma = 0$ (or $\lambda \le 0$), $\mathcal{N}(x, \sigma^2)$ (resp. $\mathrm{Lap}(x, \lambda)$) is not defined, thus we replace it by the Dirac distribution $\mathbf{d}_x$ at x instead.

relational logic reasoning about differential privacy and its relaxations, supporting *higher-order programs* and continuous random samplings. Basic proof rules can be given by Proposition 15.

For effectful operations, we import basic proof rules on noise-adding mechanisms given in prior studies ([21, 22, 40, 16]) via Theorem 1 and Proposition 16. For example, consider the Eq-relative $\mathcal{R}^+$-graded $\mathcal{R}^+$-divergence $\Delta = \langle L, l\rangle^* \mathsf{DP}$ on P. Proposition 16 with an effectful operation $c = \mathtt{lap}$ and a relational assertion (below we identify global elements in $K\mathbb{R}$ and real numbers)

$$\substack{u:\mathtt{R}\times\mathtt{R}\\ d:\mathtt{R}\times\mathtt{R}} \vdash \phi = \{(\langle x, 1/\varepsilon\rangle, \langle y, 1/\varepsilon\rangle) \mid |x-y| \leq 1\},$$

together with Theorem 1 and the prior result [21, Example 1] yields the following judgment:

$$\phi \vdash (\mathtt{lap}(u), \mathtt{lap}(d))\colon P^{[\langle L,l\rangle^* \mathsf{DP}]}(0,\epsilon)(\mathrm{Eq}K\mathbb{R}).$$

By letting diff_r be the relational assertion $\substack{u:\mathtt{R}\\ d:\mathtt{R}} \vdash \{(x,y) \mid |x-y| \leq r\}$, the above judgment is equivalent to:

$$\mathrm{diff}_1 \vdash (\mathtt{lap}(u, 1/\epsilon), \mathtt{lap}(d, 1/\epsilon)))\colon P^{[\langle L,l\rangle^* \mathsf{DP}]}(0,\epsilon)(\mathrm{Eq}K\mathbb{R}). \tag{12}$$

This rule corresponds to the rule [LapGen] of the program logic $\mathsf{apRHL+}$ ([11]) for differential privacy. For another example, by the reflexivity of DP, $\langle L, l\rangle^* \mathsf{DP}$ is also reflexive, hence we obtain the following judgments (below succ_r is the relational assertion $\substack{u:\mathtt{R}\\ d:\mathtt{R}} \vdash \{(x,y) \mid y = x+r\}$):

$$\mathrm{succ}_1 \vdash (\mathtt{lap}(u, \lambda), \mathtt{lap}(d, \lambda))\colon P^{[\langle L,l\rangle^* \mathsf{DP}]}(0,0)(\mathrm{succ}_1) \tag{13}$$

$$\mathrm{succ}_1 \vdash (\mathtt{norm}(u, \sigma), \mathtt{norm}(d, \sigma))\colon P^{[\langle L,l\rangle^* \mathsf{DP}]}(0,0)(\mathrm{succ}_1). \tag{14}$$

The judgment (13) correspond to [LapNull] of $\mathsf{apRHL+}$. Similarly, the following judgments about the DP, Rényi-DP, zero-concentrated DP of the Gaussian mechanism can be derived as (15)–(17).

$$\mathrm{diff}_1 \vdash (\mathtt{norm}(u, \sigma), \mathtt{norm}(d, \sigma))\colon P^{[\langle L,l\rangle^* \mathsf{DP}]}(\epsilon,\delta)(\mathrm{Eq}K\mathbb{R}) \tag{15}$$

$$\mathrm{diff}_r \vdash (\mathtt{norm}(u, \sigma), \mathtt{norm}(d, \sigma))\colon P^{[\langle L,l\rangle^{*\alpha} \mathsf{Re}]}(\alpha r^2/2\sigma^2)(\mathrm{Eq}K\mathbb{R}) \tag{16}$$

$$\mathrm{diff}_r \vdash (\mathtt{norm}(u, \sigma), \mathtt{norm}(d, \sigma))\colon P^{[\langle L,l\rangle^* \mathsf{zCDP}]}(0, r^2/2\sigma^2)(\mathrm{Eq}K\mathbb{R}) \tag{17}$$

In (15) we require $\sigma \geq \max((1+\sqrt{3})/2,\ \sqrt{2\log(0.66/\delta)}/\epsilon)$. The derivation is done via Proposition 16, Theorem 1 and prior studies ([51, 40, 17]).

10 Case Study II: Probabilistic Programs with Costs

We further extend the computational signature $\mathcal{C}$ in the previous section with an effectful operation $\mathtt{tick}$ such that $\Sigma_e(\mathtt{tick}) = (\mathtt{R}, 1)$. The intention of $\mathtt{tick}(r)$ is to increase cost counter by r during execution[8]. To interpret this extended metalanguage, we fill Figure 3 as follows:

1. for the CCC-SM, we take $(\mathbb{C}, T) = (\mathbf{QBS}, P_c)$ where $P_c \triangleq P(K\mathbb{R} \times -)$ is the monad for modeling probabilistic choice and cost counting (see Section 5.7).
2. interpretation of $b \in B$ is the same as Section 9,
3. interpretation of value operations is also the same as Section 9,

[8]To make examples simpler, we allow negative costs.

4. for the interpretation of effectful operations, put

$$\begin{aligned}
[\![\texttt{norm}]\!](x,\sigma) &= [(0,\mathrm{id}), \mathcal{N}(x,\sigma^2)]_{\sim_{K\mathbb{R}\times K\mathbb{R}}},\\
[\![\texttt{lap}]\!](x,\lambda) &= [(0,\mathrm{id}), \mathrm{Lap}(x,\lambda)]_{\sim_{K\mathbb{R}\times K\mathbb{R}}},\\
[\![\texttt{tick}]\!](r) &= \eta^P_{K\mathbb{R}\times[\![1]\!]}(r,*) = [\mathrm{const}(r,*),\mu]_{\sim_{K\mathbb{R}\times 1}}.
\end{aligned}$$

We derive a closed term $\texttt{ntick}\colon \texttt{R} \Rightarrow \texttt{R} \Rightarrow \texttt{T1}$ for ticking with a cost sampled from Gaussian distribution:

$$\texttt{ntick} \triangleq (\lambda s.\lambda r.\ \texttt{let}\, x = \texttt{norm}(r,s)\,\texttt{in}\,\texttt{tick}(x)).$$

The term $\texttt{ntick}\ s\ r$ adds cost counter by a random value sampled from the Gaussian distribution $\texttt{norm}(r,s^2)$.

10.1 Relational Reasoning on Probabilistic Costs

We convert the total valuation distance $\mathsf{TV} \in \mathbf{Div}(G,\mathrm{Eq},1,\mathcal{R}^+)$ to the divergence $\Delta_c \triangleq \mathsf{C}(\langle L,l\rangle^*\mathsf{TV}, K\mathbb{R}) \in \mathbf{Div}(P_c,\mathrm{Eq},1,\mathcal{R}^+)$ on P_c by Propositions 2 and 7. We also prove basic facts on effectful operations. First, the following relational judgments on $\texttt{tick}$ can be easily given:

$$\begin{aligned}
\top &\vdash (\texttt{tick}(u),\texttt{tick}(d))\colon T^{[\Delta_c]}(1)(\top) \qquad (18)\\
u = d &\vdash (\texttt{tick}(u),\texttt{tick}(d))\colon T^{[\Delta_c]}(0)(\top)
\end{aligned}$$

Remark that $\mathrm{Eq}1 = \top$ and $[\![\texttt{tick(0)}]\!] = [\![\texttt{ret}(*)]\!]$ holds. Next, in the similar way as (13), by the reflexivity of TV, we have the reflexivity of $\langle L,l\rangle^*\mathsf{TV}$, and we obtain, for each real number constant σ,λ,

$$\begin{aligned}
\mathrm{succ}_r &\vdash (\texttt{norm}(u,\sigma),\texttt{norm}(d,\sigma))\colon T^{[\Delta_c]}(0)(\mathrm{succ}_r)\\
\mathrm{succ}_r &\vdash (\texttt{lap}(u,\lambda),\texttt{lap}(d,\lambda))\colon T^{[\Delta_c]}(0)(\mathrm{succ}_r) \qquad (19)
\end{aligned}$$

We also directly verify the following judgment on $\texttt{ntick}$ using Theorem 1 and Proposition 16:

$$\mathrm{diff}_1 \vdash (\texttt{ntick}\ \sigma\ u, \texttt{ntick}\ \sigma\ d)\colon T^{[\Delta_c]}(\mathrm{Pr}_{r\sim\mathcal{N}(0,\sigma^2)}[|r|<0.5])(\top). \qquad (20)$$

10.1.1 An Example of Relational Reasoning

We give examples of verification of difference (of distributions) of costs between two runs of a probabilistic program whose output and cost depend on the input. We consider the following program:

$$M \triangleq \lambda r\colon \texttt{R}.\ \lambda t\colon \texttt{R} \to T1.\ \texttt{let}\, x = \texttt{lap}(r,5)\,\texttt{in}\,\texttt{let}\,\_ = t(r)\,\texttt{in}\,\texttt{ret}(x-r).$$

It first samples a real number x from the Laplacian distribution centered at the input r, call the (possibly effectful) closure t with r and return $x-r$. Since the return type of t is $T1$, it can only probabilistically tick the counter. We show that the following two judgments in acRL:

$$\begin{aligned}
&\vdash (M\ 0\ (\lambda x.\texttt{tick}(x)), M\ 1\ (\lambda x.\texttt{tick}(x)))\colon T^{[\Delta_c]}(1)(\mathrm{Eq}[\![\texttt{R}]\!]), \qquad \text{(A)}\\
&\vdash (M\ 0\ (\texttt{ntick}(2)), M\ 1\ (\texttt{ntick}(2)))\colon T^{[\Delta_c]}(0.20)(\mathrm{Eq}[\![\texttt{R}]\!]) \qquad \text{(B)}
\end{aligned}$$

In judgment (A), we pass the tick operation $t = \lambda x.\texttt{tick}(x)$ itself to $M\ 0$ and $M\ 1$. By the fundamental property of $T^{[\Delta_c]}$, the difference of costs between two runs of $M\ 0\ t$ and $M\ 1\ t$

is 1, because each of these programs reports cost 0 and 1 deterministically. In contrast, in judgment (B), we pass to M 0 and M 1 the probabilistic tick function $t' = \mathtt{ntick}(2)$ that ticks a real number sampled from the Gaussian distribution with variance $2^2 = 4$. Therefore the cost reported by the runs of programs $M\ 0\ t'$ and $M\ 1\ t'$ follow the Gaussian distributions $\mathcal{N}(0, 4)$ and $\mathcal{N}(1, 4)$, whose difference by TV is bounded by 0.20.

We first show (A). By (18) and 2 of Proposition 15, we have,

$$\mathrm{succ}_1 \vdash (\mathtt{tick}(u), \mathtt{tick}(d)) \colon T^{[\mathsf{\Delta}_c]}(1)(\top). \tag{21}$$

By (21), and 4, 5 of Proposition 15, we obtain,

$$\begin{aligned}\mathrm{succ}_1 \vdash (&\mathtt{let}\,\_ = \mathtt{tick}(u)\,\mathtt{in}\,\mathtt{ret}(u),\\ &\qquad \mathtt{let}\,\_ = \mathtt{tick}(d)\,\mathtt{in}\,\mathtt{ret}(d-1)) \colon T^{[\mathsf{\Delta}_c]}(1)(\mathrm{Eq}[\![\mathrm{R}]\!]).\end{aligned} \tag{22}$$

By (19), (22), and 1 and 5 of Proposition 15 again, we conclude (A).

To show (B), it suffices to replace (21) by the following judgment proved by (20), the inequality $\Pr_{r \sim \mathcal{N}(0,4)}[|r| < 0.5] \leq 0.20$ and 2 of Proposition 15:

$$\mathrm{succ}_1 \vdash (\mathtt{ntick}\ 2\ u, \mathtt{ntick}\ 2\ d) \colon T^{[\mathsf{\Delta}_c]}(0.20)(\top).$$

The rest of proof is the same as (A).

11 Related Work

This work is based on the frameworks for verifying the differential privacy of probabilistic programs using relational logic, summarized in Table 5. Composable divergences employed in these frameworks include the one for differential privacy, plus its recent relaxations, such as, Rényi DP, zero-concentrated DP, and truncated-concentrated DP ([16, 17, 40]).

Table 5: Approximate Probabilistic Relational Logic

Work	Monad	Relation	Lifting Method	Supported divergences
[8, 10, 12]	Dist	$\mathbf{BRel}(\mathbf{Set})$	coupling	DP
[13]	Dist	$\mathbf{BRel}(\mathbf{Set})$	coupling	f-divergences
[52]	Giry	$\mathbf{Span}(\mathbf{Meas})$	coupling (spans)	composable ones
[51]	Giry	$\mathbf{BRel}(\mathbf{Meas})$	codensity	DP
This work	Generic	$\mathbf{BRel}(\mathbb{C})$	codensity	composable ones

The key semantic structure in these frameworks is *graded relational liftings* of the probability distribution monad. Barthe et al. gave a graded relational lifting of the distribution monad based on the existence of two witnessing probability distributions (called *coupling*) ([12]). Since then, *coupling-based* liftings have been refined and used in several works ([8, 10, 13, 52]). They can be systematically constructed from *composable* divergences on the probability distribution monad ([13]). One advantage of coupling-based liftings is that, to relate two probability distributions, it suffices to exhibit a coupling; this is exploited in the mechanized verification of differential privacy of programs ([1, 2]). These coupling-based liftings, however, are developed upon discrete probability distributions, and measure-theoretic probability distributions, such as Gaussian or Cauchy distributions, were not supported until the work ([52]).

The relational Hoare logic supporting sampling from continuous probability measures is given in the study by [51]. In his work, the graded relational lifting for (ϵ, δ)-DP is given in the style of *codensity lifting* ([33]), which does not rely on the existence of coupling. Yet, it has been an open question [52, Section VIII] how to extend his graded relational lifting to support various relaxations of differential privacy. This paper answers to this question as Theorem 8. Later, coupling-based liftings has also been extended to support samplings from continuous probability measures ([52]). This extension is achieved by redefining the concept of binary relations as *spans* of measurable functions. Comparison of these approaches is in the next section.

The verification of differential privacy in functional programming languages has also been pursued ([48, 23, 9, 5]). [48] introduced a linear functional programming language with a graded monadic type that supports reasoning about ϵ-differential privacy. Later, Gaboardi et al. strengthen Reed-Pierce type system with dependent types ([23]). A category-theoretic account of Reed and Pierce type system is given in [5], where general (ϵ, δ)-differential privacy is also supported. These works basically regard types as metric spaces, allowing us to reason about *sensitivity* of programs with respect to inputs. The coupling-based lifting techniques are also employed in the relational models of higher-order probabilistic programming language ([9]).

The study [5] gives a categorical definition of composable divergences in a general framework called *weakly closed refinements of symmetric monoidal closed categories* [5, Definition 1]. A comparison is given in Section 6.1.2.

[39] introduced a quantitative refinement of algebraic theory called *quantitative equational theory*, and studied variety theorem for quantitative algebras. [6] discussed tensor products of quantitative equational theories. QETs and divergences on monads share the common interest of measuring quantitative differences between computational effects. Divergences on monads are derived as a generalization of the composability condition of statistical divergences studied by [13]. To make a precise connection between these two concepts, in Section 6.3, we have given an adjunction between QETs of type Ω over X and X-generated divergences on the free monad T_Ω. The adjunction cuts down to the isomorphism between *unconditional* QETs of type Ω over X and X-generated divergences on T_Ω.

The use of metric-like spaces in the semantics is seen in several recent work. [25] studies quantitative refinements of Abramsky's applicative bisimilarity for Reed-Pierce type system. He introduces a monadic operational semantics of the language and formalized quantitative applicative bisimilarity using monad liftings to the category of quantale-valued relations. [15] also used metric-like spaces to study bisimulations and up-to techniques in the category of quantale-valued relations. In this work our interest is relational program verification of effectful programs, and it is carried out in the relational category $\mathbf{BRel}(\mathbb{C})$, rather than $\mathbf{Div}_{\mathcal{Q}}(\mathbb{C})$. The quantitative difference of computational effects measured by a divergence Δ is represented by the binary relation $\tilde{\Delta}$ graded by upper bounds of distance.

12 Future Work

The framework for relational cost analysis given in ([47])(extension of $\mathsf{RelCost}$ ([18])) consists of the relational logic verifying the difference of costs between two programs and the unary logic verifying the lower and upper bound of costs (i.e. cost intervals) in one program. We expect that the relational logic can be reformulated by an instantiation of acRL with the divergence NCI on $P(\mathbb{N} \times -)$ (or its variant). However to reformulate the unary logic, we want a *unary version of divergence* on $P(\mathbb{N} \times -)$ for cost intervals. To establish the connection between the unary logic and relational logic, we want a conversion from the unary version of divergence (for cost intervals) to NCI (for cost difference).

There might be many other examples and applications of divergences on monads. In this paper, we mainly discussed examples of divergences with basic endorelations Top and Eq, but various other basic endorelations can be considered.

13 Measurable Spaces and Quasi-Borel Spaces

Measurable Spaces. For the treatment of continuous probability distributions, we employ the category **Meas** of measurable spaces and measurable functions. For a measurable space I we write $|I|$ and Σ_I for the underlying set and σ-algebra of I respectively. The category **Meas** is a (well-pointed) CC, and it has all small limits and small colimits that are strictly preserved by the forgetful functor $|-|\colon \mathbf{Meas} \to \mathbf{Set}$. It is naturally isomorphic to the global element functor $\mathbf{Meas}(1, -)$.

Standard Borel Spaces. A standard Borel space is a special measurable space $(|\Omega|, \Sigma_\Omega)$ whose σ-algebra Σ_Ω is the coarsest one containing the topology σ_Ω of a Polish space $(|\Omega|, \sigma_\Omega)$. In particular, the real line $\mathbb{R}$ forms a standard Borel space. In fact, a measurable space Ω is standard Borel if and only if there are $\gamma\colon \Omega \to \mathbb{R}$ and $\gamma'\colon \mathbb{R} \to \Omega$ in **Meas** forming a section-retraction pair, that is, $\gamma' \circ \gamma = \mathrm{id}_\Omega$. For example, $[0,1]$, $[0,\infty]$, $\mathbb{N}$, $\mathbb{R}^k$ ($k \in \mathbb{N}$) are standard Borel.

The Giry Monad. We recall the Giry monad G ([26]). For every measurable space I, GI is the set $|GI|$ of all probability measures over I with the coarsest σ-algebra induced by functions $\mathrm{ev}_A\colon |GI| \to [0,1]$ ($A \in \Sigma_X$) defined by $\mathrm{ev}_A(\mu) = \mu(A)$. The unit $\eta_I\colon I \to GI$ assigns to each $x \in I$ the Dirac distribution $\mathbf{d}_x$ centered at x. For every $f\colon I \to GJ$, the Kleisli extension $f^\sharp\colon GI \to GJ$ is given by $(f^\sharp(\mu))(A) = \int_x f(x)(A)\ d\mu(x)$ for each $\mu \in GI$. We also denote by G_s the subprobabilistic variant of G (called sub-Giry monad), where the underlying set $|G_s I|$ of $G_s I$ is relaxed to the set of subprobaility measures over I.

The Giry monad G (resp. the sub-Giry monad G_s) carries a (commutative) strength $\theta_{I,J}\colon I \times GJ \to G(I \times J)$ over the CC $(\mathbf{Meas}, 1, (\times))$. It computes the product of measures $((x,\mu) \mapsto \mathbf{d}_x \otimes \mu)$. Therefore $(\mathbf{Meas}, G)$ and $(\mathbf{Meas}, G_s)$ are (well-pointed) CC-SMs.

Quasi-Borel Spaces. The category **Meas** is not suitable for the semantics of *higher-order* programming languages since it is not Cartesian closed ([4]). For the treatment of higher-order probabilistic programs with continuous distributions, we employ the Cartesian closed category **QBS** of quasi-Borel spaces and morphisms between them, together with the probability monad P on **QBS** ([28]). A quasi-Borel space is a pair $I = (|I|, M_I)$ of a set $|I|$ and a subset M_I of the function space $\mathbb{R} \Rightarrow |I|$ satisfying

1. for $\alpha \in M_I$ and a measurable function $f\colon \mathbb{R} \to \mathbb{R}$, $\alpha \circ f \in M_I$.
2. for any $x \in I$, $(\lambda r \in \mathbb{R}.x) \in M_I$.
3. for all $P\colon \mathbb{R} \to \mathbb{N}$ and a family $\{\alpha_i\}_{i \in \mathbb{N}}$ of functions $\alpha_i \in M_I$, $(\lambda r \in \mathbb{R}.\alpha_{P(r)}(r)) \in M_I$.

A morphism $f\colon (|I|, M_I) \to (|J|, M_J)$ is a function $f\colon |I| \to |J|$ such that $f \circ \alpha \in M_J$ holds for all $\alpha \in M_I$. The category **QBS** is a (well-pointed) CCC, and has all countable products and coproducts that are strictly preserved by the forgetful functor $|-|\colon \mathbf{QBS} \to \mathbf{Set}$. It is naturally isomorphic to the global element functor $\mathbf{QBS}(1, -)$.

Connection to Measurable Spaces: an Adjunction We can convert measurable spaces and quasi-Borel spaces using an adjunction $L \dashv K\colon \mathbf{Meas} \to \mathbf{QBS}$. They are given by

$$LI \triangleq (|I|, \{U \subseteq |I| \mid \forall \alpha \in M_X.\alpha^{-1}(I) \in \Sigma_\mathbb{R}\}) \qquad\qquad Lf \triangleq f$$

$$KI \triangleq (|I|, \mathbf{Meas}(\mathbb{R}, I)) \qquad\qquad Kf \triangleq f$$

For any standard Borel space $\Omega \in \mathbf{Meas}$, we have $LK\Omega = \Omega$. The right adjoint K is full-faithful when restricted to the standard Borel spaces [28, Proposition 15-(2)]. The right adjoint K preserves countable coproducts and function spaces (if exists) of standard Borel spaces [28, Proposition 19].

Probability Measures and the Probability Monad. A probability measure on a quasi-Borel space I is a pair $(\alpha, \mu) \in M_I \times G\mathbb{R}$. We introduce an equivalence relation $\sim_I$ over probability measures on I by

$$(\alpha, \mu) \sim_I (\beta, \nu) \iff \mu(\alpha^{-1}(-)) = \nu(\beta^{-1}(-)).$$

Using this, we introduce a probability monad P on $\mathbf{QBS}$ as follows:

- On objects, we define $P : \mathbf{Obj}(\mathbf{QBS}) \to \mathbf{Obj}(\mathbf{QBS})$ by

$$|P(I)| \triangleq (M_I \times G\mathbb{R})/\sim_I, \quad M_{P(I)} \triangleq \{\lambda r.[(\alpha, g(r))]_{\sim_I} \mid \alpha \in M_I, g \in \mathbf{Meas}(\mathbb{R}, G\mathbb{R})\}.$$

- The unit is defined by $\eta_I(x) \triangleq [\lambda r.x, \mu]_{\sim_I}$ for an arbitrary $\mu \in G\mathbb{R}$.
- The Kleisli extension of $f \colon I \to P(J)$ is defined by $f^\sharp[\alpha, \mu]_{\sim_I} \triangleq [\beta, g^\sharp\mu]$ where there are $\beta \in M_J$ and $g \in \mathbf{Meas}(\mathbb{R}, G\mathbb{R})$ satisfying $f \circ \alpha = \lambda r \in \mathbb{R}.[\beta, g(r)]_{\sim_J}$ by definition of $M_{P(J)}$.

The monad P is (commutative) strong with respect to the CCC $(\mathbf{QBS}, 1, (\times))$.

Acknowledgments

Tetsuya Sato carried out this research under the support by JST ERATO HASUO Metamathematics for Systems Design Project (No. JPMJER1603) and JSPS KAKENHI Grant Number 20K19775, Japan. Shin-ya Katsumata carried out this research under the support by JST ERATO HASUO Metamathematics for Systems Design Project (No. JPMJER1603) and JSPS KAKENHI Grant Number 18H03204, Japan. The authors are grateful to Ichiro Hasuo providing the opportunity of collaborating in that project. The authors are grateful to Satoshi Kura, Justin Hsu, Marco Gaboardi, Borja Balle and Gilles Barthe for fruitful discussions.

References

[1] Aws Albarghouthi and Justin Hsu. Constraint-based synthesis of coupling proofs. In *Computer Aided Verification - 30th International Conference, CAV 2018, Proceedings, Part I*, volume 10981 of *LNCS*, pages 327–346. Springer, 2018.

[2] Aws Albarghouthi and Justin Hsu. Synthesizing coupling proofs of differential privacy. *PACMPL*, 2(POPL):58:1–58:30, 2018.

[3] Thorsten Altenkirch, James Chapman, and Tarmo Uustalu. Monads need not be endofunctors. *Log. Methods Comput. Sci.*, 11(1), 2015.

[4] Robert J. Aumann. Borel structures for function spaces. *Illinois J. Math.*, 5(4):614–630, 12 1961.

[5] Arthur Azevedo de Amorim, Marco Gaboardi, Justin Hsu, and Shin-ya Katsumata. Probabilistic relational reasoning via metrics. In *34th Annual ACM/IEEE Symposium on Logic in Computer Science, LICS 2019*, pages 1–19. IEEE, 2019.

[6] Giorgio Bacci, Radu Mardare, Prakash Panangaden, and Gordon Plotkin. Tensor of Quantitative Equational Theories. In Fabio Gadducci and Alexandra Silva, editors, *9th Conference on Algebra and Coalgebra in Computer Science (CALCO 2021)*, volume 211 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 7:1–7:17, Dagstuhl, Germany, 2021. Schloss Dagstuhl – Leibniz-Zentrum für Informatik.

[7] Borja Balle, Gilles Barthe, Marco Gaboardi, Justin Hsu, and Tetsuya Sato. Hypothesis testing interpretations and renyi differential privacy. In Silvia Chiappa and Roberto Calandra, editors, *Proceedings of the Twenty Third International Conference on Artificial Intelligence and Statistics (AISTATS 2020)*, volume 108 of *Proceedings of Machine Learning Research*, pages 2496–2506, Online, 26–28 Aug 2020. PMLR.

[8] Gilles Barthe, Marco Gaboardi, Emilio Jesús Gallego Arias, Justin Hsu, César Kunz, and Pierre-Yves Strub. Proving differential privacy in Hoare logic. In *IEEE 27th Computer Security Foundations Symposium, CSF 2014*, pages 411–424. IEEE Computer Society, 2014.

[9] Gilles Barthe, Marco Gaboardi, Emilio Jesús Gallego Arias, Justin Hsu, Aaron Roth, and Pierre-Yves Strub. Higher-order approximate relational refinement types for mechanism design and differential privacy. In *Proceedings of the 42nd Annual ACM SIGPLAN-SIGACT Symposium on Principles of Programming Languages, POPL 2015, Mumbai, India, January 15-17, 2015*, pages 55–68. ACM, 2015.

[10] Gilles Barthe, Marco Gaboardi, Benjamin Grégoire, Justin Hsu, and Pierre-Yves Strub. Proving differential privacy via probabilistic couplings. In *Proceedings of the 31st Annual ACM/IEEE Symposium on Logic in Computer Science, LICS '16*, pages 749–758. ACM, 2016.

[11] Gilles Barthe, Benjamin Grégoire, Justin Hsu, and Pierre-Yves Strub. Coupling proofs are probabilistic product programs. In *Proceedings of the 44th ACM SIGPLAN Symposium on Principles of Programming Languages*, POPL 2017, pages 161–174, New York, NY, USA, 2017. Association for Computing Machinery.

[12] Gilles Barthe, Boris Köpf, Federico Olmedo, and Santiago Zanella Béguelin. Probabilistic relational reasoning for differential privacy. In *Proceedings of the 39th ACM SIGPLAN-SIGACT Symposium on Principles of Programming Languages, POPL 2012*, pages 97–110. ACM, 2012.

[13] Gilles Barthe and Federico Olmedo. Beyond differential privacy: Composition theorems and relational logic for f-divergences between probabilistic programs. In *Automata, Languages, and Programming - 40th International Colloquium, ICALP 2013, Proceedings, Part II*, volume 7966 of *LNCS*, pages 49–60. Springer, 2013.

[14] Nick Benton. Simple relational correctness proofs for static analyses and program transformations. *SIGPLAN Not.*, 39(1):14–25, January 2004.

[15] Filippo Bonchi, Barbara König, and Daniela Petrisan. Up-To Techniques for Behavioural Metrics via Fibrations. In *29th International Conference on Concurrency Theory (CONCUR 2018)*, volume 118 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 17:1–17:17, Dagstuhl, Germany, 2018. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik.

[16] Mark Bun, Cynthia Dwork, Guy N. Rothblum, and Thomas Steinke. Composable and versatile privacy via truncated CDP. In *Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing*, STOC 2018, pages 74–86, New York, NY, USA, 2018. Association for Computing Machinery.

[17] Mark Bun and Thomas Steinke. Concentrated differential privacy: Simplifications, extensions, and lower bounds. In *Theory of Cryptography*, pages 635–658, Berlin, Heidelberg, 2016. Springer Berlin Heidelberg.

[18] Ezgi Çiçek, Gilles Barthe, Marco Gaboardi, Deepak Garg, and Jan Hoffmann. Relational cost analysis. *SIGPLAN Not.*, 52(1):316–329, January 2017.

[19] Imre Csiszár. Eine informationstheoretische Ungleichung und ihre Anwendung auf den beweis der ergodizitat von markoffschen ketten. *Magyar. Tud. Akad. Mat. Kutato Int. Kozl.*, 8:85–108, 1963.

[20] Imre Csiszár. Information-type measures of difference of probability distributions and indirect observations. *Studia Sci. Math. Hungar.*, 2:299–318, 1967.

[21] Cynthia Dwork, Frank McSherry, Kobbi Nissim, and Adam Smith. Calibrating noise to sensitivity in private data analysis. In *Theory of Cryptography*, volume 3876 of *LNCS*, pages 265–284. Springer Berlin Heidelberg, 2006.

[22] Cynthia Dwork and Aaron Roth. The algorithmic foundations of differential privacy. *Foundations and Trends® in Theoretical Computer Science*, 9(3-4):211–407, 2013.

[23] Marco Gaboardi, Andreas Haeberlen, Justin Hsu, Arjun Narayan, and Benjamin C. Pierce. Linear dependent types for differential privacy. In *The 40th Annual ACM SIGPLAN-SIGACT Symposium on Principles of Programming Languages, POPL '13*, pages 357–370. ACM, 2013.

[24] Marco Gaboardi, Shin-ya Katsumata, Dominic Orchard, and Tetsuya Sato. Graded hoare logic and its categorical semantics. In Nobuko Yoshida, editor, *Programming Languages and Systems - 30th European Symposium on Programming, ESOP 2021, Held as Part of the European Joint Conferences on Theory and Practice of Software, ETAPS 2021, Luxembourg City, Luxembourg, March 27 - April 1, 2021, Proceedings*, volume 12648 of *Lecture Notes in Computer Science*, pages 234–263. Springer, 2021.

[25] Francesco Gavazzo. Quantitative behavioural reasoning for higher-order effectful programs: Applicative distances. In *Proceedings of the 33rd Annual ACM/IEEE Symposium on Logic in Computer Science*, LICS '18, pages 452–461, New York, NY, USA, 2018. Association for Computing Machinery.

[26] Michèle Giry. A categorical approach to probability theory. In B. Banaschewski, editor, *Categorical Aspects of Topology and Analysis*, volume 915 of *LNM*, pages 68–85. Springer, 1982.

[27] Rob Hall. *New Statistical Applications for Differential Privacy*. PhD thesis, Machine Learning Department School of Computer Science Carnegie Mellon University, 2012.

[28] Chris Heunen, Ohad Kammar, Sam Staton, and Hongseok Yang. A convenient category for higher-order probability theory. In *32nd Annual ACM/IEEE Symposium on Logic in Computer Science, LICS 2017*, pages 1–12, 2017.

[29] B. Jacobs. *Categorical Logic and Type Theory.* Elsevier, 1999.

[30] Peter Kairouz, Sewoong Oh, and Pramod Viswanath. The composition theorem for differential privacy. In *Proceedings of the 32nd International Conference on Machine Learning, ICML 2015, Lille, France, 6-11 July 2015*, pages 1376–1385, 2015.

[31] Shin-ya Katsumata. Parametric effect monads and semantics of effect systems. In *The 41st Annual ACM SIGPLAN-SIGACT Symposium on Principles of Programming Languages, POPL '14*, pages 633–646. ACM, 2014.

[32] Shin-ya Katsumata and Tetsuya Sato. Preorders on monads and coalgebraic simulations. In Frank Pfenning, editor, *Foundations of Software Science and Computation Structures*, pages 145–160, Berlin, Heidelberg, 2013. Springer Berlin Heidelberg.

[33] Shin-ya Katsumata, Tetsuya Sato, and Tarmo Uustalu. Codensity lifting of monads and its dual. *Logical Methods in Computer Science*, 14(4), 2018.

[34] Max Kelly. *Basic Concepts of Enriched Category Theory*, volume 64. Cambridge University Press, 1982. Republished in: Reprints in Theory and Applications of Categories, No. 10 (2005) pp.1-136.

[35] Friedrich Liese and Igor Vajda. On divergences and informations in statistics and information theory. *IEEE Transactions on Information Theory*, 52(10):4394–4412, Oct 2006.

[36] John M. Lucassen and David K. Gifford. Polymorphic effect systems. In *Conference Record of the Fifteenth Annual ACM Symposium on Principles of Programming Languages*, pages 47–57. ACM Press, 1988.

[37] Saunders Mac Lane. *Categories for the Working Mathematician (Second Edition)*, volume 5 of *Graduate Texts in Mathematics.* Springer, 1998.

[38] R. Mardare, P. Panangaden, and G. Plotkin. On the axiomatizability of quantitative algebras. In *2017 32nd Annual ACM/IEEE Symposium on Logic in Computer Science (LICS)*, pages 1–12, Los Alamitos, CA, USA, jun 2017. IEEE Computer Society.

[39] Radu Mardare, Prakash Panangaden, and Gordon Plotkin. Quantitative algebraic reasoning. In *Proceedings of the 31st Annual ACM/IEEE Symposium on Logic in Computer Science*, LICS '16, page 700–709, New York, NY, USA, 2016. Association for Computing Machinery.

[40] Ilya Mironov. Rényi differential privacy. In *2017 IEEE 30th Computer Security Foundations Symposium (CSF)*, pages 263–275, Aug 2017.

[41] John C. Mitchell and Andre Scedrov. Notes on sconing and relators. In *Computer Science Logic, 6th Workshop, CSL '92*, volume 702 of *LNCS*, pages 352–378. Springer, 1992.

[42] Eugenio Moggi. Notions of computation and monads. *Information and Computation*, 93(1):55–92, 1991.

[43] Tetsuzo Morimoto. Markov processes and the H-theorem. *Journal of the Physical Society of Japan*, 18(3):328–331, 1963.

[44] Hanne Riis Nielson and Flemming Nielson. *Semantics with Applications: An Appetizer.* Springer-Verlag, Berlin, Heidelberg, 2007.

[45] Federico Olmedo. *Approximate Relational Reasoning for Probabilistic Programs*. PhD thesis, Technical University of Madrid, 2014.

[46] Shiva Prasad and Kasiviswanathan Adam Smith. A note on differential privacy: Defining resistance to arbitrary side information. *Journal of Privacy and Confidentiality*, 6(1), 2014.

[47] Ivan Radiček, Gilles Barthe, Marco Gaboardi, Deepak Garg, and Florian Zuleger. Monadic refinements for relational cost analysis. *Proc. ACM Program. Lang.*, 2(POPL):36:1–36:32, December 2017.

[48] Jason Reed and Benjamin C. Pierce. Distance makes the types grow stronger: a calculus for differential privacy. In *Proceeding of the 15th ACM SIGPLAN international conference on Functional programming, ICFP 2010*, pages 157–168. ACM, 2010.

[49] J. J. M. M. Rutten. Elements of generalized ultrametric domain theory. *Theor. Comput. Sci.*, 170(1-2):349–381, December 1996.

[50] Tetsuya Sato. Identifying all preorders on the subdistribution monad. In Bart Jacobs, Alexandra Silva, and Sam Staton, editors, *Proceedings of the 30th Conference on the Mathematical Foundations of Programming Semantics, MFPS 2014, Ithaca, NY, USA, June 12-15, 2014*, volume 308 of *Electronic Notes in Theoretical Computer Science*, pages 309–327. Elsevier, 2014.

[51] Tetsuya Sato. Approximate relational hoare logic for continuous random samplings. In *The Thirty-second Conference on the Mathematical Foundations of Programming Semantics, MFPS 2016*, volume 325 of *Electronic Notes in Theoretical Computer Science*, pages 277–298. Elsevier, 2016.

[52] Tetsuya Sato, Gilles Barthe, Marco Gaboardi, Justin Hsu, and Shin-ya Katsumata. Approximate span liftings: Compositional semantics for relaxations of differential privacy. In *34th Annual ACM/IEEE Symposium on Logic in Computer Science, LICS 2019*, pages 1–14. IEEE, 2019.

[53] Ross Street. The formal theory of monads. *Journal of Pure and Applied Algebra*, 2(2):149 – 168, 1972.

[54] Larry Wasserman and Shuheng Zhou. A statistical framework for differential privacy. *Journal of the American Statistical Association*, 105(489):375–389, 2010.

A Proofs for Section 5 (Examples of Divergences on Monads)

Proposition 17. *The family* $\mathsf{C}' = \{\mathsf{C}'_I \colon (\mathbb{N} \times I)^2 \to \mathcal{N}\}_{I \in \mathbf{Set}}$ *of* $\mathcal{N}$*-divergences defined by*

$$\mathsf{C}'_I((i,x),(j,y)) \triangleq \begin{cases} |i-j| & x = y \\ \infty & x \neq y \end{cases}.$$

is a Eq*-relative* $\mathcal{N}$*-divergence on the monad* $\mathbb{N} \times -$.

Proof. The monotonicity of C' is obvious.

We show the Eq-unit-reflexivity of C'. For all $(x,y) \in \mathrm{Eq}I$ (that is, $x = y \in I$), we have

$$\mathsf{C}'_I(\eta_I(x), \eta_I(y)) = \mathsf{C}'_I((0,x),(0,y)) = 0.$$

We show the Eq-composability of C'. Let $(i,x),(j,y) \in \mathbb{N} \times I$ and $f,g \colon I \to \mathbb{N} \times J$. We write $f(z) = (i_z, f_z)$ and $g(z) = (j_z, g_z)$ for each $z \in Z$.

- If $x = y$ and $x_z = y_z$ for all $z \in I$, we have

$$\begin{aligned} \mathsf{C}'_J(f^\sharp(i,x), g^\sharp(j,y)) &= \mathsf{C}'_J(i + i_x, f_x), (j + j_x, g_x)) \\ &= |(i + i_x) - (j + j_x)| \leq |i - j| + |i_x - j_x| \\ &\leq \mathsf{C}'_I((i,x),(j,y)) + \sup_{(x,y) \in \mathrm{Eq}I (\iff x = y \in I)} \mathsf{C}'_J(f(x), g(y)) \end{aligned}$$

- If $x \neq y$ or $f_z \neq g_z$ for some $z \in I$, we have

$$\mathsf{C}'_J(f^\sharp(i,x), g^\sharp(j,y)) \leq \infty = \mathsf{C}'_I((i,x),(j,y)) + \sup_{(x,y) \in \mathrm{Eq}I (\iff x = y \in I)} \mathsf{C}'_J(f(x), g(y)).$$

This completes the proof. □

Proposition 18. *The family* $\mathsf{NC} = \{\mathsf{NC}_I \colon (P(\mathbb{N} \times I))^2 \to \mathcal{N}\}_{I \in \mathbf{Set}}$ *of* $\mathcal{N}$*-divergences defined by*

$$\mathsf{NC}_I(A,B) \triangleq \sup_{(i,x) \in A, (j,x) \in B} |i - j|$$

is a Top*-relative* $\mathcal{N}$*-divergence on the monad* $P(\mathbb{N} \times -)$.

Proof. The monotonicity of NC is obvious.

We show the Top-unit-reflexivity of NC . For all $(x,y) \in \mathrm{Top}I$ (that is, $x, y \in I$), we have

$$\mathsf{NC}_I(\eta_I(x), \eta_I(y)) = \mathsf{NC}_I(\{(0,x)\}, \{(0,y)\}) = |0 - 0| = 0.$$

We show the Top-composability of NC. For all $f, g \colon I \to P(\mathbb{N} \times J)$ and $A, B \in P(\mathbb{N} \times I)$, we have

$$\begin{aligned} \mathsf{NC}_J(f^\sharp A, g^\sharp B) &= \sup\{|i - j| \mid (i,x) \in f^\sharp(A), (j,y) \in g^\sharp(B)\} \\ &= \sup\left\{ |i_1 + i_2 - j_1 - j_2| \;\middle|\; \begin{array}{l} (i_1, x) \in A, (j_1, y) \in B, \\ (i_2, x') \in f(x), (j_2, y') \in g(y) \end{array} \right\} \\ &\leq \sup\{|i_1 - j_1| \mid (i_1, x) \in A, (j_1, y) \in B\} \\ &\qquad + \sup_{(x,y) \in \mathrm{Top}I (\iff x,y \in I)} \{|i_2 - j_2| \mid (i_2, x') \in f(x), (j_2, y') \in g(y)\} \\ &= \mathsf{NC}_I(A,B) + \sup_{(x,y) \in \mathrm{Top}I (\iff x,y \in I)} \mathsf{NC}_J(f(x), g(y)). \end{aligned}$$

This completes the proof. □

Proposition 19. *The family* $\mathsf{NCI} = \{\mathsf{NCI}_I \colon (P(\mathbb{N}\times I))^2 \to \mathcal{N}\}_{I\in\mathbf{Set}}$ *of* $\mathcal{Z}$*-divergences defined by*

$$\mathsf{NCI}_I(A,B) \triangleq \sup_{(i,x)\in A,(j,y)\in B} i-j$$

is a Top*-relative* $\mathcal{Z}$*-divergence on the monad* $P(\mathbb{N}\times -)$*.*

Proof. The monotonicity of NCI is obvious.

We show the Top-unit-reflexivity of NCI . For all $(x,y)\in \mathrm{Top}I$ (that is, $x,y\in I$), we have

$$\mathsf{NCI}_I(\eta_I(x),\eta_I(y)) = \mathsf{NCI}_I(\{(0,x)\},\{(0,y)\}) = 0-0 = 0.$$

We show the Top-composability of NCI. For all $f,g\colon I\to P(\mathbb{N}\times J)$ and $A,B\in P(\mathbb{N}\times I)$, we have

$$\begin{aligned}
\mathsf{NCI}_J(f^\sharp A, g^\sharp B) &= \sup\{i-j \mid (i,x)\in f^\sharp(A)\wedge (j,y)\in g^\sharp(B)\}\\
&= \sup\left\{ i_1+i_2-j_1-j_2 \;\middle|\; \begin{array}{l}(i_1,x)\in A,(j_1,y)\in B,\\ (i_2,x')\in f(x),(j_2,y')\in g(y)\end{array}\right\}\\
&\le \sup\{i_1-j_1 \mid (i_1,x)\in A,(j_1,y)\in B\}\\
&\qquad + \sup_{(x,y)\in\mathrm{Top}I(\iff x,y\in I)} \{i_2-j_2 \mid (i_2,x')\in f(x),(j_2,y')\in g(y)\}\\
&= \mathsf{NCI}_I(A,B) + \sup_{(x,y)\in\mathrm{Top}I(\iff x,y\in I)} \mathsf{NCI}_J(f(x),g(y)).
\end{aligned}$$

This completes the proof. $\square$

Proof. (Proof of Proposition 1) We have Eq-unit reflexivity because the reflexivity ${}^f\mathsf{Div}_I(\mu,\mu) = 0$ is obtained from $f(1)=0$. We show Eq-composability. To show this, we prove a bit stronger statement. Consider three positive weight functions $f,f_1,f_2\ge 0$ with $f(1)=f_1(1)=f_2(1)=0$. Assume that there are some $\alpha,\beta,\beta'\in\mathbb{R}$ satisfying the following conditions:

(A')

for all $x,y,z,w\in[0,1]$, $0\le(\beta' z+(1-\beta')x)+\gamma x f_1(z/x)$ and

$$\begin{aligned}
xyf(zw/xy) \le{}& (\beta w+(1-\beta)y)xf_1(z/x) + (\beta' z+(1-\beta')x)yf_2(w/y)\\
&+\gamma xyf_1(z/x)f_2(w/y)+\alpha(x-z)(w-y).
\end{aligned}$$

Let $\mu_1,\mu_2\in G_sI$, and let $h,k\colon I\to G_sJ$. We want to show the composability in the sense of [45, Definition 5.2]:

$$\begin{aligned}
&{}^f\mathsf{Div}_J(h^\sharp\mu_1,k^\sharp\mu_2)\\
&\le {}^{f_1}\mathsf{Div}_I(\mu_1,\mu_2) + \sup_{x\in I}{}^{f_2}\mathsf{Div}_J(h(x),k(x)) + \gamma\,{}^{f_1}\mathsf{Div}_I(\mu_1,\mu_2)\cdot\sup_{x\in I}{}^{f_2}\mathsf{Div}_J(h(x),k(x)).
\end{aligned} \tag{23}$$

We first fix a *measurable partition* $\{A_i\}_{i=0}^n$ of J, that is a family $\{A_i\}_{i=0}^n$ of measurable subsets $A_i\in\Sigma_J$ satisfying $i\ne j \implies A_i\cap A_j=\emptyset$ and $\bigcup_{i=0}^n A_i = J$. For each $0\le i\le n$, we fix two monotone increasing sequences $\{h_l^i\}_{l=0}^\infty$ and $\{k_l^i\}_{l=0}^\infty$ of simple functions that converge uniformly to measurable functions $h(-)(A_i)\colon I\to[0,1]$ and $k(-)(A_i)\colon I\to[0,1]$ respectively. The above composability (23) is then equivalent to

$$\begin{aligned}
&\lim_{l\to\infty}\sum_{i=0}^n \left(\int_X k_l^i\,d\mu_2\right) f\left(\frac{\int_X h_l^i\,d\mu_1}{\int_X k_l^i\,d\mu_2}\right)\\
&\le {}^{f_1}\mathsf{Div}_I(\mu_1,\mu_2) + \sup_{x\in I}{}^{f_2}\mathsf{Div}_J(h(x),k(x)) + \gamma\,{}^{f_1}\mathsf{Div}_I(\mu_1,\mu_2)\sup_{x\in I}{}^{f_2}\mathsf{Div}_J(h(x),k(x)).
\end{aligned} \tag{24}$$

We fix $l \in \mathbb{N}$. We suppose $h_l^i = \sum_{j=0}^m \alpha_j^i \chi_{B_j}$ and $k_l^i = \sum_{j=0}^m \beta_j^i \chi_{B_j}$ for some $\alpha_j^i, \beta_j^i \in [0,1]$ $(0 \leq j \leq m)$ and a measurable partition $\{B_j\}_{j=0}^m$ of I.

Thanks to the condition **(A')**, we calculate as follows:

$$\begin{aligned}
&\sum_{i=0}^{n}(\int_X k_l^i \ d\mu_2) f\left(\frac{\int_X h_l^i \ d\mu_1}{\int_X k_l^i \ d\mu_2}\right) \\
&\leq \sum_{i=0}^{n}\sum_{j=0}^{m} \beta_j^i \mu_2(B_j) f\left(\frac{\alpha_j^i \mu_1(B_j)}{\beta_j^i \mu_2(B_j)}\right) \\
&\leq \quad \underbrace{\sum_{i=0}^{n}\sum_{j=0}^{m} (\beta\alpha_j^i + (1-\beta)\beta_j^i) \ \mu_2(B_j) f_1\left(\frac{\mu_1(B_j)}{\mu_2(B_j)}\right)}_{\triangleq V_1} \\
&\quad + \underbrace{\sum_{i=0}^{n}\sum_{j=0}^{m} \left(\beta' \mu_1(B_j) + (1-\beta')\mu_2(B_j)) + \gamma\mu_2(B_j) f_1\left(\frac{\mu_1(B_j)}{\mu_2(B_j)}\right)\right) \beta_j^i f_2\left(\frac{\alpha_j^i}{\beta_j^i}\right)}_{\triangleq V_2} \\
&\quad + \underbrace{\sum_{i=0}^{n}\sum_{j=0}^{m} \alpha(\mu_2(B_j) - \mu_1(B_j))(\alpha_j^i - \beta_j^i)}_{\triangleq V_3}
\end{aligned}$$

We evaluate the above three subexpressions V_1, V_2, V_3 as follows.

We evaluate V_1 as follows:

$$\begin{aligned}
V_1 &\leq \left(\sup_{0 \leq j \leq m} \sum_{i=0}^{n} (\beta\alpha_j^i + (1-\beta)\beta_j^i)\right) \cdot \sum_{j=0}^{m} \mu_2(B_j) f_1\left(\frac{\mu_1(B_j)}{\mu_2(B_j)}\right) \\
&= \sup_{x \in I}\left(\beta \sum_{i=0}^{n} h_l^i(x) + (1-\beta)\sum_{i=0}^{n} k_l^i(x)\right) \cdot \sum_{j=0}^{m} \mu_2(B_j) f_1\left(\frac{\mu_1(B_j)}{\mu_2(B_j)}\right) \\
&\leq \sup_{x \in I}\left(\beta \sum_{i=0}^{n} h_l^i(x) + (1-\beta)\sum_{i=0}^{n} k_l^i(x)\right) \cdot {}^{f_1}\mathsf{Div}_I(\mu_1, \mu_2) \\
&\xrightarrow{l \to \infty} \sup_{x \in I} \left(\beta h(x)(J) + (1-\beta) k(x)(J)\right) \cdot {}^{f_1}\mathsf{Div}_I(\mu_1, \mu_2) \\
&\leq {}^{f_1}\mathsf{Div}_I(\mu_1, \mu_2)
\end{aligned}$$

Here, the first inequality is given from the non-negativity of each $\mu_2(B_j) f_1\left(\frac{\mu_1(B_j)}{\mu_2(B_j)}\right)$; the equality is given by definition of α_j^i and β_j^i; the second inequality can be given by the continuity of ${}^{f_1}\mathsf{Div}$ ([35, Theorem 16]; [52, Theorem 3] for the sub-Giry monad G_s):

$${}^{f_1}\mathsf{Div}_I(\mu_1, \mu_2) = \sup\left\{ \sum_{j=0}^{m} \mu_2(B_j) f_1\left(\frac{\mu_1(B_j)}{\mu_2(B_j)}\right) \middle| \{B_j\}_{j=0}^m\text{: measurable partition of } I \right\};$$

the last inequality is derived by $\beta h(x)(J) + (1-\beta) k(x)(J) \in [0,1]$ from the assumption that either $\beta \in [0,1]$ or $h(x)(J) = k(x)(J)$ for all $x \in I$ holds.

We next evaluate V_2 as follows:

$$
\begin{aligned}
V_2 &\le \left(\sup_{0\le j\le m}\sum_{i=0}^{n}\beta_j^i f_2\left(\frac{\alpha_j^i}{\beta_j^i}\right)\right)\sum_{j=0}^{m}\left(\beta'\mu_1(B_j)+(1-\beta')\mu_2(B_j))+\gamma\mu_2(B_j)f_1\left(\frac{\mu_1(B_j)}{\mu_2(B_j)}\right)\right)\\
&= \left(\sup_{x\in I}\sum_{i=0}^{n}k_l^i(x)f_2\left(\frac{h_l^i(x)}{k_l^i(x)}\right)\right)\left(\beta'\mu_1(I)+(1-\beta')\mu_2(I)+\gamma\sum_{j=0}^{m}\mu_2(B_j)f_1\left(\frac{\mu_1(B_j)}{\mu_2(B_j)}\right)\right)\\
&\le \left(\sup_{x\in I}\sum_{i=0}^{n}k_l^i(x)f_2\left(\frac{h_l^i(x)}{k_l^i(x)}\right)\right)\left(\beta'\mu_1(I)+(1-\beta')\mu_2(I)+\gamma^{f_1}\mathsf{Div}_I(\mu_1,\mu_2)\right)\\
&\xrightarrow{l\to\infty}\left(\sup_{x\in I}\sum_{i=0}^{n}k(x)(A_i)f_2\left(\frac{h(x)(A_i)}{k(x)(A_i)}\right)\right)\left(\beta'\mu_1(I)+(1-\beta')\mu_2(I)+\gamma^{f_1}\mathsf{Div}_I(\mu_1,\mu_2)\right)\\
&\le \sup_{x\in I}{}^{f_2}\mathsf{Div}_J(h(x),k(x))\left(\beta'\mu_1(I)+(1-\beta')\mu_2(I)+\gamma^{f_1}\mathsf{Div}_I(\mu_1,\mu_2)\right)\\
&\le \sup_{x\in I}{}^{f_2}\mathsf{Div}_J(h(x),k(x))\cdot\gamma^{f_1}\mathsf{Div}_I(\mu_1,\mu_2)\\
&= \gamma^{f_1}\mathsf{Div}_I(\mu_1,\mu_2)\cdot\sup_{x\in I}{}^{f_2}\mathsf{Div}_J(h(x),k(x)).
\end{aligned}
$$

Here, the first inequality is derived from the non-negativity of each

$$
(\beta'\mu_1(B_j)+(1-\beta')\mu_2(B_j))+\gamma\mu_2(B_j)f_1\left(\frac{\mu_1(B_j)}{\mu_2(B_j)}\right); \tag{25}
$$

the first equality is given by definition of α_j^i and β_j^i and the countable additivity of μ_1 and μ_2; the second inequality is given by the continuity of ${}^{f_1}\mathsf{Div}$ and $0\le\gamma$; the last inequality is derived by $\beta'\mu_1(I)+(1-\beta')\mu_2(I)\in[0,1]$ from the assumption that either $\beta'\in[0,1]$ or $\mu_1(I)=\mu_2(I)$ holds. We prove the third inequality. Since f_2 is convex function, and sequences $\{h_l^i(x)\}_{l=0}^{\infty}$ and $\{k_l^i(x)\}_{l=0}^{\infty}$ are monotone increasing at each $x\in I$, By Jensen's inequality, the sequence $\left\{\sum_{i=0}^{n}k_l^i(x)f_2\left(h_l^i(x)/k_l^i(x)\right)\right\}_{l=0}^{\infty}$ is monotone increasing for each $x\in I$. Then, the sequence $\left\{\sup_{x\in I}\sum_{i=0}^{n}k_l^i(x)f_2\left(h_l^i(x)/k_l^i(x)\right)\right\}_{l=0}^{\infty}$ of supremums is also monotone increasing, because each $\sum_{i=0}^{n}k_{l+1}^i(x)f_2\left(h_{l+1}^i(x)/k_{l+1}^i(x)\right)$ is always greater than $\sum_{i=0}^{n}k_l^i(x)f_2\left(h_l^i(x)/k_l^i(x)\right)$. Hence,

$$
\begin{aligned}
\lim_{l\to\infty}\sup_{x\in I}\sum_{i=0}^{n}k_l^i(x)f_2\left(\frac{h_l^i(x)}{k_l^i(x)}\right) &= \sup_{l\in\mathbb{N}}\sup_{x\in I}\sum_{i=0}^{n}k_l^i(x)f_2\left(\frac{h_l^i(x)}{k_l^i(x)}\right)\\
&= \sup_{x\in I}\sup_{l\in\mathbb{N}}\sum_{i=0}^{n}k_l^i(x)f_2\left(\frac{h_l^i(x)}{k_l^i(x)}\right)\\
&= \sup_{x\in I}\sum_{i=0}^{n}k(x)(A_i)f_2\left(\frac{h(x)(A_i)}{k(x)(A_i)}\right)\\
&\le \sup_{x\in I}{}^{f_2}\mathsf{Div}(h(x),k(x)).
\end{aligned}
$$

Finally, we evaluate V_3 as follows:

$$
V_3=\sum_{j=0}^{m}\alpha(\mu_2(B_j)-\mu_1(B_j))(\sum_{i=0}^{n}\alpha_j^i-\beta_j^i)
$$

$$
\begin{aligned}
&= \alpha \left(\int_I h_l^i \; d\mu_2 - \int_I k_l^i \; d\mu_2 + \int_I k_l^i \; d\mu_1 - \int_I h_l^i \; d\mu_1 \right) \\
&\xrightarrow{l \to \infty} \alpha \left(\int_I h(-)(J) \; d\mu_2 - \int_I k(-)(J) \; d\mu_2 + \int_I k(-)(J) \; d\mu_1 - \int_I h(-)(J) \; d\mu_1 \right).
\end{aligned}
$$

Here, if either $\alpha = 0$ or $h(x)(J) = k(x)(J)$ for any $x \in I$ holds then the limit will be 0. To sum up the above evaluations of V_1, V_2, V_3, we obtain the inequality (24) if we have either

1. $\mu_1(I) = \mu_2(I) = 1$ and $\forall x \in I.\ h(x)(J) = k(x)(J) = 1$, or
2. $\alpha = 0$ and $\beta, \beta \in [0, 1]$.

This completes the proof. □

Parameters for Proposition 1 for for weight functions of TV, KL, HD and Chi are shown in Table 4. Below, we check the conditions in Proposition 1.

- For the weight function $f(t) = |t - 1|/2$ of TV, the tuple $(\gamma, \alpha, \beta, \beta') = (0, 0, 1, 0)$ satisfies for all $x, y, z, w \in [0, 1]$, we have

$$
\begin{aligned}
0 &\leq w + xf(z/x), \\
xyf(zw/xy) = |zw - xy|/2 &\leq |zw - wx| + |xw - xy|/2 = wxf(z/x) + xf(|w/y)/2.
\end{aligned}
$$

- For the weight function $f(t) = t \log(t) - t + 1$ of KL, the tuple $(\gamma, \alpha, \beta, \beta') = (0, -1, 1, 1)$ satisfies for all $x, y, z, w \in [0, 1]$, we have

$$
\begin{aligned}
&0 \leq z + xf(z/x), \\
&\ xy((zw/xy)\log(zw/xy) - zw/xy + 1) \\
&\ = zw \log(w/y) + zw \log(z/x) - zw + xy \\
&\ = xw((z/x)\log(z/x) - z/x + 1) + zy((w/y)\log(w/y) - w/y + 1) - (x - z)(w - y).
\end{aligned}
$$

- For the weight function $f(t) = (\sqrt{t} - 1)^2/2$ of HD, the tuple $(\gamma, \alpha, \beta, \beta') = (0, -1/4, 1/2, 1/2)$ satisfies for all $x, y, z, w \in [0, 1]$,

$$
\begin{aligned}
0 &\leq (z + x)/2 + f(z/x), \\
xyf(zw/xy) &= (zw + xy)/2 - ((x + z) - (\sqrt{x} - \sqrt{z})^2)((y + w) - (\sqrt{y} - \sqrt{w})^2)/4 \\
&= (zw + xy)/2 - ((x + z) - xf(z/x))((y + w) - yf(w/y))/4 \\
&\leq (y + w)/2 \cdot xf(z/x) + (x + z)/2 \cdot yf(w/y) - (x - z)(w - y)/4.
\end{aligned}
$$

- For the weight function $f(t) = (t - 1)^2/2$ of Chi, The tuple $(\gamma, \alpha, \beta, \beta') = (1, -2, 2, 2)$ satisfies for all $x, y, z, w \in [0, 1]$,

$$
\begin{aligned}
&0 \leq (2z - x) + xf(z/x) = (2z - x) + ((z/x) - 1)(z - x) = z + (z^2/x), \\
&\ xyf(zw/xy) = z^2w^2/xy + xy - 2zw \\
&\ = (xf(z/x) + 2z - x)(yf(w/y) + 2w - y) - 2zw + xy \\
&\ = (2w - y)xf(z/x) + (2z - x)yf(w/y) + xyf(z/x)f(w/y) - 2(x - z)(w - y).
\end{aligned}
$$

Proof. (Proof of Proposition 2)

We first show the monotonicity of $\langle p, \lambda\rangle^* \Delta$. Assume $m \le m'$. From the monotonicity of the original Δ, we obtain for each $\nu_1, \nu_2 \in U^{\mathbb{C}}(SI)$),

$$\begin{aligned}(\langle p, \lambda\rangle^* \Delta)^m_I(\nu_1, \nu_2) &= \Delta^m_{pI}((U^{\mathbb{D}}\lambda_I)(\nu_1), (U^{\mathbb{D}}\lambda_I)(\nu_2)) \\ &\ge \Delta^{m'}_{pI}((U^{\mathbb{D}}\lambda_I)(\nu_1), (U^{\mathbb{D}}\lambda_I)(\nu_2)) \\ &= (\langle p, \lambda\rangle^* \Delta)^{m'}_I(\nu_1, \nu_2).\end{aligned}$$

Second, we show the F-unit-reflexivity of $\langle p, \lambda\rangle^* \Delta$. For $FI = (I, I, R_{FI})$, we have $EpI = (p = (pI, pI, R_{FI})$ for all $I \in \mathbb{C}$. We can calculate for all $(x, y) \in R_F$,

$$\begin{aligned}(\langle p, \lambda\rangle^* \Delta)^{1_M}_I(\eta^S_I \bullet x, \eta^S_I \bullet y) &= \Delta^{1_M}_{pI}(U^{\mathbb{D}}\lambda_I \circ U^{\mathbb{C}}\eta^S_I \circ x, U^{\mathbb{D}}\lambda_I \circ U^{\mathbb{C}}\eta^S_I \circ y) \\ &= \Delta^{1_M}_{pI}((\lambda_I \circ p\eta^S_I) \bullet x, (\lambda_I \circ p\eta^S_I) \bullet y) \\ &= \Delta^{1_M}_{pI}(\eta^T_{pI} \bullet x, \eta^T_{pI} \bullet y) \le 0.\end{aligned}$$

Finally, we show the F-composability of $\langle p, \lambda\rangle^* \Delta$. For all $J \in \mathbb{C}$, $c_1, c_2 \in U^{\mathbb{C}}TI$, and $f_1, f_2 \colon I \to SJ$ we can calculate

$$\begin{aligned}(\langle p, \lambda\rangle^* \Delta)^{mn}_J(f_1^\sharp \bullet c_1, f_2^\sharp \bullet c_2) &= \Delta^{mn}_{pJ}(U^{\mathbb{D}}\lambda_J \circ U^{\mathbb{D}}p(f_1^\sharp) \circ c_1, U^{\mathbb{D}}\lambda_J \circ U^{\mathbb{D}}p(f_2^\sharp) \circ c_2) \\ &= \Delta^{mn}_{pJ}(U^{\mathbb{D}}((\lambda_J \circ pf_1)^\sharp) \circ U^{\mathbb{D}}\lambda_I \circ c_1, U^{\mathbb{D}}((\lambda_J \circ pf_2)^\sharp) \circ U^{\mathbb{D}}\lambda_I \circ c_2) \\ &= \Delta^{mn}_{pJ}((\lambda_J \circ pf_1)^\sharp \bullet (\lambda_I \bullet c_1), (\lambda_J \circ pf_2)^\sharp \bullet (\lambda_I \bullet c_2)) \\ &\le \Delta^m_{pI}(\lambda_I \bullet c_1, \lambda_I \bullet c_2) + \sup_{(x,y) \in EpI} \Delta^n_{pJ}((\lambda_J \circ pf_1) \bullet x, (\lambda_J \circ pf_2) \bullet x) \\ &= (\langle p, \lambda\rangle^* \Delta)^m_I(c_1, c_2) + \sup_{(x,y) \in FI} (\langle p, \lambda\rangle^* \Delta)^n_J(f_1 \bullet x, f_2 \bullet y).\end{aligned}$$

To prove the second equality, we calculate

$$\begin{aligned}U^{\mathbb{D}}\lambda_J \circ U^{\mathbb{D}}p(f_i^\sharp) &= U^{\mathbb{D}}(\lambda_J \circ p\mu^S_J \circ pSf_i) = U^{\mathbb{D}}(\mu^T_{pJ} \circ T\lambda_J \circ \lambda_{SJ} \circ pSf_i) \\ &= U^{\mathbb{D}}(\mu^T_{pJ} \circ T\lambda_J \circ Tpf_i \circ \lambda_I) = U^{\mathbb{D}}((\lambda_J \circ pf_i)^\sharp \circ \lambda_I).\end{aligned}$$

This completes the proof. □

Proof. (Proof of Proposition 3)

It suffices to show Top-unit reflexivity and Top-composability:

$$\Delta^{\mathsf{lip}, d_S}_I(\eta_I(x), \eta_I(y)) = \sup_{s', s \in S} \frac{d_S(\pi_2(s, x), \pi_2(s', y))}{d_S(s, s')} = \frac{d_S(s, s')}{d_S(s, s')} = 1,$$

$$\begin{aligned}&\Delta^{\mathsf{lip}, d_S}_J(F_1^\sharp(f_1), F_1^\sharp(f_2)) \\ &= \sup_{s', s \in S} \frac{d_S(\pi_2(F_1(\pi_1 f_1(s))(\pi_2 f_1(s))), \pi_2(F_2(\pi_1 f_2(s'))(\pi_2 f_2(s'))))}{d_S(s, s')} \\ &= \sup_{s', s \in S} \frac{d_S(\pi_2 f_1(s), \pi_2 f_2(s'))}{d_S(s, s')} \cdot \frac{d_S(\pi_2(F_1(\pi_1 f_1(s))(\pi_2 f_1(s))), \pi_2(F_2(\pi_1 f_2(s'))(\pi_2 f_2(s'))))}{d_S(\pi_2 f_1(s), \pi_2 f_2(s'))} \\ &\le \sup_{s', s \in S} \frac{d_S(\pi_2 f_1(s), \pi_2 f_2(s'))}{d_S(s, s')} \cdot \sup_{t', t \in S} \frac{d_S(\pi_2(F_1(\pi_1 f_1(s))(t)), \pi_2(F_2(\pi_1 f_2(s'))(t')))}{d_S(t, t')} \\ &\le \Delta^{\mathsf{lip}, d_S}_I(f_1, f_2) \cdot \sup_{x, y \in I} \Delta^{\mathsf{lip}, d_S}_J(F_1(x), F_2(y))\end{aligned}$$

Here $F_1, F_2 \colon I \to T_S J$ and $f_1, f_2 \in T_S I$. □

Proof. (Proof of Proposition 4)

It suffices to show Eq-unit reflexivity and Eq-composability:

$$
\begin{aligned}
\Delta_I^{\mathsf{met},d_S}(\eta_I(x),\eta_I(x)) &= \sup_{s\in S} d_S(\pi_2(x,s),\pi_2(x,s)) = \sup_{s\in S} d_S(s,s) = 0.\\
\Delta_J^{\mathsf{met},d_S}(F_1^\sharp(f_1),F_1^\sharp(f_2)) &= \sup_{s\in S} d_S(\pi_2(F_1(\pi_1 f_1(s))(\pi_2 f_1(s))),\pi_1(F_2(\pi_1 f_2(s))(\pi_2 f_2(s))))\\
&\leq \sup_{s\in S} d_S(\pi_2(F_1(\pi_1 f_1(s))(\pi_2 f_1(s))),\pi_2(F_2(\pi_1 f_1(s))(\pi_2 f_1(s))))\\
&\quad + \sup_{s\in S} d_S(\pi_2(F_2(\pi_1 f_1(s))(\pi_2 f_1(s))),\pi_2(F_2(\pi_1 f_1(s))(\pi_2 f_2(s))))\\
&\leq \sup_{x\in I} \Delta_J^{\mathsf{met},d_S}(F_1(x),F_2(x)) + \Delta_I^{\mathsf{met},d_S}(f_1,f_2)
\end{aligned}
$$

Here $F_1,F_2\colon I\to T_S J$ and $f_1,f_2\in T_S I$. Without loss of generality, we may assume $\pi_1 f_1=\pi_1 f_2$ holds and $\pi_2 f_1$ and $\pi_2 f_2$ are nonexpansive, and for every $x\in I$, $\pi_1 F_1(x)=\pi_1 F_2(x)$ holds and $\pi_2 F_1(x)$ and $\pi_2 F_2(x)$ are nonexpansive. ☐

Proof. (Proof of Proposition 5) We first show the Eq-unit reflexivity of $d^{T_S(-)}$. For any $s\in S$, we calculate

$$
\begin{aligned}
d_{T_S I}(\eta_I(x),\eta_I(x)) &= \sup_{s\in S}\max\left(d_I(\pi_1(x,s),\pi_1(x,s)),d_S(\pi_2(x,s),\pi_2(x,s)\right)\\
&= \sup_{s\in S}\max(d_I(x,x),d_S(s,s)) = 0.
\end{aligned}
$$

We next show the Eq-composability of $d^{T_S(-)}$. For any $f_1,f_2\in T_S(I,d_I)$ and nonexpansive functions $F_1,F_2\colon(I,d_I)\to T_S(J,d_J)$, we compute

$$
\begin{aligned}
d^{T_S J}(F_1^\sharp(f_1),F_1^\sharp(f_2)) &= \sup_{s\in S}\max\begin{pmatrix} d_J(\pi_1(F_1(\pi_1 f_1(s))(\pi_2 f_1(s))),\pi_1(F_2(\pi_1 f_2(s))(\pi_2 f_2(s))),\\ d_S(\pi_2(F_1(\pi_1 f_1(s))(\pi_2 f_1(s))),\pi_2(F_2(\pi_1 f_2(s))(\pi_2 f_2(s)))\end{pmatrix}\\
&\leq \sup_{s\in S}\max\begin{pmatrix} d_J(\pi_1(F_1(\pi_1 f_1(s))(\pi_2 f_1(s))),\pi_1(F_2(\pi_1 f_1(s))(\pi_2 f_1(s))),\\ d_J(\pi_1(F_2(\pi_1 f_1(s))(\pi_2 f_1(s))),\pi_1(F_2(\pi_1 f_2(s))(\pi_2 f_2(s))),\\ d_S(\pi_2(F_1(\pi_1 f_1(s))(\pi_2 f_1(s))),\pi_2(F_2(\pi_1 f_1(s))(\pi_2 f_1(s))),\\ d_S(\pi_2(F_2(\pi_1 f_1(s))(\pi_2 f_1(s))),\pi_2(F_2(\pi_1 f_2(s))(\pi_2 f_2(s)))\end{pmatrix}\\
&= \sup_{s\in S}\max\begin{pmatrix} d_J(\pi_1(F_2(\pi_1 f_1(s))(\pi_2 f_1(s))),\pi_1(F_2(\pi_1 f_2(s))(\pi_2 f_2(s))),\\ d_S(\pi_2(F_2(\pi_1 f_1(s))(\pi_2 f_1(s))),\pi_2(F_2(\pi_1 f_2(s))(\pi_2 f_2(s))),\\ d_J(\pi_1(F_1(\pi_1 f_1(s))(\pi_2 f_1(s))),\pi_1(F_2(\pi_1 f_1(s))(\pi_2 f_1(s))),\\ d_S(\pi_2(F_1(\pi_1 f_1(s))(\pi_2 f_1(s))),\pi_2(F_2(\pi_1 f_1(s))(\pi_2 f_1(s)))\end{pmatrix}\\
&\leq \sup_{s\in S}\max\begin{pmatrix} \hfill d_I(\pi_1(f_1(s)),\pi_1(f_2(s))),\\ \hfill d_S(\pi_2(f_1(s)),\pi_2(f_2(s))),\\ \sup_{x\in I}\sup_{s'\in S}\max\begin{pmatrix} d_J(\pi_1(F_1(x)(s')),\pi_1(F_2(x)(s')),\\ d_S(\pi_2(F_1(x)(s')),\pi_2(F_2(x)(s'))\end{pmatrix}\end{pmatrix}\\
&= \max\begin{pmatrix} \sup_{s\in S}\max(d_I(\pi_1(f_1(s)),\pi_1(f_2(s))),d_S(\pi_2(f_1(s)),\pi_2(f_2(s)))),\\ \sup_{x\in I}\sup_{s'\in S}\max\begin{pmatrix} d_J(\pi_1(F_1(x)(s')),\pi_1(F_2(x)(s')),\\ d_S(\pi_2(F_1(x)(s')),\pi_2(F_2(x)(s'))\end{pmatrix}\end{pmatrix}\\
&= \max(d_{T_S I}(f_1,f_2),\sup_{x\in I} d^{T_S J}(F_1(x),F_2(x)).
\end{aligned}
$$

We note here that the nonexpansivity of $F_2 \colon (I, d_I) \to (S, d_S) \Rightarrow (S, d_S) \times (J, d_J)$ is equivalent to the one of its uncurrying $\overline{F_2} \colon (S, d_S) \times (I, d_I) \to (S, d_S) \times (J, d_J)$. $\square$

Proof. (Proof of Proposition 6) We first show the Dist_0-unit reflexivity of Δ^{Dist_0}. For $(x_1, x_2) \in \mathsf{Dist}_0(I, d_I)$ (i.e. $d_I(x_1, x_2) = 0$), we calculate

$$\begin{aligned} \Delta^{\mathsf{Dist}_0}_{(I,d_I)}(\eta_I(x_1), \eta_I(x_2)) &= \sup_{d_S(s_1,s_2)=0} \max\left(d_I(\pi_1(x_1, s_2), \pi_1(x_2, s_2)), d_S(\pi_2(x_1, s_1), \pi_2(x_2, s_2)\right) \\ &= \sup_{d_S(s_1,s_2)=0} \max(d_I(x_1, x_2), d_S(s_1, s_2)) = 0. \end{aligned}$$

Next, we show the Dist_0-composability of Δ^{Dist_0}. For any $f_1, f_2 \in T_S(I, d_I)$ and nonexpansive functions $F_1, F_2 \colon (I, d_I) \to T_S(J, d_J)$, we compute

$$\begin{aligned}
&\Delta^{\mathsf{Dist}_0}_J(F_1^\sharp(f_1), F_1^\sharp(f_2)) \\
&= \sup_{d_S(s_1,s_2)=0} \max \begin{pmatrix} d_J(\pi_1(F_1(\pi_1 f_1(s_1))(\pi_2 f_1(s_1))), \pi_1(F_2(\pi_1 f_2(s_2))(\pi_2 f_2(s_2))), \\ d_S(\pi_2(F_1(\pi_1 f_1(s_1))(\pi_2 f_1(s_1))), \pi_2(F_2(\pi_1 f_2(s_2))(\pi_2 f_2(s_2))) \end{pmatrix} \\
&\le \sup_{d_S(s_1,s_2)=0} \max \begin{pmatrix} d_J(\pi_1(F_1(\pi_1 f_1(s_1))(\pi_2 f_1(s_1))), \pi_1(F_2(\pi_1 f_1(s_1))(\pi_2 f_1(s_1))), \\ d_J(\pi_1(F_2(\pi_1 f_1(s_1))(\pi_2 f_1(s_1))), \pi_1(F_2(\pi_1 f_2(s_2))(\pi_2 f_2(s_2))), \\ d_S(\pi_2(F_1(\pi_1 f_1(s_1))(\pi_2 f_1(s_1))), \pi_2(F_2(\pi_1 f_1(s_1))(\pi_2 f_1(s_1))), \\ d_S(\pi_2(F_2(\pi_1 f_1(s_1))(\pi_2 f_1(s_1))), \pi_2(F_2(\pi_1 f_2(s_2))(\pi_2 f_2(s_2))) \end{pmatrix} \\
&= \sup_{d_S(s_1,s_2)=0} \max \begin{pmatrix} d_J(\pi_1(F_2(\pi_1 f_1(s_1))(\pi_2 f_1(s_1))), \pi_1(F_2(\pi_1 f_2(s_2))(\pi_2 f_2(s_2))), \\ d_S(\pi_2(F_2(\pi_1 f_1(s_1))(\pi_2 f_1(s_1))), \pi_2(F_2(\pi_1 f_2(s_2))(\pi_2 f_2(s_2))), \\ d_J(\pi_1(F_1(\pi_1 f_1(s_1))(\pi_2 f_1(s_1))), \pi_1(F_2(\pi_1 f_1(s_1))(\pi_2 f_1(s_1))), \\ d_S(\pi_2(F_1(\pi_1 f_1(s_1))(\pi_2 f_1(s_1))), \pi_2(F_2(\pi_1 f_1(s_1))(\pi_2 f_1(s_1))) \end{pmatrix} \\
&\le \sup_{d_S(s_1,s_2)=0} \max \begin{pmatrix} d_I(\pi_1(f_1(s_1)), \pi_1(f_2(s_2))), \\ d_S(\pi_2(f_1(s_1)), \pi_2(f_2(s_2))), \\ \sup_{(x_1,x_2)\in \mathsf{Dist}_0(I,d_I)} \sup_{d_S(s_1',s_2')=0} \max \begin{pmatrix} d_J(\pi_1(F_1(x)(s_1')), \pi_1(F_2(x)(s_2')), \\ d_S(\pi_2(F_1(x)(s_1')), \pi_2(F_2(x)(s_2')) \end{pmatrix} \end{pmatrix} \\
&= \max \begin{pmatrix} \sup_{d_S(s_1',s_2')=0} \max(d_I(\pi_1(f_1(s_1)), \pi_1(f_2(s_2))), d_S(\pi_2(f_1(s_1)), \pi_2(f_2(s_2)))), \\ \sup_{(x_1,x_2)\in \mathsf{Dist}_0(I,d_I)} \sup_{d_S(s_1',s_2')=0} \max \begin{pmatrix} d_J(\pi_1(F_1(x_1)(s_1')), \pi_1(F_2(x_2)(s_2')), \\ d_S(\pi_2(F_1(x_1)(s_1')), \pi_2(F_2(x_2)(s_2')) \end{pmatrix} \end{pmatrix} \\
&= \max(\Delta^{\mathsf{Dist}_0}_I(f_1, f_2), \sup_{(x_1,x_2)\in \mathsf{Dist}_0(I,d_I)} \Delta^{\mathsf{Dist}_0}_J(F_1(x), F_2(x)).
\end{aligned}$$

This completes the proof. $\square$

Proof. (Proof of Proposition 7) The monotonicity of $\mathsf{C}(\Delta, N)$ is obvious since $M = 1$.

We show the Eq-unit reflexivity of $\mathsf{C}(\Delta, N)$. For all $x \in UI$, we have

$$\begin{aligned} T\pi_1 \bullet (\eta_I^{T(N\times -)} \bullet x) &= (T\pi_1 \circ \eta_I^{T(N\times -)}) \bullet x = (T\pi_1 \circ T\eta_I^{(N\times -)} \circ \eta_I^T) \bullet x \\ &= (T(1_N \circ !_I) \circ \eta_I^T) \bullet x \ = (\eta^T \circ 1_N \circ !_I) \bullet x \\ &= \eta^T \bullet ((1_N \circ !_I) \bullet x) \quad = \eta^T \bullet (1_N \bullet (!_I \bullet x)) \\ &= \eta^T \bullet 1_N \end{aligned}$$

Hence,

$$
\begin{aligned}
\mathsf{C}(\Delta, N)_I(\eta^{T(N\times -)} \bullet x, \eta^{T(N\times -)} \bullet x) &= \mathsf{C}(\Delta, N)_I(\eta^{T(N\times -)} \bullet x, \eta^{T(N\times -)} \bullet x) \\
&= \Delta_N(T\pi_1 \bullet (\eta^{T(N\times -)} \bullet x), T\pi_1 \bullet (\eta^{T(N\times -)} \bullet x) \\
&= \Delta_N(\eta^T \bullet 1_N, \eta^T \bullet 1_N) \\
&\leq 0_{\mathcal{Q}}.
\end{aligned}
$$

We next show the Eq-composability of $\mathsf{C}(\Delta, N)$. For any $f\colon I \to T(N \times I)$, we define $h_f\colon N \times I \to T(N)$ by $h_f = T(\star) \circ \theta_{N,N} \circ (N \times (T\pi_1 \circ f))$. Then, we have $T\pi_1 \bullet f^{\sharp(T(N\times I))} \bullet \nu = h_f^{\sharp T} \bullet \nu$ for any $\nu \in U(T(N \times I))$. First, for all $m, n \in UN$, we have

$$
\begin{aligned}
(T(\star) \circ (\eta_{N\times N})_n) \bullet m = T(\star) \bullet (\eta_{N\times N} \bullet \langle n, m\rangle) &= (T(\star) \circ \eta_{N\times N}) \bullet \langle n, m\rangle) \\
= (\eta_N \circ \star) \bullet \langle n, m\rangle) \qquad &= \eta_N \bullet (\star \bullet \langle n, m\rangle) \\
= \eta_N \bullet (\star_n \bullet m) \qquad &= (\eta_N \circ \star_n) \bullet m.
\end{aligned}
$$

From this and the equality (1), we can calculate as follows:

$$
\begin{aligned}
& h_f \bullet \langle n, i\rangle \\
&= (T(\star) \circ \theta_{N,N} \circ (N \times (T\pi_1 \circ f))) \bullet \langle n, i\rangle && = (T(\star) \circ \theta_{N,N}) \bullet ((N \times (T\pi_1 \circ f)) \bullet \langle n, i\rangle) \\
&= (T(\star) \circ \theta_{N,N}) \bullet (U(N \times (T\pi_1 \circ f))(n, i)) && = (T(\star) \circ \theta_{N,N}) \bullet ((U(N) \times U(T\pi_1 \circ f))(n, i)) \\
&= (T(\star) \circ \theta_{N,N}) \bullet \langle U(N)(n), U(T\pi_1 \circ f)(i)\rangle && = (T(\star) \circ \theta_{N,N}) \bullet \langle n, (T\pi_1 \circ f) \bullet i\rangle \\
&= T(\star) \bullet (\theta_{N,N} \bullet \langle n, (T\pi_1 \circ f) \bullet i\rangle) && = T(\star) \bullet ((\theta_{N,N})_n \bullet ((T\pi_1 \circ f) \bullet i)) \\
&= T(\star) \bullet (((\eta_{N\times N})_n)^\sharp \bullet ((T\pi_1 \circ f) \bullet i)) && = (T(\star) \circ ((\eta_{N\times N})_n)^\sharp) \bullet ((T\pi_1 \circ f) \bullet i) \\
&= (T(\star) \circ (\eta_{N\times N})_n)^\sharp \bullet ((T\pi_1 \circ f) \bullet i) && = (\eta_N \circ (\star_n))^\sharp \bullet ((T\pi_1 \circ f) \bullet i).
\end{aligned}
$$

From the assumption $\Delta_{N\times I}(c_1, c_2) \leq \mathsf{C}(\Delta, N)_I(c_1, c_2)$, the Eq-unit-reflexivity and Eq-composability of the original divergence Δ, we obtain the Eq-composability of $\mathsf{C}(\Delta, N)$ as follows:

$$
\begin{aligned}
& \mathsf{C}(\Delta, N)_J(f_1^{\sharp(T(N\times I))} \bullet c_1, f_2^{\sharp(T(N\times I))} \bullet c_2) \\
&= \Delta_N(T\pi_1 \bullet f_1^{\sharp(T(N\times I))} \bullet c_1, T\pi_1 \bullet f_2^{\sharp(T(N\times I))} \bullet c_2) \\
&= \Delta_N(h_{f_1}^{\sharp T} \bullet c_1, h_{f_2}^{\sharp T} \bullet c_2) \\
&\leq \Delta_{N\times I}(c_1, c_2) + \sup_{\langle n,i\rangle \in U(N\times I)} \Delta_N(h_{f_1} \bullet \langle n, i\rangle, h_{f_2} \bullet \langle n, i\rangle) \\
&= \Delta_{N\times I}(c_1, c_2) \\
&\qquad + \sup_{\langle n,i\rangle \in U(N\times I)} \Delta_N((\eta_N \circ (\star)_n)^{\sharp T} \bullet ((T\pi_1 \circ f) \bullet i), (\eta_N \circ (\star)_n)^{\sharp T} \bullet ((T\pi_1 \circ f) \bullet i)) \\
&\leq \Delta_{N\times I}(c_1, c_2) \\
&\qquad + \sup_{\langle n,i\rangle \in U(N\times I)} \begin{pmatrix} \Delta_N((T\pi_1 \circ f) \bullet i, (T\pi_1 \circ f) \bullet i) \\ \qquad + \sup_{m \in UN} \Delta_N(\eta_N \circ (\star)_n) \bullet m, (\eta_N \circ (\star)_n) \bullet m) \end{pmatrix} \\
&\leq \Delta_{N\times I}(c_1, c_2) + \sup_{\langle n,i\rangle \in U(N\times I)} \Delta_N((T\pi_1 \circ f) \bullet i, (T\pi_1 \circ f) \bullet i) \\
&= \Delta_{N\times I}(c_1, c_2) + \sup_{i \in UI} \Delta_N((T\pi_1 \circ f_1) \bullet i, (T\pi_1 \circ f_2) \bullet i) \\
&\leq \mathsf{C}(\Delta, N)_I(c_1, c_2) + \sup_{i \in UI} \mathsf{C}(\Delta, N)_J(f_1 \bullet i, f_2 \bullet i).
\end{aligned}
$$

This completes the proof. □

Proof. (Proof of Proposition 8)

We consider a preorder $\sqsubseteq$ on a monad T. We define the $\mathcal{B}$-divergence $\Delta^{\sqsubseteq}$ on TI by

$$\Delta^{\sqsubseteq}_I(c_1, c_2) \triangleq \begin{cases} 0 & c_1 \not\sqsubseteq_I c_2 \\ 1 & c_1 \sqsubseteq_I c_2 \end{cases}$$

Each $\tilde{\Delta}(1)I$ is a preorder because $\tilde{\Delta}(1)I = \sqsubseteq_I$ holds for each I.

The Eq-unit reflexivity of $\Delta^{\sqsubseteq}$ is derived from the reflexivity of $\sqsubseteq$. For all set I and $c \in TI$,

$$(\Delta^{\sqsubseteq}_I(c, c) \leq 1) \iff (c \sqsubseteq_I c).$$

Since $\sqsubseteq$ is a preorder on T, for all set I, J, $c_1, c_2 \in TI$ and $f, g \colon I \to TJ$,

$$\begin{aligned}
&(\Delta^{\sqsubseteq}_I(c_1, c_2) \times \sup_{x \in I} \Delta^{\sqsubseteq}_J(f(x), g(x))) = 1 \\
\iff& (\Delta^{\sqsubseteq}_I(c_1, c_2) = 1) \wedge (\sup_{x \in I} \Delta^{\sqsubseteq}_J(f(x), g(x)) = 1) \\
\iff& (c_1 \sqsubseteq_I c_2) \wedge (\forall x \in I.\ f(x) \sqsubseteq_J g(x)) \\
\implies& (f^\sharp(c_1) \sqsubseteq_J f^\sharp(c_2)) \wedge (f^\sharp(c_2) \sqsubseteq_J g^\sharp(c_2)) \\
\implies& (f^\sharp(c_1) \sqsubseteq_J g^\sharp(c_2)) \\
\iff& (\Delta^{\sqsubseteq}_J(f^\sharp(c_1), g^\sharp(c_2)) = 1)
\end{aligned}$$

Hence, we have the Eq-composability

$$\Delta^{\sqsubseteq}_J(f^\sharp(c_1), g^\sharp(c_2)) \leq \Delta^{\sqsubseteq}_I(c_1, c_2) \times \sup_{x \in I} \Delta^{\sqsubseteq}_J(f(x), g(x)).$$

Conversely, we consider an Eq-relative $\mathcal{B}$-divergence Δ on T such that each $\tilde{\Delta}(1)I$ is a preorder. We show that the family $\sqsubseteq^\Delta = \{\sqsubseteq^\Delta_I\}_{I \in \mathbf{Set}}$ defined by $\sqsubseteq^\Delta_I \triangleq \tilde{\Delta}(1)I$ forms a preorder on monad T.

Each component $\sqsubseteq^\Delta_I$ of $\sqsubseteq^\Delta$ at set I is a preorder on the set TI. We here note that the divergence Δ must be reflexive (i.e. $\Delta_I(c, c) \leq 1$ for all $I \in \mathbf{Set}, c \in TI$) because of the reflexivity of $\sqsubseteq^\Delta_I$:

$$(\Delta_I(c, c) \leq 1) \iff (c \sqsubseteq^\Delta_I c), \quad \text{for all } I \in \mathbf{Set}, c \in TI.$$

From the reflexivity and Eq-composability of Δ, we have for all $c_1, c_2, c \in TI$ and $f, g \colon I \to TJ$,

$$\forall c_1, c_2 \in TI, f \colon I \to TJ\ .\ \Delta_J(f^\sharp(c_1), f^\sharp(c_2)) \leq \Delta_I(c_1, c_2), \tag{26}$$

$$\forall c \in TI, f, g \colon I \to TJ\ .\ \Delta_J(f^\sharp(c), g^\sharp(c)) \leq \sup_{x \in I} \Delta_J(f(x), g(x)). \tag{27}$$

They are equivalent to the substitutivity and congruence of $\sqsubseteq^\Delta$ respectively:

$$\begin{aligned}
(26) &\iff \forall c_1, c_2 \in TI, f \colon I \to TJ\ .\ (c_1 \sqsubseteq^\Delta_I c_2 \implies f^\sharp(c_1) \sqsubseteq^\Delta_J f^\sharp(c_2)), \\
(27) &\iff \forall c \in TI, f, g \colon I \to TJ\ .\ (\forall x \in I.\ f(x) \sqsubseteq^\Delta_J g(x) \implies f^\sharp(c) \sqsubseteq^\Delta_J g^\sharp(c)).
\end{aligned}$$

Finally, the above conversions $\Delta^{(-)}$ and $\sqsubseteq^{(-)}$ are mutually inverse:

$$\begin{aligned}
\Delta^{\sqsubseteq^{\Delta'}}_I(c_1, c_2) \leq 1 &\iff c_1 \sqsubseteq^{\Delta'}_I c_2 \iff \Delta'_I(c_1, c_2) \leq 1, \\
c_1 \sqsubseteq^{\Delta^{\sqsubseteq'}}_I c_2 &\iff \Delta^{\sqsubseteq'}_I(c_1, c_2) \leq 1 \iff c_1 \sqsubseteq'_I c_2.
\end{aligned}$$

This completes the proof. $\square$

B Proofs for Section 6 (Properties of Divergences on Monads)

Proof. (Proof of Theorem 2) First, it is easy to see that the inequality (3) is equivalent to Δ satisfying E-unit reflexivity.

We next show that the inequality (4) is equivalent to Δ satisfying E-composability.

(only if) Since $U1 = \{\mathrm{id}_1\}$, we have $R_{E1} = \{(\mathrm{id}_1, \mathrm{id}_1)\}$. Therefore it holds $d_{1,J}(c_1, c_2) = \Delta_J(c_1, c_2)$. By letting $I = 1$ in the inequality (4), we obtain the E-composability:

$$\begin{aligned} & d_{1,K}(f_1 \circ_{\mathbb{C}_T} c_1, f_2 \circ_{\mathbb{C}_T} c_2) \leq d_{J,K}(f_1, f_2) + d_{1,J}(c_1, c_2) \\ \iff \quad & \Delta_K(f_1^\sharp \circ c_1, f_2^\sharp \circ c_2) \leq \sup_{(x_1,x_2) \in EI} \Delta_K(f_1 \bullet x_1, f_2 \bullet x_2) + \Delta_J(c_1, c_2). \end{aligned}$$

(if) From the E-composability, for any $f_1, f_2 : I \to TJ$ and $g_1, g_2 : J \to TK$ and $(x_1, x_2) \in EI$, we have

$$\Delta_K(g_1^\sharp \bullet (f_1 \bullet x_1), g_2^\sharp \bullet (f_2 \bullet x_2)) \leq d_{J,K}(g_1, g_2) + \Delta_J(f_1 \bullet x_1, f_2 \bullet x_2).$$

Next, for any $(x_1, x_2) \in EI$, we have $\Delta_J(f_1 \bullet x_1, f_2 \bullet x_2) \leq d_{I,J}(f_1, f_2)$. Thus by monotonicity of $(+)$ we have

$$\Delta_K(g_1^\sharp \bullet f_1 \bullet x_1, g_2^\sharp \bullet f_2 \bullet x_2) \leq d_{J,K}(g_1, g_2) + d_{I,J}(f_1, f_2).$$

By discharging $(x_1, x_2) \in EI$, we conclude

$$d_{I,K}(g_1^\sharp \circ f_1, g_2^\sharp \circ f_2) \leq d_{J,K}(g_1, g_2) + d_{I,J}(f_1, f_2).$$

□

Proof. (Proof of Theorem 4) $[\Delta]$ is a graded variant of *codensity lifting* performed along the fibration $V_{\mathcal{Q},\mathbb{C}} : \mathbf{Div}_{\mathcal{Q}}(\mathbb{C}) \to \mathbb{C}$ ([33]; see also Definition 15). Proving that it is a graded lifting of T is routine. We show $\Delta^m_I = [\Delta]m(E'I)$. The direction $[\Delta]m(E'I) \leq \Delta^m_I$ is easy. We show the converse. From the composability of Δ, for any $c_1, c_2 \in U(TI)$, $J \in \mathbb{C}$, $n \in M$ and $f \in \mathbf{Div}_{\mathcal{Q}}(\mathbb{C})(E'I, \Delta^n_J)$, we have

$$\Delta^{m \cdot n}_J(f^\sharp \bullet c_1, f^\sharp \bullet c_2) \leq \Delta^m_I(c_1, c_2) + \sup_{(x_1,x_2) \in EI} \Delta^n_J(f \bullet x_1, f \bullet x_2).$$

Next, the nonexpansivity of f is equivalent to

$$\sup_{(x_1,x_2) \in EI} \Delta^n_J(f \bullet x_1, f \bullet x_2) \leq 0.$$

Therefore we conclude $\Delta^{m \cdot n}_J(f^\sharp \bullet c_1, f^\sharp \bullet c_2) \leq \Delta^m_I(c_1, c_2)$. By discharging J, n, f, we conclude the inequality $[\Delta]m(EI) \leq \Delta^m_I$. □

Proof. (Proof of Theorem 5) Let $\Delta \in \mathbf{Div}(T, E, M, \mathcal{Q})$. We have already shown that $[\Delta]$ is an M-graded $\mathcal{Q}$-divergence lifting of T. We show that $[\Delta]$ is E-strong (this proof does not need the closedness of $\mathbb{C}$). Let $X \triangleq (I, d) \in \mathbf{Div}_{\mathcal{Q}}(\mathbb{C})$ and $J \in \mathbb{C}$ be objects. We first rewrite the goal:

$$\begin{aligned} & \theta \in \mathbf{Div}_{\mathcal{Q}}(\mathbb{C})(X \otimes [\Delta]m(E'J), [\Delta]m(X \otimes E'J)) \\ \iff & \begin{pmatrix} \forall x_1, x_2 \in UI, c_1, c_2 \in U(TJ) \ . \\ \quad d_{[\Delta]m(X \otimes E'J)}(\theta \bullet \langle x_1, c_1 \rangle, \theta \bullet \langle x_2, c_2 \rangle) \leq d(x_1, x_2) + d_{[\Delta]m(E'J)}(c_1, c_2) \end{pmatrix} \end{aligned}$$

$$\iff \begin{pmatrix} \forall x_1, x_2 \in UI, c_1, c_2 \in U(TJ), K \in \mathbb{C}, n \in M, f \in \mathbf{Div}_{\mathcal{Q}}(\mathbb{C})(X \otimes E'J, \Delta^n_K) \ . \\ \Delta^{m \cdot n}_K (f^\sharp \bullet \theta \bullet \langle x_1, c_1 \rangle, f^\sharp \bullet \theta \bullet \langle x_2, c_2 \rangle) \le d(x_1, x_2) + d_{[\Delta] m(E'J)}(c_1, c_2) \end{pmatrix}$$

$$\overset{\dagger}{\iff} \begin{pmatrix} \forall x_1, x_2 \in UI, c_1, c_2 \in U(TJ), K \in \mathbb{C}, n \in M, f \in \mathbf{Div}_{\mathcal{Q}}(\mathbb{C})(X \otimes E'J, \Delta^n_K) \ . \\ \Delta^{m \cdot n}_K ((f_{x_1})^\sharp \bullet c_1, (f_{x_2})^\sharp \bullet c_2) \le d(x_1, x_2) + d^m_{[\Delta]}(c_1, c_2) \end{pmatrix}.$$

In the step $\overset{\dagger}{\iff}$, we used the equality (1). To show this goal, we proceed as follows. Let $x_1, x_2 \in UI, c_1, c_2 \in U(TJ), K \in \mathbb{C}, n \in M$ and $f \in \mathbf{Div}_{\mathcal{Q}}(\mathbb{C})(X \otimes E'J, \Delta^n_K)$. First, from the composability of Δ, we obtain

$$\Delta^{m \cdot n}_K ((f_{x_1})^\sharp \bullet c_1, (f_{x_2})^\sharp \bullet c_2) \le \Delta^m_J (c_1, c_2) + \sup_{(y_1, y_2) \in EJ} \Delta^n_K (f_{x_1} \bullet y_1, f_{x_2} \bullet y_2).$$

We look at summands of the right hand side. First, we easily obtain $\Delta^m_J(c_1, c_2) \le d_{[\Delta] m(E'J)}(c_1, c_2)$. Next, from the nonexpansivity of f, for any $x_1, x_2 \in UI, y_1, y_2 \in UJ$, we have

$$\Delta^n_K (f_{x_1} \bullet y_1, f_{x_2} \bullet y_2) = \Delta^n_K (f \bullet \langle x_1, y_1 \rangle, f \bullet \langle x_2, y_2 \rangle) \le d(x_1, x_2) + E'J(y_1, y_2).$$

Because $x + \top = \top$, we obtain

$$\forall x_1, x_2 \in UI \ . \sup_{(y_1, y_2) \in EJ} \Delta^n_K (f_{x_1} \bullet y_1, f_{x_2} \bullet y_2) \le d(x_1, x_2).$$

Therefore we obtain the goal:

$$\Delta^{m \cdot n}_K ((f_{x_1})^\sharp \bullet c_1, (f_{x_2})^\sharp \bullet c_2) \le d_{[\Delta] m(E'J)}(c_1, c_2) + d(x_1, x_2) = d(x_1, x_2) + d_{[\Delta] m(E'J)}(c_1, c_2).$$

Next, let $\dot{T} \in \mathbf{SGDLift}(T, E, M, \mathcal{Q})$. We show that $\langle \dot{T} \rangle \in \mathbf{Div}(T, E, M, \mathcal{Q})$.
The unit law of $\dot{T}$ immediately entails

$$\eta_I \in \mathbf{Div}_{\mathcal{Q}}(\mathbb{C})(E'I, \dot{T}1(E'I)).$$

Next, under the assumption on $(\mathbb{C}, T)$ and $\mathcal{Q}$, in $\mathbf{Div}_{\mathcal{Q}}(\mathbb{C})$ the functor $(-) \otimes E'I$ has a right adjoint $E'I \multimap (-)$ above the adjunction $(-) \times I \dashv I \Rightarrow (-)$ (Lemma 1). Therefore each component of the internal Kleisli extension morphism kl given in (6) are nonexpansive morphisms in $\mathbf{Div}_{\mathcal{Q}}(\mathbb{C})$:

$$\begin{array}{c} \langle \dot{T} \rangle m(E'I) \otimes (E'I \multimap \langle \dot{T} \rangle n(E'J)) \\ \big\downarrow{\scriptstyle \langle \pi_2, \pi_1 \rangle} \\ (E'I \multimap \langle \dot{T} \rangle n(E'J)) \otimes \langle \dot{T} \rangle m(E'I) \\ \big\downarrow{\scriptstyle \theta} \\ \langle \dot{T} \rangle m((E'I \multimap \langle \dot{T} \rangle n(E'J)) \otimes E'I) \\ \big\downarrow{\scriptstyle \mathrm{ev}^\sharp} \\ \langle \dot{T} \rangle (m \cdot n)(E'J) \end{array}$$

Therefore we conclude

$$\mathrm{kl} \in \mathbf{Div}_{\mathcal{Q}}(\mathbb{C})(\langle \dot{T} \rangle m(E'I) \otimes (E'I \multimap \langle \dot{T} \rangle n(E'J)), \langle \dot{T} \rangle (m \cdot n)(E'J)).$$

We also easily have monotonicity: $\langle\dot{T}\rangle m(E'I) \leq \langle\dot{T}\rangle n(E'I)$ for $m \leq n$ by condition 1 of graded divergence lifting. We thus conclude that $\langle\dot{T}\rangle mE'I \in \mathbf{Div}(T, E, M, \mathcal{Q})$.

We finally show $\dot{T} \preceq [\langle\dot{T}\rangle]$. Let $c_1, c_2 \in U(TI)$. We show

$$\sup_{n\in M, J\in\mathbb{C}, f\in\mathbf{Div}_{\mathcal{Q}}(\mathbb{C})(X,\dot{T}n(E'J))} d_{\dot{T}(m\cdot n)(E'J)}(f^\sharp(c_1), f^\sharp(c_2)) \leq d_{\dot{T}mX}(c_1, c_2). \tag{28}$$

Let $n \in M, J \in \mathbb{C}, f \in \mathbf{Div}_{\mathcal{Q}}(\mathbb{C})(X, \dot{T}n(E'J))$. Since $\dot{T}$ is an M-graded $\mathcal{Q}$-divergence lifting of T, we obtain

$$f^\sharp \in \mathbf{Div}_{\mathcal{Q}}(\mathbb{C})(\dot{T}mX, \dot{T}(m\cdot n)(E'J)).$$

This implies the inequality $d_{\dot{T}(m\cdot n)(E'J)}(f^\sharp(c_1), f^\sharp(c_2)) \leq d_{\dot{T}mX}(c_1, c_2)$ in $\mathcal{Q}$. By taking the sup for n, J, f, we obtain the inequality (28). □

Proof. (Proof of Proposition 9) We write $|1| = \{*\}$. We first check the measurable isomorphism $G_s1 \cong [0,1]$. The measurable functions $\mathrm{ev}_{\{*\}} \colon G_s1 \to [0,1]$ ($\nu \mapsto \nu(*)$) and the function $H \colon |[0,1]| \to |G_s1|$ ($r \mapsto r \cdot \mathbf{d}_*$) are mutually inverse. For any (Borel-)measurable $U \in \Sigma_{[0,1]}$, we have $H^{-1}(\mathrm{ev}_{\{*\}}^{-1}(U)) = U$ and $H^{-1}(\mathrm{ev}_{\emptyset}^{-1}(U)) = [0,1]$ if $0 \in U$ and $H^{-1}(\mathrm{ev}_{\emptyset}^{-1}(U)) = \emptyset$ otherwise. Since all generators of Σ_{G_s1} are $\mathrm{ev}_{\{*\}}^{-1}(U)$ and $\mathrm{ev}_{\emptyset}^{-1}(U)$ where $U \in \Sigma_{[0,1]}$, we conclude the measurability of H. Thus, $f \colon I \to [0,1]$ corresponds bijectively to $H \circ f \colon I \to G_s1$, and

$$\int_I f d\nu_1 = \int_I \mathrm{ev}_{\{*\}} \circ H \circ f d\nu_1 = ((H \circ f)^\sharp \nu_1)(\{*\}).$$

We then obtain, for all $I \in \mathbf{Meas}$, $\nu_1, \nu_2 \in G_sI$

$$\begin{aligned}
\mathsf{DP}^\varepsilon_I(\nu_1, \nu_2) &= \sup_{S\in\Sigma_I} (\nu_1(S) - \exp(\varepsilon)\nu_2(S)) \\
&\leq \sup_{f_S \colon I\to[0,1]} (\int_I f_S d\nu_1 - \exp(\varepsilon) \int_I f_S d\nu_2) \\
&= \sup_{f_S \colon I\to G_s[0,1]} (((H \circ f_S)^\sharp \nu_1)(*) - \exp(\varepsilon)((H \circ f_S)^\sharp \nu_2)(*)) \\
&\leq \sup_{f_S \colon I\to G_s[0,1]} \sup_{S'\in\Sigma_1(\iff S'=\{*\},\emptyset)} ((H \circ f_S)^\sharp \nu_1)(S') - \exp(\varepsilon)((H \circ f_S)^\sharp \nu_2)(S')) \\
&= \sup_{f_S \colon I\to G_s[0,1]} \mathsf{DP}^\varepsilon_1((H \circ f_S)^\sharp \nu_1, (H \circ f_S)^\sharp \nu_2) \\
&= \sup_{g \colon I\to G_s1} \mathsf{DP}^\varepsilon_1(g^\sharp \nu_1, g^\sharp \nu_2) \\
&\leq \mathsf{DP}^\varepsilon_I(\nu_1, \nu_2).
\end{aligned}$$

The first inequality is given by $\nu(S) = \int_I \chi_S d\nu$ where $\chi_S \colon I \to [0,1]$ is the indicator function of S defined by $\chi_S(x) = 1$ when $x \in S$ and $\chi_S(x) = 0$ otherwise. The last inequality is given by the data-processing inequality which is given by the reflexivity and Eq-composability of DP. □

Proof. (Proof of Proposition 10) We first prove that TV is not 1-generated. We write $|2| = \{0, 1\}$. We define $\nu_1, \nu_2 \in G_s2$ by

$$\nu_1 = \frac{1}{2} \cdot \mathbf{d}_0 + \frac{1}{2} \cdot \mathbf{d}_1, \qquad \nu_2 = \frac{1}{3} \cdot \mathbf{d}_0 + \frac{2}{3} \cdot \mathbf{d}_1.$$

Then the total variation distance between them is calculated by

$$\mathsf{TV}_2(\nu_1, \nu_2) = \frac{1}{2}\left(\left|\frac{1}{2} - \frac{1}{3}\right| + \left|\frac{1}{2} - \frac{2}{3}\right|\right) = \frac{1}{6}.$$

On the other hand, for any $f\colon 2 \to G_s 1$, we have

$$\begin{aligned}
\mathsf{TV}_1(f^\sharp(\nu_1), f^\sharp(\nu_2)) &= \frac{1}{2}\left|\frac{1}{2}f(0) + \frac{1}{2}f(1) - \frac{1}{3}f(0) - \frac{2}{3}f(1)\right| \\
&= \frac{1}{2}\left|\frac{1}{6}f(0) - \frac{1}{6}f(1)\right| \\
&= \frac{1}{12}\left|f(0) - f(1)\right| \\
&\leq \frac{1}{12}.
\end{aligned}$$

This implies that TV is not 1-generated.

Next, we prove that TV is 2-generated. From the data-processing inequality TV which is given by the reflexivity and Eq-composability of TV, we obtain for any $\nu_1, \nu_2 \in G_s I$,

$$\mathsf{TV}_I(\nu_1, \nu_2) \geq \sup_{g\colon I \to G_s 2} \mathsf{TV}_2(g^\sharp\nu_1, g^\sharp\nu_2).$$

We show that the above inequality becomes the equality for some g.

We fix $\nu_1, \nu_2 \in G_s I$, a base measure μ over I satisfying the absolute continuity $\nu_1, \nu_2 \ll \mu$ and the Radon-Nikodym derivatives (density functions) $\frac{d\nu_1}{d\mu}, \frac{d\nu_2}{d\mu}$ of ν_1, ν_2 with respect to μ respectively.

Let $A = (\frac{d\nu_1}{d\mu} - \frac{d\nu_2}{d\mu})^{-1}([0,\infty))$ and $B = I \setminus A$. We define $g\colon I \to G_s 2$ by $g(x) = \mathbf{d}_0$ if $x \in B$ and $g(x) = \mathbf{d}_1$ otherwise. Then for any $\nu \in G_s I$ we have

$$(g^\sharp\nu)(\{0\}) = \int_I g(-)(\{0\}) d\nu = \int_A g(-)(\{0\}) d\nu + \int_B g(-)(\{0\}) d\nu = \int_A 1 d\nu + \int_B 0 d\nu = \nu(A).$$

Similarly we have $(g^\sharp\nu)(\{1\}) = \nu(B)$. Therefore, we obtain

$$\begin{aligned}
\frac{1}{2}\mathsf{TV}_I(\mu_1, \mu_2) &= \frac{1}{2}\int_I \left|\frac{d\nu_1}{d\mu}(x) - \frac{d\nu_2}{d\mu}(x)\right| \, d\mu(x) \\
&= \frac{1}{2}\int_A \frac{d\nu_1}{d\mu}(x) - \frac{d\nu_2}{d\mu}(x) \, d\mu(x) + \frac{1}{2}\int_B \frac{d\nu_2}{d\mu}(x) - \frac{d\nu_1}{d\mu}(x) \, d\mu(x) \\
&= \frac{1}{2}(\nu_1(A) - \nu_2(A) + \nu_2(B) - \nu_1(B)) \\
&= \frac{1}{2}((g^\sharp\nu_1)(\{0\}) - (g^\sharp\nu_2)(\{0\}) + (g^\sharp\nu_2)(\{1\}) - (g^\sharp\nu_2)(\{1\})) \\
&= \frac{1}{2}(|(g^\sharp\nu_1)(\{0\}) - (g^\sharp\nu_2)(\{0\})| + |(g^\sharp\nu_2)(\{1\}) - (g^\sharp\nu_2)(\{1\})|) \\
&= \mathsf{TV}_2(g^\sharp(\mu_1), g^\sharp(\mu_2))
\end{aligned}$$

We then conclude that Δ^{TV} is 2-generated. $\square$

Proof. (Proof of Proposition 11) For all set J and $c_1, c_2 \in TJ$, we have

$$\begin{aligned}
\Delta_J^{[\leq]^\Omega}(c_1, c_2) = 1 &\iff c_1 [\leq]_J^\Omega c_2 \\
&\iff \bigwedge_{g\colon J \to T\Omega} g^\sharp(c_1) \leq g^\sharp(c_2) \\
&\iff \bigwedge_{g\colon J \to T\Omega} g^\sharp(c_1) \, [\leq]_\Omega^\Omega \, g^\sharp(c_2)
\end{aligned}$$

$$\iff \sup_{g\colon J\to T\Omega} \Delta([\leq]^{\Omega})_{\Omega}(g^{\sharp}(c_1), g^{\sharp}(c_2)) = 1.$$

This implies that $\Delta^{[\leq]^{\Omega}}$ is Ω-generated. $\square$

Lemma 2. *For any $U \in \mathbf{QET}(\Omega, X)$, the function $d[U]\colon (T_\Omega X)^2 \to \mathcal{R}^+$ defined by*

$$d[U](t,u) \triangleq \inf \left\{ \varepsilon \in \mathbb{Q}^+ \;\middle|\; \emptyset \vdash t =_\varepsilon u \in U \right\}$$

is a CS-EPMet on $T_\Omega X$ such that $d[U](t,u) \in \mathbb{Q}^+ \implies \emptyset \vdash t =_{d[U](t,u)} u$.

Proof. We first check the axioms of extended pseudometric.

By (Ref), U contains $\emptyset \vdash t =_0 t$ for each $t \in T_\Omega X$. Hence $d[U](t,t) = 0$ holds for all $t \in T_\Omega X$.

By (Sym) and (Cut), $\emptyset \vdash t =_\varepsilon u$ if and only if $\emptyset \vdash u =_\varepsilon t$. Hence, for all $t, u \in T_\Omega X$,

$$d[U](t,u) = \inf \left\{ \varepsilon \in \mathbb{Q}^+ \;\middle|\; \emptyset \vdash t =_\varepsilon u \right\} = \inf \left\{ \varepsilon \in \mathbb{Q}^+ \;\middle|\; \emptyset \vdash u =_\varepsilon t \right\} = d[U](u,t)$$

By (Tri) and (Cut), if $\emptyset \vdash t =_\varepsilon u$ and $\emptyset \vdash u =_{\varepsilon'} v$ then $\emptyset \vdash t =_{\varepsilon+\varepsilon'} v$. Hence, for all $t, u, v \in T_\Omega X$,

$$\begin{aligned} d[U](t,v) &= \inf \left\{ \varepsilon^* \in \mathbb{Q}^+ \;\middle|\; \emptyset \vdash t =_{\varepsilon^*} v \right\} \\ &\leq \inf \{ \varepsilon + \varepsilon' \mid \emptyset \vdash t =_\varepsilon u \wedge \emptyset \vdash u =_{\varepsilon'} v \} \\ &\leq \inf \left\{ \varepsilon \in \mathbb{Q}^+ \;\middle|\; \emptyset \vdash t =_\varepsilon u \right\} + \inf \left\{ \varepsilon' \in \mathbb{Q}^+ \;\middle|\; \emptyset \vdash u =_{\varepsilon'} v \right\} \\ &= d[U](t,u) + d[U](u,v). \end{aligned}$$

We next check the substitutivity. Let $t, u \in T_\Omega X$ and $h\colon X \to T_\Omega X$. By (Subst), we have

$$\emptyset \vdash t =_\varepsilon u \in U \implies \emptyset \vdash h^\sharp(t) =_\varepsilon h^\sharp(u) \in U.$$

Since ε is arbitrary, we conclude the substitutivity as follows:

$$\begin{aligned} d[U](h^\sharp(t), h^\sharp(u)) &= \inf \left\{ \varepsilon \in \mathbb{Q}^+ \;\middle|\; \emptyset \vdash h^\sharp(t) =_\varepsilon h^\sharp(u) \in U \right\} \\ &\leq \inf \left\{ \varepsilon \in \mathbb{Q}^+ \;\middle|\; \emptyset \vdash t =_\varepsilon u \in U \right\} \\ &= d[U](t,u). \end{aligned}$$

Next, we check the congruence, Let $t \in T_\Omega I$ and $h_1, h_2\colon I \to T_\Omega X$ By applying (Nonexp) and (Cut) inductively by unfolding the structure of t,

$$\forall i \in I \;.\; \emptyset \vdash h_1(i) =_\varepsilon h_2(i) \in U \implies \emptyset \vdash h_1^\sharp(t) =_\varepsilon h_2^\sharp(t) \in U. \tag{29}$$

If $\sup_{i\in I} d[U](h_1(i), h_2(i)) \leq \varepsilon'$ for some $\varepsilon' \in \mathbb{Q}^+$, then we have $d[U](h_1(i), h_2(i)) \leq \varepsilon'$ for all $i \in I$. By (Max),(Cut) and definition of d_X^U, we have $\vdash h_1(i) =_{\varepsilon'} h_2(i) \in U$ for all $i \in I$. Hence,

$$\sup_{i\in I} d[U](h_1(i), h_2(i)) \leq \varepsilon' \implies \forall i \in I \;.\; \emptyset \vdash h_1(i) =_{\varepsilon'} h_2(i) \in U. \tag{30}$$

From the above two implications (29) and (30), We conclude the congruence as follows:

$$\begin{aligned} d[U](h_1^\sharp(t), h_2^\sharp(t)) &= \inf \left\{ \varepsilon' \in \mathbb{Q}^+ \;\middle|\; \emptyset \vdash h_1^\sharp(t) =_{\varepsilon'} h_2^\sharp(t) \in U \right\} \\ &\leq \inf \left\{ \varepsilon' \in \mathbb{Q}^+ \;\middle|\; \forall i \in I \;.\; \emptyset \vdash h_1(i) =_{\varepsilon'} h_2(i) \in U \right\} \\ &\leq \inf \left\{ \varepsilon' \in \mathbb{Q}^+ \;\middle|\; \sup_{i\in I} d_X^U(h_1(i), h_2(i)) \leq \varepsilon' \right\} \end{aligned}$$

$$= \sup_{i \in I} d[U](h_1(i), h_2(i)).$$

Finally, we assume $d[U](t,u) \in \mathbb{Q}^+$. By definition of $d[U](t,u)$, for any $\varepsilon \in \mathbb{Q}^+$ such that $d[U](t,u) < \varepsilon$, there is $\varepsilon' \in \mathbb{Q}^+$ satisfying $d[U](t,u) \leq \varepsilon' < \varepsilon$ and $t =_{\varepsilon'} u \in U$. Since $\varepsilon \in \mathbb{Q}^+$ is arbitrary, by (Max) and (Cut), we conclude

$$\forall \varepsilon \in \mathbb{Q}^+ \ . \ (d[U](t,u) < \varepsilon \implies t =_\varepsilon u \in U).$$

Since $d[U](t,u) \in \mathbb{Q}^+$, by (Arch) and (Cut), we have $t =_{d[U](t,u)} u \in U$. ☐

The monotonicity of $d[-]\colon (\mathbf{QET}(\Omega, X), \subseteq) \to (\mathbf{CSEPMet}(T_\Omega, X), \preceq)$ is easy to prove:

$$\begin{aligned}
U \subseteq V \implies & \forall t, u \in T_\Omega X \ . \ \inf \left\{ \varepsilon \in \mathbb{Q}^+ \;\middle|\; \emptyset \vdash t =_\varepsilon u \in U \right\} \geq \inf \left\{ \varepsilon \in \mathbb{Q}^+ \;\middle|\; \emptyset \vdash t =_\varepsilon u \in V \right\} \\
\iff & \forall t, u \in T_\Omega X \ . \ d[U](t,u) \geq d[V](t,u) \\
\iff & d[U] \preceq d[V].
\end{aligned}$$

Lemma 3. *Let T be a monad on* **Set***, and let $X \in$* **Set***. For any $d \in \mathbf{CSEPMet}(T, X)$, the family $\mathrm{Gen}(d) = \{\mathrm{Gen}(d)_I \colon (TX)^2 \to \mathcal{R}^+\}$ defined by*

$$\mathrm{Gen}(d)_I(c_1, c_2) = \sup_{k \colon I \to TX} d(k^\sharp(c_1), k^\sharp(c_2))$$

is an X-generated Eq*-relative $\mathcal{R}^+$-divergence on T where each $\mathrm{Gen}(d)_I$ is a pseudometric.*

Proof. From the reflexivity of d, we have the reflexivity of $\mathrm{Gen}(d)_I$: for each $c \in TI$,

$$\mathrm{Gen}(d)_I(c, c) = \sup_{k \colon I \to TX} d(k^\sharp(c), k^\sharp(c)) = 0.$$

Hence, the Eq-unit-reflexivity of Δ^d is already proved from the (proper) reflexivity. From the symmetry of d, we have the symmetry of Δ^d_I: for each $c_1, c_2 \in TI$,

$$\begin{aligned}
\mathrm{Gen}(d)_I(c_1, c_2) &= \sup_{k \colon I \to TX} d(k^\sharp(c_1), k^\sharp(c_2)) \\
&= \sup_{k \colon I \to TX} d(k^\sharp(c_2), k^\sharp(c_1)) \\
&= \mathrm{Gen}(d)_I(c_2, c_1).
\end{aligned}$$

From the triangle-inequality of d, we have the triangle-inequality of $\mathrm{Gen}(d)_I$: for all $c_1, c_2, c_3 \in TI$,

$$\begin{aligned}
\mathrm{Gen}(d)_I(c_1, c_3) &= \sup_{k \colon I \to TX} d(k^\sharp(c_1), k^\sharp(c_3)) \\
&\leq \sup_{k \colon I \to TX} d(k^\sharp(c_1), k^\sharp(c_2)) + d(k^\sharp(c_2), k^\sharp(c_3)) \\
&\leq \sup_{k \colon I \to TX} d(k^\sharp(c_1), k^\sharp(c_2)) + \sup_{k \colon I \to TX} d(k^\sharp(c_2), k^\sharp(c_3)) \\
&= \mathrm{Gen}(d)_I(c_1, c_2) + \mathrm{Gen}(d)_I(c_2, c_3).
\end{aligned}$$

From the reflexivity, congruence and substitutivity of d and the triangle-inequality of Δ^d_I, we next show the composability. Let $c_1, c_2 \in TI$ and $f_1, f_2 \colon I \to TJ$. We obtain,

$$\mathrm{Gen}(d)_J(f_1^\sharp(c_1), f_2^\sharp(c_2))$$

$$
\begin{aligned}
&\leq \mathrm{Gen}(d)_J(f_1^\sharp(c_1), f_1^\sharp(c_2)) + \mathrm{Gen}(d)_J(f_1^\sharp(c_2), f_2^\sharp(c_2)) \\
&= \sup_{k\colon J\to TX} d((k^\sharp \circ f_1)^\sharp(c_1), (k^\sharp \circ f_1)^\sharp(c_2)) + \sup_{k\colon J\to TX} d((k^\sharp \circ f_1)^\sharp(c_2), (k^\sharp \circ f_2)^\sharp(c_2)) \\
&\leq \sup_{k\colon J\to TX} d(f_1^\sharp(c_1), f_1^\sharp(c_2)) + \sup_{k\colon J\to TX} \sup_{i\in I} d(k^\sharp \circ f_1(i), k^\sharp \circ f_2(i)) \\
&= d(f_1^\sharp(c_1), f_1^\sharp(c_2)) + \sup_{k\colon J\to TX} \sup_{i\in I} d(k^\sharp \circ f_1(i), k^\sharp \circ f_2(i)) \\
&\leq \sup_{f_1\colon I\to TX} d(f_1^\sharp(c_1), f_1^\sharp(c_2)) + \sup_{i\in I} \sup_{k\colon J\to TX} d(k^\sharp \circ f_1(i), k^\sharp \circ f_2(i)) \\
&= \mathrm{Gen}(d)_I(c_1, c_2) + \sup_{i\in I} \mathrm{Gen}(d)_J(f_1(i), f_2(i)).
\end{aligned}
$$

Finally we show the X-generatedness of $\mathrm{Gen}(d)$ by definition

$$
\begin{aligned}
\mathrm{Gen}(d)I(c_1, c_2) &= \sup_{k\colon I\to TX} d(k^\sharp(c_1), k^\sharp(c_2)) \\
&= \sup_{h\colon X\to TX} \sup_{k\colon I\to TX} d(h^\sharp(k^\sharp(c_1)), h^\sharp(k^\sharp(c_2))) \\
&= \sup_{k\colon I\to TX} \sup_{h\colon X\to TX} d(h^\sharp(k^\sharp(c_1)), h^\sharp(k^\sharp(c_2))) \\
&= \sup_{k\colon I\to TX} \mathrm{Gen}(d)_X(k^\sharp(c_1), k^\sharp(c_2))
\end{aligned}
$$

This completes the proof. $\square$

The monotonicity of $\mathrm{Gen}\colon (\mathbf{CSEPMet}(T_\Omega, X), \preceq) \to (\mathbf{DivEPMet}(T_\Omega, X), \preceq)$ is easy to prove:

$$
\begin{aligned}
d \preceq d' \implies & \forall c_1, c_2 \in TI \ .\ \sup_{k\colon I\to TX} d(k^\sharp(c_1), k^\sharp(c_2)) \geq \sup_{k\colon I\to TX} d'(k^\sharp(c_1), k^\sharp(c_2)) \\
\iff & \forall c_1, c_2 \in TI \ .\ \mathrm{Gen}(d)_I(c_1, c_2) \geq \mathrm{Gen}(d')_I(c_1, c_2) \\
\iff & \mathrm{Gen}(d) \preceq \mathrm{Gen}(d').
\end{aligned}
$$

Proof. (Proof of Theorem 6) We first show $(\mathrm{Gen}(-))_X = \mathrm{id}$. Let $d \in \mathbf{CSEPMet}(T_\Omega, X)$. We fix arbitrary $t, u \in T_\Omega X$. From the substitutivity of d, we have $d(k^\sharp(t), k^\sharp(u)) \leq d(t, u)$, but we can take $k = \eta_X$, we obtain

$$
\mathrm{Gen}(d)XC(t, u) = \sup_{k\colon X\to TX} d(k^\sharp(t), k^\sharp(u)) = d(t, u).
$$

Since d, t, u are arbitrary, we conclude $(\mathrm{Gen}(-))_X = \mathrm{id}$.

We show $\mathrm{Gen}((-)_X) = \mathrm{id}$. Let $\Delta \in \mathbf{DivEPMet}(T_\Omega, X)$. By the X-generatedness of Δ, we have for all set I and $t, u \in T_\Omega I$,

$$
\mathrm{Gen}((\Delta)_X)_I(t, u) = \sup_{k\colon I\to TX} \Delta_X(k^\sharp(t), k^\sharp(u)) = \Delta_I(t, u).
$$

Since Δ, I, t, u are arbitrary, we conclude $(\mathrm{Gen}(-))_X = \mathrm{id}$.

We show the adjointness: $U[d] \subseteq V \iff d \geq d[V]$ for any $V \in \mathbf{QET}(\Omega, X)$ and $d \in \mathbf{CSEPMet}(T_\Omega, X)$.

$$
\begin{aligned}
U[d] \subseteq V \iff & \overline{\{\emptyset \vdash t =_\varepsilon u \mid \varepsilon \in \mathbb{Q}^+, d(t, u) \leq \varepsilon\}}^{\mathrm{QET}(\Omega, X)} \subseteq V \\
\iff & \forall t, u \in T_\Omega, \varepsilon \in \mathbb{Q}^+ \ .\ d(t, u) \leq \varepsilon \implies \emptyset \vdash t =_\varepsilon u \in V
\end{aligned}
$$

$$
\begin{aligned}
&\iff \forall t, u \in T_\Omega, \varepsilon \in \mathbb{Q}^+ \ . \ d(t,u) \le \varepsilon \implies \inf\left\{\varepsilon' \in \mathbb{Q}^+ \;\middle|\; \emptyset \vdash t =_{\varepsilon'} u \in V\right\} \le \varepsilon \\
&\iff \forall t, u \in T_\Omega \ . \ \inf\left\{\varepsilon' \in \mathbb{Q}^+ \;\middle|\; \emptyset \vdash t =_{\varepsilon'} u \in V\right\} \le d(t,u) \\
&\iff d \ge d[V]
\end{aligned}
$$

We notice that since V is closed under (Max), (Arch) and (Cut), we have the equivalence

$$
\begin{aligned}
&\inf\left\{\varepsilon' \in \mathbb{Q}^+ \;\middle|\; \emptyset \vdash t =_{\varepsilon'} u \in V\right\} \le \varepsilon \\
&\implies (\forall \varepsilon' \in \mathbb{Q}^+ \ . \ \varepsilon' > \varepsilon \implies \emptyset \vdash t =_{\varepsilon'} u \in V) \\
&\implies \emptyset \vdash t =_\varepsilon u \in V \\
&\implies \inf\left\{\varepsilon' \in \mathbb{Q}^+ \;\middle|\; \emptyset \vdash t =_{\varepsilon'} u \in V\right\} \le \varepsilon.
\end{aligned}
$$

We finally show $d[U[-]] = \mathrm{id}_{\mathbf{CSEPMet}(T_\Omega, X)}$. From the adjointness, $d[U[d]] \le d$ holds for each $d \in \mathbf{CSEPMet}(T_\Omega, X)$. We can rewrite $d \le d[U[d]]$ as follows:

$$
\begin{aligned}
d \le d[U[d]] &\iff \forall t, u \in T_\Omega \ . \ d(t,u) \le d[U[d]](t,u) \\
&\iff \forall t, u \in T_\Omega \ . \ d(t,u) \le \inf\left\{\varepsilon \in \mathbb{Q}^+ \;\middle|\; \emptyset \vdash t =_\varepsilon u \in U[d]\right\} \\
&\iff \forall t, u \in T_\Omega, \varepsilon \in \mathbb{Q}^+ \ . \ \implies d(t,u) \le \varepsilon \\
&\iff \{\emptyset \vdash t =_\varepsilon u \in u[d]\} \subseteq \{\emptyset \vdash t =_\varepsilon u | d(t,u) \le \varepsilon\}.
\end{aligned}
$$

Thanks to the minimality of $U[d]$, it suffices to have a QET $V \in \mathbf{QET}(\Omega, X)$ such that

$$
\{\emptyset \vdash t =_\varepsilon u \in V\} = \{\emptyset \vdash t =_\varepsilon u | d(t,u) \le \varepsilon\}.
$$

Inspired from the definition of models of QET ([6]), we define V as follows:

$$
\begin{aligned}
&\Gamma \vdash t =_\varepsilon u \in V \\
&\iff \forall \sigma \colon X \to T_\Omega X \ . \ \left(\left(\forall t' =_{\varepsilon'} u' \in \Gamma \ . \ d(\sigma^\sharp(t'), \sigma^\sharp(u')) \le \varepsilon'\right) \implies d(\sigma^\sharp(t), \sigma^\sharp(u)) \le \varepsilon\right).
\end{aligned}
$$

By the substitutivity of d and the definition of V, we obtain for all $t, u \in T_\Omega X$ and $\varepsilon \in \mathbb{Q}^+$,

$$
\emptyset \vdash t =_\varepsilon u \in V \iff (\forall \sigma \colon X \to T_\Omega X \ . \ d(\sigma^\sharp(t), \sigma^\sharp(u)) \le \varepsilon) \iff d(t,u) \le \varepsilon.
$$

We check that V satisfies all rules of QET:

(Ref) Immediate from the reflexivity of d.

(Sym) Immediate from the symmetry of d.

(Tri) Immediate from the triangle-inequality of d.

(Max) Immediate from the transitivity of ordering $\le$ and the monotonicity of $+$.

(Arch) Immediate from the Archimedean property and the completeness of $[0, \infty]$.

(Nonexp) Let $f \colon |I| \in \Omega$. We then take a term $t_f \in T_\Omega I$ corresponding to f. Let $t, s : I \to T_\Omega X$ be functions. We fix an arbitrary $\sigma \colon X \to T_\Omega X$. Assume $d(\sigma^\sharp(t(i)), \sigma^\sharp(s(i))) \le \varepsilon$ for each $i \in I$. Then this asserts $\sup_{i \in I} d(\sigma^\sharp(t(i)), \sigma^\sharp(s(i))) \le \varepsilon$. From the congruence of d, we conclude

$$
d(\sigma^\sharp(f(t(i)|i \in I)), \sigma^\sharp(f(s(i)|i \in I))) = d(\sigma^\sharp(t^\sharp(t_f)), \sigma^\sharp(s^\sharp(t_f))) \le \sup_{i \in I} d(\sigma^\sharp(t(i)), \sigma^\sharp(s(i))) \le \varepsilon.
$$

(Subst) Immediate by definition of V:

$$
\begin{aligned}
&\Gamma \vdash t =_\varepsilon u \in V \\
&\iff \forall \sigma \colon X \to T_\Omega X \ . \ \left(\left(\forall t' =_{\varepsilon'} u' \in \Gamma \ . \ d(\sigma^\sharp(t'), \sigma^\sharp(u')) \le \varepsilon'\right) \implies d(\sigma^\sharp(t), \sigma^\sharp(u)) \le \varepsilon\right)
\end{aligned}
$$

$$\implies \forall \sigma' \colon X \to T_\Omega X \;.\; \forall \sigma \colon X \to T_\Omega X \;.\; \left(\begin{array}{r} \left(\forall t' =_{\varepsilon'} u' \in \Gamma \;.\; d(\sigma^\sharp(\sigma'^\sharp(t')), \sigma^\sharp(\sigma'^\sharp(u'))) \le \varepsilon'\right) \\ \implies d(\sigma^\sharp(\sigma'^\sharp(t)), \sigma^\sharp(\sigma'^\sharp(u))) \le \varepsilon \end{array} \right)$$

$$\implies \forall \sigma' \colon X \to T_\Omega X \;.\; \forall \sigma \colon X \to T_\Omega X \;.\; \left(\begin{array}{r} \left(\forall t'' =_{\varepsilon'} u'' \in \sigma'(\Gamma) \;.\; d(\sigma^\sharp(t''), \sigma^\sharp(u'')) \le \varepsilon'\right) \\ \implies d(\sigma^\sharp(\sigma'^\sharp(t)), \sigma^\sharp(\sigma'^\sharp(u))) \le \varepsilon \end{array} \right)$$

$$\iff \forall \sigma' \colon X \to T_\Omega X \;.\; \sigma'(\Gamma) \vdash \sigma'^\sharp(t) =_\varepsilon \sigma'^\sharp(u) \in V.$$

(Cut) Immediate.
(Assumpt) Immediate. □

Proof. (Proof of Theorem 7) Since the range of $U[-]$ is a subset of $\mathbf{UQET}(\Omega, X)$, we may define the following monotone restrictions of $U[-]$ and $d[-]$:

$$U'[-] \colon (\mathbf{CSEPMet}(T_\Omega, X), \preceq) \to (\mathbf{UQET}(\Omega, X), \subseteq) \quad U'[d] \triangleq U[d] \quad (d \in \mathbf{CSEPMet}(T_\Omega, X)),$$
$$d'[-] \colon (\mathbf{UQET}(\Omega, X), \subseteq) \to (\mathbf{CSEPMet}(T_\Omega, X), \preceq) \quad d'[V] \triangleq d[V] \quad (V \in \mathbf{UQET}(\Omega, X)).$$

By Theorem 6, we have $U'[-] \vdash d'[-]$ and $d'[U'[-]] = \mathrm{id}$. We show $U'[d'[-]] = \mathrm{id}$. Let $V \in \mathbf{UQET}(\Omega, X)$. There exists $S \subseteq \{\emptyset \vdash t =_\varepsilon u \mid t, u \in T_\Omega X, \varepsilon \in \mathbb{Q}^+\}$ such that $V = \overline{S}^{\mathrm{QET}(\Omega,X)}$. We check $U'[d'[V]] = V$. By the adjunction $U'[-] \dashv d'[-]$, we have $U'[d'[V]] \subseteq V$ which is equivalent to $d'[V] \preceq d'[V]$. It suffices to check $V \subseteq U'[d'[V]]$. We have

$$\begin{aligned} &\emptyset \vdash t =_\varepsilon u \in S \\ &\implies \emptyset \vdash t =_\varepsilon u \in V \\ &\implies d'[V](t, u) = \inf\{\varepsilon' \in \mathbb{Q}^+ \mid \emptyset \vdash t =_{\varepsilon'} u \in V\} \le \varepsilon \end{aligned}$$

From the monotonicity of the closure $\overline{(-)}^{\mathrm{QET}(\Omega,X)}$, we conclude

$$V = \overline{S}^{\mathrm{QET}(\Omega,X)} \subseteq \overline{\{\emptyset \vdash t =_\varepsilon u \mid d'[V](t, u) \le \varepsilon\}}^{\mathrm{QET}(\Omega,X)} = U'[d'[V]].$$

Since $V \in \mathbf{UQET}(\Omega, X)$ is arbitrary, we have $U'[d'[-]] = \mathrm{id}$. □

C Proofs for Section 7 (Graded Strong Relational Liftings for Divergences)

Lemma 4. *Let $(\mathbb{C}, T)$ be a CC-SM and $\Delta = \{\Delta_I^m \colon (U(TI))^2 \to \mathcal{Q}\}_{m \in M, I \in \mathbb{C}}$ be a doubly-indexed family of $\mathcal{Q}$-divergences satisfying monotonicity on m (Definition 6). Then $T^{[\Delta]}$ is an $M \times \mathcal{Q}$-graded relational lifting of T (satisfies conditions 1–3 of Definition 15).*

Proof. (Condition 1) We first show that $(\mathrm{id}_{TX_1}, \mathrm{id}_{TX_2}) \in \mathbf{BRel}(\mathbb{C})(T^{[\Delta]}(m, v)X, T^{[\Delta]}(n, w)X)$ for all X whenever $m \le n$ and $v \le w$. From the monotonicity of Δ, for all $I \in \mathbb{C}$, $c_1', c_2' \in U(TI)$, $n' \in M$,$w' \in \mathcal{Q}$, we have

$$\begin{aligned} &(c_1', c_2') \in \tilde{\Delta}(m \cdot n', v + w')I \\ &\iff \Delta_I^{m \cdot n'}(c_1', c_2') \le v + w' \implies \Delta_I^{n \cdot n'}(c_1', c_2') \le v + w' \implies \Delta_I^{n \cdot n'}(c_1', c_2') \le w + w' \\ &\iff (c_1', c_2') \in \tilde{\Delta}(n \cdot n', w + w')I. \end{aligned}$$

Therefore, for any $(c_1, c_2) \in T^{[\Delta]}(m, v)X$, we obtain $(c_1, c_2) \in T^{[\Delta]}(n, w)X$ as follows:

$$\begin{aligned}
&(c_1, c_2) \in T^{[\Delta]}(m, v)X \\
&\iff \forall I \in \mathbb{C}, n' \in M, w' \in \mathcal{Q}, (k_1, k_2) : X \dot{\to} \tilde{\Delta}(n', w')I \ . \ (k_1^\sharp \bullet c_1, k_2^\sharp \bullet c_2) \in \tilde{\Delta}(m \cdot n', v + w')I \\
&\implies \forall I \in \mathbb{C}, n' \in M, w' \in \mathcal{Q}, (k_1, k_2) : X \dot{\to} \tilde{\Delta}(n', w')I \ . \ (k_1^\sharp \bullet c_1, k_2^\sharp \bullet c_2) \in \tilde{\Delta}(n \cdot n', w + w')I \\
&\iff (c_1, c_2) \in T^{[\Delta]}(n, w)X.
\end{aligned}$$

(Condition 2) We next show $(\eta_{X_1}, \eta_{X_2}) : X \dot{\to} T^{[\Delta]}(1, 0)X$. From the definition of morphisms in $\mathbf{BRel}(\mathbb{C})$, for all $(x_1, x_2) \in X$, we have $(\eta_{X_1} \bullet x_1, \eta_{X_2} \bullet x_2) \in T^{[\Delta]}(1, 0)X$ as follows:

$$\begin{aligned}
&(x_1, x_2) \in X \\
&\implies \forall I \in \mathbb{C}, n \in M, w \in \mathcal{Q}, (k_1, k_2) : X \dot{\to} \tilde{\Delta}(n, w)I \ . \ (k_1 \bullet x_1, k_2 \bullet x_2) \in \tilde{\Delta}(n, w)I \\
&\iff \forall I \in \mathbb{C}, n \in M, w \in \mathcal{Q}, (k_1, k_2) : X \dot{\to} \tilde{\Delta}(n, w)I \ . \ ((k_1^\sharp \circ \eta_{X_1}) \bullet x_1), (k_2^\sharp \circ \eta_{X_2}) \bullet x_2) \in \tilde{\Delta}(n, w)I \\
&\iff \forall I \in \mathbb{C}, n \in M, w \in \mathcal{Q}, (k_1, k_2) : X \dot{\to} \tilde{\Delta}(n, w)I \ . \ (k_1^\sharp \bullet (\eta_{X_1} \bullet x_1), k_2^\sharp \bullet (\eta_{X_2} \bullet x_2)) \in \tilde{\Delta}(n, w)I \\
&\iff (\eta_{X_1} \bullet x_1, \eta_{X_2} \bullet x_2) \in T^{[\Delta]}(1, 0)X.
\end{aligned}$$

(Condition 3) Finally, we show that $(f_1^\sharp, f_2^\sharp) : T^{[\Delta]}(n, w)X \dot{\to} T^{[\Delta]}(n \cdot m, w + v)Y$ holds for any $(f_1, f_2) : X \dot{\to} T^{[\Delta]}(m, v)Y$ and $(n, w) \in M \times \mathcal{Q}$. For all $(f_1, f_2) : X \dot{\to} T^{[\Delta]}(m, v)Y$, we have

$$\begin{aligned}
&(f_1, f_2) : X \dot{\to} T^{[\Delta]}(m, v)Y \\
&\iff \forall (x_1, x_2) \in X \ . \ (f_1 \bullet x_1, f_2 \bullet x_2) \in T^{[\Delta]}(m, v)Y \\
&\iff \begin{pmatrix} \forall (x_1, x_2) \in X, I \in \mathbb{C}, n' \in M, w' \in \mathcal{Q}, (k_1, k_2) : Y \dot{\to} \tilde{\Delta}(n', w')I \ . \\ (k_1^\sharp \bullet (f_1 \bullet x_1), k_2^\sharp \bullet (f_2 \bullet x_2)) \in \tilde{\Delta}(m \cdot n', v + w')I \end{pmatrix} \\
&\iff \begin{pmatrix} \forall (x_1, x_2) \in X, I \in \mathbb{C}, n' \in M, w' \in \mathcal{Q}, (k_1, k_2) : Y \dot{\to} \tilde{\Delta}(n', w')I \ . \\ ((k_1^\sharp \circ f_1) \bullet x_1), (k_2^\sharp \circ f_2) \bullet x_2) \in \tilde{\Delta}(m \cdot n', v + w')I \end{pmatrix} \\
&\iff \begin{pmatrix} \forall I \in \mathbb{C}, n' \in M, w' \in \mathcal{Q}, (k_1, k_2) : Y \dot{\to} \tilde{\Delta}(n', w')I \ . \\ (k_1^\sharp \circ f_1, k_2^\sharp \circ f_2) : X \dot{\to} \tilde{\Delta}(m \cdot n', v + w')I \end{pmatrix}. \qquad \text{(a)}
\end{aligned}$$

For all $(c_1, c_2) \in T^{[\Delta]}(n, w)X$, we have

$$\begin{aligned}
&(c_1, c_2) \in T^{[\Delta]}(n, w)X \\
&\iff \begin{pmatrix} \forall I \in \mathbb{C}, n' \in M, w' \in \mathcal{Q}, (l_1, l_2) : X \dot{\to} \tilde{\Delta}(n', w')I \ . \\ (l_1^\sharp \bullet c_1, l_2^\sharp \bullet c_2) \in \tilde{\Delta}(n \cdot n', w + w')I \end{pmatrix}. \qquad \text{(b)}
\end{aligned}$$

We here fix $(f_1, f_2) : X \dot{\to} T^{[\Delta]}(m, v)Y$. We show $(f_1^\sharp, f_2^\sharp)$: $T^{[\Delta]}(n, w)X \dot{\to} T\Delta(n \cdot m, w + v)Y$. We also fix $I \in \mathbb{C}$, $n'' \in M$, $w'' \in \mathcal{Q}$ and $(k_1, k_2) : Y \dot{\to} \tilde{\Delta}(n'', w'')I$. From (a), we obtain

$$(k_1^\sharp \circ f_1, k_2^\sharp \circ f_2) : X \dot{\to} \tilde{\Delta}(m \cdot n'', v + w'')I.$$

Therefore, by instantiating (b) with $(n', w') = (m \cdot n'', v + w'')$ and $(l_1, l_2) = (k_1^\sharp \circ f_1, k_2^\sharp \circ f_2)$, for all $(c_1, c_2) \in T^{[\Delta]}(n, w)X$, we have

$$((k_1^\sharp \circ f_1)^\sharp \bullet c_1, (k_2^\sharp \circ f_2)^\sharp \bullet c_2) \in \tilde{\Delta}(n \cdot m \cdot n'', w + v + w'')I.$$

Since $(c_1, c_2) \in T^{[\Delta]}(n, w)X$, $I \in \mathbb{C}$, $n'' \in M$, $w'' \in \mathcal{Q}$ and $(k_1, k_2) : Y \dot{\to} \tilde{\Delta}(n'', w'')I$ are arbitrary, we conclude $(f_1^\sharp, f_2^\sharp) \colon T^{[\Delta]}(n, w)X \dot{\to} T\Delta(n \cdot m, w + v)$ as follows:

$$
\begin{aligned}
&\begin{pmatrix} \forall (c_1, c_2) \in T^{[\Delta]}(n, w)X, I \in \mathbb{C}, m'' \in M, v'' \in \mathcal{Q}, (k_1, k_2) : Y \dot{\to} \tilde{\Delta}(m'', v'')I \ . \\ ((k_1^\sharp \circ f_1)^\sharp \bullet c_1, (k_2^\sharp \circ f_2)^\sharp \bullet c_2) : X \dot{\to} \tilde{\Delta}(n \cdot m \cdot m'', w + v + v'')I \end{pmatrix} \\
\iff &\begin{pmatrix} \forall (c_1, c_2) \in T^{[\Delta]}(n, w)X, I \in \mathbb{C}, m'' \in M, v'' \in \mathcal{Q}, (k_1, k_2) : Y \dot{\to} \tilde{\Delta}(m'', v'')I \ . \\ (k_1^\sharp \bullet (f_1^\sharp \bullet c_1), k_2^\sharp \bullet (f_2^\sharp \bullet c_2)) : X \dot{\to} \tilde{\Delta}(n \cdot m \cdot m'', w + v + v'')I \end{pmatrix} \\
\iff &\forall (c_1, c_2) \in T^{[\Delta]}(n, w)X \ . \ (f_1^\sharp \bullet c_1, f_2^\sharp \bullet c_2) \in T^{[\Delta]}(n \cdot m, w + v)Y \\
\iff &(f_1^\sharp, f_2^\sharp) \colon T^{[\Delta]}(n, w)X \dot{\to} T\Delta(n \cdot m, w + v).
\end{aligned}
$$

This completes the proof. $\square$

Proof. (Proof of Proposition 13) By Theorem 8 and the assumption $\forall I, J \in \mathbb{C} \ . \ EI \dot{\times} EJ \subseteq E(I \times J)$, we obtain for all $(x_1, x_2) \in EI$ and $c_1, c_2 \in U(TI)$,

$$
\begin{aligned}
&(\langle x_1, c_1 \rangle, \langle x_2, c_2 \rangle) \in EI \dot{\times} \tilde{\Delta}(m, v)J \\
\iff &(\langle x_1, c_1 \rangle, \langle x_2, c_2 \rangle) \in EI \dot{\times} T^{[\Delta]}(m, v)(EJ) \\
\implies &(\theta_{I,J} \bullet \langle x_1, c_1 \rangle, \theta_{I,J} \bullet \langle x_2, c_2 \rangle) \in T^{[\Delta]}(m, v)(EI \dot{\times} EJ) \\
\implies &(\theta_{I,J} \bullet \langle x_1, c_1 \rangle, \theta_{I,J} \bullet \langle x_2, c_2 \rangle) \in T^{[\Delta]}(m, v)E(I \times J) \\
\iff &(\theta_{I,J} \bullet \langle x_1, c_1 \rangle, \theta_{I,J} \bullet \langle x_2, c_2 \rangle) \in \tilde{\Delta}(m, v)(I \times J).
\end{aligned}
$$

This completes the proof. $\square$

D Proofs for Section 9 (Case Study I: Higher-Order Probabilistic Programs)

Lemma 5. *The mapping*

$$
(x, \sigma) \mapsto \begin{cases} \mathcal{N}(x, \sigma^2) & \sigma^2 > 0 \\ \mathbf{d}_x & \sigma = 0 \end{cases}
$$

forms a measurable function of type $\mathbb{R} \times \mathbb{R} \to G\mathbb{R}$.

Proof. We show that for all $A \in \Sigma_{\mathbb{R}}$, the mapping $f_A(x, \sigma) = \mathcal{N}(x, \sigma^2)(A)$ forms a measurable function of type $\mathbb{R} \times \mathbb{R}_{\neq 0} \to [0, 1]$ where $\mathbb{R}_{\neq 0}$ is the subspace of $\mathbb{R}$ whose underlying set is $\{r \in \mathbb{R} | r \neq 0\}$. We have

$$
\mathcal{N}(x, \sigma^2)(A) = \sum_{k \in \mathbb{Z}} \mathcal{N}(x, \sigma^2)(A \cap [k, k+1]) = \sum_{k \in \mathbb{Z}} \int_{A \cap [k, k+1]} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x - r)^2}{\sigma^2}\right) dr
$$

The mapping $h(x, \sigma, r) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-r)^2}{\sigma^2}\right)$ forms a continuous function of type $\mathbb{R} \times \mathbb{R}_{\neq 0} \times \mathbb{R} \to \mathbb{R}$, hence it is uniformly continuous on the compact set $I_1 \times I_2 \times [k, k+1]$ where I_1 and I_2 are arbitrary closed intervals in $\mathbb{R}$ and $\mathbb{R}_{\neq 0}$ respectively. Then, for all $0 < \varepsilon$, there exists $0 < \delta$ such that $|h(x, \sigma, r) - h(x', \sigma', r')| < \varepsilon$ holds wherever $|x - x'| + |\sigma - \sigma'| + |r - r'| < \delta$. Hence, for all $0 < \varepsilon$, there is $0 < \delta$ such that whenever $|x - x'| + |\sigma - \sigma'| < \delta$,

$$
\left| \int_{A \cap [k, k+1]} h(x, \sigma, r) dr - \int_{A \cap [k, k+1]} h(x', \sigma', r) dr \right| \leq | \int_{[k, k+1]} |h(x, \sigma, r) - h(x', \sigma', r')| dr \leq \varepsilon.
$$

Since the closed intervals I_1 and I_2 are arbitrary, we conclude that the function $f_{A\cap[k,k+1]}\colon \mathbb{R}\times\mathbb{R}_{\neq 0} \to [0,1]$ is continuous, hence measurable. Hence, the mapping $f_A = \sum_{k\in\mathbb{Z}} f_{A\cap[k,k+1]}$ is measurable. Since A is arbitrary and $f_A(x,\sigma^2) = \mathrm{ev}_A \circ \mathcal{N}(x,\sigma^2)$, the mapping $g(x,\sigma^2) = \mathcal{N}(x,\sigma^2)$ forms a measurable function of type $\mathbb{R}\times\mathbb{R}_{\neq 0} \to G\mathbb{R}$. The rest of proof is routine. $\square$

Corollary 1. $[\![\texttt{norm}]\!] \in \mathbf{QBS}(K\mathbb{R}\times K\mathbb{R}, PK\mathbb{R})$.

Lemma 6 (Measurability of $[\![\texttt{lap}]\!]$)**.** *The mapping*

$$(x,\lambda) \mapsto \begin{cases} \mathrm{Lap}(x,\lambda) & \lambda > 0 \\ \mathbf{d}_x & \lambda \leq 0 \end{cases}$$

forms a measurable function of type $\mathbb{R}\times\mathbb{R} \to G\mathbb{R}$*.*

Proof. We have, for all $A \in \Sigma_{\mathbb{R}}$,

$$\mathrm{Lap}(x,\lambda)(A) = \int_A \frac{1}{2\lambda} \exp\left(-\frac{|x-r|}{\lambda}\right) dr$$

The density function $h(x,\lambda,r) = \frac{1}{2\lambda}\exp\left(-\frac{|x-r|}{\lambda}\right)$ is continuous function of type $\mathbb{R}\times\mathbb{R}_{0\leq}\times\mathbb{R} \to \mathbb{R}$ where $\mathbb{R}_{0\leq}$ is the subspace of $\mathbb{R}$ whose underlying set is $\{r \in \mathbb{R} | 0 \leq r\}$. The measurability of $\mathrm{Lap}(x,\lambda)$ is proved in the same way as $\mathcal{N}(x,\sigma^2)$. The rest of proof is routine. $\square$

Corollary 2. $[\![\texttt{lap}]\!] \in \mathbf{QBS}(K\mathbb{R}\times K\mathbb{R}, PK\mathbb{R})$.