


PAPER

# Divergences on monads for relational program logics

Tetsuya Sato<sup>1</sup>  and Shin-ya Katsumata<sup>2</sup> 

<sup>1</sup>Tokyo Institute of Technology, 2-12-1 Ookayama, Meguro-ku, Tokyo, Japan, <sup>2</sup>National Institute of Informatics, 2-1-2 Chiyoda-ku, Tokyo, Japan

**Corresponding author:** Shin-ya Katsumata; Email: [s-katsumata@nii.ac.jp](mailto:s-katsumata@nii.ac.jp)

(Received 28 November 2021; revised June 2023; accepted 29 June 2023; first published online 31 July 2023)

## Abstract

Several relational program logics have been introduced for integrating reasoning about relational properties of programs and measurement of quantitative difference between computational effects. Toward a general framework for such logics, in this paper, we formalize the concept of quantitative difference between computational effects as *divergences on monads*, then develop a relational program logic called *approximate computational relational logic* (acRL for short). It supports generic computational effects and divergences on them. The semantics of the acRL is given by *graded strong relational liftings* constructed from divergences on monads. We derive two instantiations of the acRL: (1) for the verification of various kinds of differential privacy of higher-order functional probabilistic programs and (2) the other for measuring difference of distributions of cost between higher-order functional probabilistic programs with a cost counting operator.

**Keywords:** Relational Hoare logics; graded monads; relational liftings

## 1. Introduction

Comparing behavior of programs is one of the fundamental activities in verification and analysis of programs, and many concepts, techniques, and formal systems have been proposed for this purpose, such as *product program construction* (Barthe et al., 2011, 2004; Zaks and Pnueli, 2008), *relational Hoare logic* (Benton, 2004), *higher-order relational refinement types* (Barthe et al., 2015), and so on.

Several recent relational program logics integrate compositional reasoning about relational properties of programs and over-approximating *quantitative differences* between computational effects of programs; the latter is done in the style of *effect system* (Lucassen and Gifford, 1988). 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 (Barthe et al., 2012). A judgment of apRHL is of the form  $c \sim_{\epsilon, \delta} c' : \Phi \Rightarrow \Psi$ , and its intuitive meaning is that for any pair of initial states  $(\rho, \rho')$  related by  $\Phi$ , the  $\epsilon$ -distance between two probability distributions of final states  $\llbracket c \rrbracket \rho$  and  $\llbracket c' \rrbracket \rho'$  is below  $\delta$ , and final states are related by  $\Psi$ . Another relational program logic that measures differences between computational effects of programs is Çiçek et al.'s RelCost (Çiçek et al., 2017). The target of the reasoning is a higher-order functional programming language equipped with cost counting effect. When we derive a judgment  $\Delta; \Psi; \Gamma \vdash M_1 \ominus M_2 \lesssim n : \Phi$  in RelCost, its 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 differences between computational effects into relational program logic. We aim to mathematically formulate this perspective in this paper. The central concept in our formulation is a *divergence on a monad*. It specifies quantitative differences between computational effects modeled by a monad, and it may be seen as a generalization of various statistical divergences, such as the Kullback–Leibler divergence and the total variation distance, to general computational effects. We then present a construction of the semantics of a relational program logic that integrates the measurement of quantitative differences between computational effects from a divergence on a monad. This construction works with general monads and divergences on them and can be applied to derive the semantics of the approximate relational program logic for differential privacy as an instance. The detail of the development of our formulation is as follows:

- We introduce a structure called the *divergence on a monad* for measuring quantitative differences between computational effects in Section 4. Divergences on monads are a mild generalization of Barthe and Olmedo’s composable divergences on probability distributions, such as the Kullback–Leibler divergence and the total variation distance on probability distributions (Barthe and Olmedo, 2013). We visit examples of divergences on monads in Section 5 and give a method to transfer divergences on a monad to those on another monad through monad opfunctors.
- We study mathematical aspects of divergences on monads in Section 6. We first reformulate them as structures in the category of divergences. We show that divergences on monads can be lifted to the category of divergences as monads, and that these liftings yield relative monads corresponding to the original divergences on monads. We next introduce the concept of *generatedness* of divergences on monads. It is a direct generalization of the same concept studied for differential privacy by Balle et al. (2020). We then associate them to the concept of *quantitative equational theories* studied by Mardare et al. (2016). We show that unconditional quantitative equational theories on a set  $X$  of variables bijectively correspond to  $X$ -generated pseudometrics on free monads.
- The key structure to integrate divergences on monads and relational program logics is something called the *graded strong relational liftings* of a monad. We present a general construction of such liftings from divergences on monads in Section 7. This construction makes it possible to give semantics of relational program logics with quantitative measurement of differences between computational effects for various 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 differences 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 (Mac Lane, 1998) and Moggi’s model of computational effects (Moggi, 1991). The definition of monad (Mac Lane, 1998, Chapter VI) and Kleisli category (Mac Lane, 1998, Section VI.5) are omitted.

In this paper, a Cartesian category (CC for short) is specified by a category  $\mathbb{C}$  with a designated terminal object  $1$  and a binary product functor  $(\times) : \mathbb{C}^2 \rightarrow \mathbb{C}$ . 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  $!_I : I \rightarrow 1$ . A Cartesian closed category (CCC for short) is a

CC  $(\mathbb{C}, 1, (\times))$  with a specified exponential functor  $(\Rightarrow): \mathbb{C}^{\text{op}} \times \mathbb{C} \rightarrow \mathbb{C}$ . The associated evaluation morphism and currying operation is denoted by  $\text{ev}, \lambda(-)$ , respectively. There are plenty of examples of  $\text{C}(\text{C})\text{Cs}$ . For the models of probabilistic computation, we will use the CC **Meas** of measurable spaces and the CCC **QBS** of quasi-Borel spaces (Heunen et al., 2017), which we briefly recall in Section 2.1.

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

A monad  $(T, \eta, \mu)$  on a category  $\mathbb{C}$  determines the operation  $(-)^{\sharp}: \mathbb{C}(I, TJ) \rightarrow \mathbb{C}(TI, TJ)$  called *Kleisli extension*. It is defined by  $f^{\sharp} \triangleq \mu_J \circ Tf$ . A monad may be equivalently given by a *Kleisli triple* (Moggi, 1991, Definition 1.2) that axiomatizes the triple  $(T, \eta, (-)^{\sharp})$ . A *strong monad* on a CC  $(\mathbb{C}, 1, (\times))$  is a pair of a monad  $(T, \eta, \mu)$  and a natural transformation  $\theta_{I,J}: I \times TJ \rightarrow T(I \times J)$  called *strength*. It should satisfy four axioms; see (Moggi, 1991, Definition 3.2) for detail.

A *Cartesian category with a strong monad* (CC-SM for short) is a pair of a CC and a strong monad on it. A Cartesian closed category with a strong monad (CCC-SM for short) is similarly defined. 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  $T$  (Moggi, 1991, 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, Proposition 17, and Theorem 39.

### 2.1 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  $\Sigma_I$  for the underlying set and  $\sigma$ -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  $|-|: \mathbf{Meas} \rightarrow \mathbf{Set}$ , which 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  $\sigma$ -algebra  $\Sigma_{\Omega}$  is the coarsest one including the topology  $\sigma_{\Omega}$  of a Polish space  $(|\Omega|, \sigma_{\Omega})$ . In particular, the real line  $\mathbb{R}$  forms a standard Borel space. In fact, a measurable space  $\Omega$  is standard Borel if and only if there are  $\gamma: \Omega \rightarrow \mathbb{R}$  and  $\gamma': \mathbb{R} \rightarrow \Omega$  in **Meas** forming a section–retraction pair, that is,  $\gamma' \circ \gamma = \text{id}_{\Omega}$ . For example,  $[0, 1]$ ,  $[0, \infty]$ ,  $\mathbb{N}$ , and  $\mathbb{R}^k$  ( $k \in \mathbb{N}$ ) are standard Borel.

**The Giry monad.** We recall the Giry monad  $G$  (Giry, 1982). For every measurable space  $I$ ,  $GI$  is the measurable space given by the following data: the underlying set  $|GI|$  is the set of all probability measures over  $I$ , and the  $\sigma$ -algebra is the coarsest one induced by functions  $\text{ev}_A: |GI| \rightarrow [0, 1]$  ( $A \in \Sigma_X$ ) defined by  $\text{ev}_A(\mu) = \mu(A)$ . The unit  $\eta_I: I \rightarrow GI$  assigns to each  $x \in I$  the Dirac measure  $\mathbf{d}_x$  centered at  $x$ . For every  $f: I \rightarrow GJ$ , the Kleisli extension  $f^{\sharp}: GI \rightarrow GJ$  is given by  $(f^{\sharp}(\mu))(A) = \int_x f(x)(A) d\mu(x)$  for each  $\mu \in GI$ . By  $G_s$ , we mean 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 subprobability measures over  $I$ .

The Giry monad  $G$  (resp. the sub-Giry monad  $G_s$ ) carries a (commutative) strength  $\theta_{I,J}: I \times GJ \rightarrow G(I \times J)$  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 (Aumann, 1961)<sup>1</sup>. 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** (Heunen et al., 2017). 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  $f : \mathbb{R} \rightarrow \mathbb{R}, \alpha \circ f \in M_I$ .
- (2) for any  $x \in I, (\lambda r \in \mathbb{R}.x) \in M_I$ .
- (3) for all  $P : \mathbb{R} \rightarrow \mathbb{N}$  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  $f : (|I|, M_I) \rightarrow (|J|, M_J)$  is a function  $f : |I| \rightarrow |J|$  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  $|-| : \mathbf{QBS} \rightarrow \mathbf{Set}$ . 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  $L \dashv K : \mathbf{Meas} \rightarrow \mathbf{QBS}$ . They are given by:

$$\begin{aligned}
 LI &\triangleq (|I|, \{U \subseteq |I| \mid \forall \alpha \in M_X.\alpha^{-1}(U) \in \Sigma_{\mathbb{R}}\}) & Lf &\triangleq f \\
 KI &\triangleq (|I|, \mathbf{Meas}(\mathbb{R}, I)) & Kf &\triangleq f
 \end{aligned}$$

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 (Heunen et al., 2017, Proposition 15-(2)). The right adjoint  $K$  preserves countable coproducts and function spaces (if exists) of standard Borel spaces (Heunen et al., 2017, Proposition 19).

**Probability Measures and the Probability Monad.** A probability measure on a quasi-Borel space  $I$  is a pair  $(\alpha, \mu) \in M_I \times \mathbb{G}\mathbb{R}$ . 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 **QBS** as follows:

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

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

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

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

### 2.2 Category of binary relations

We define the category **BRel**( $\mathbb{C}$ ) of binary relations over  $\mathbb{C}$ -objects. This category is equivalent to *subcones* of  $\mathbb{C}^2$  (Mitchell and Scedrov, 1992). It offers an underlying category for relational reasoning about programs interpreted in  $\mathbb{C}$ .

- An object in **BRel**( $\mathbb{C}$ ) is a triple  $(I_1, I_2, R)$  where  $R \subseteq UI_1 \times UI_2$ .
- A morphism from  $(I_1, I_2, R)$  to  $(J_1, J_2, S)$  in **BRel**( $\mathbb{C}$ ) is a pair of  $\mathbb{C}$ -morphisms  $f_1 : I_1 \rightarrow J_1$  and  $f_2 : I_2 \rightarrow J_2$  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  $(f_1, f_2): (X_1, X_2) \rightarrow (Y_1, Y_2)$  in  $\mathbb{C}^2$ , by

$$(f_1, f_2): X \dot{\rightarrow} Y$$

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  $p_{\mathbb{C}}: \mathbf{BRel}(\mathbb{C}) \rightarrow \mathbb{C}^2$  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  $X \leq Y \iff R_X \subseteq R_Y$ .

When  $\mathbb{C}$  is a  $\mathbf{C}(\mathbf{C})\mathbf{C}$ , so is  $\mathbf{BRel}(\mathbb{C})$  (Mitchell and Scedrov, 1992, Proposition 4.3). We specify a terminal object, a binary product functor and an exponential functor (in case  $\mathbb{C}$  is a  $\mathbf{CCC}$ ) on  $\mathbf{BRel}(\mathbb{C})$  by:

$$\begin{aligned} \dot{1} &\triangleq (1, 1, \{(id_1, id_1)\}) \\ X \dot{\times} Y &\triangleq (X_1 \times Y_1, X_2 \times Y_2, \{(\langle x_1, y_1 \rangle, \langle x_2, y_2 \rangle) \mid (x_1, x_2) \in X, (y_1, y_2) \in Y\}) \\ X \dot{\Rightarrow} Y &\triangleq (X_1 \Rightarrow Y_1, X_2 \Rightarrow Y_2, \{(f_1, f_2) \mid \forall (x_1, x_2) \in X. (ev \circ \langle f_1, x_1 \rangle, ev \circ \langle f_2, x_2 \rangle) \in Y\}). \end{aligned}$$

### 3. Divergences on Objects

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

**Definition 1.** A divergence domain  $\mathcal{Q} = (Q, \leq, 0, (+))$  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 as follows:

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

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

**Definition 2.** Let  $\mathbb{C}$  be a  $\mathbf{CC}$ . A  $\mathcal{Q}$ -divergence on an object  $I \in \mathbb{C}$  is a function  $d: (UI)^2 \rightarrow \mathcal{Q}$ . For two  $\mathcal{Q}$ -divergences  $d, d'$  on  $I$ , by  $d \leq_I d'$ , we mean  $\forall x_1, x_2 \in UI. d'(x_1, x_2) \leq d(x_1, x_2)$ .

The binary relation  $\leq_I$  is a partial order on the set of  $\mathcal{Q}$ -divergences on  $I$ .

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

**Definition 3.** Let  $\mathbb{C}$  be a  $\mathbf{CC}$ . 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  $f: I \rightarrow J$  such that for any  $x_1, x_2 \in UI$ ,  $e(f \bullet x_1, f \bullet x_2) \leq d(x_1, x_2)$  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  $V_{\mathcal{Q},\mathbb{C}} : \mathbf{Div}_{\mathcal{Q}}(\mathbb{C}) \rightarrow \mathbb{C}$  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  $V_{\mathcal{Q},\mathbf{Set}} : \mathbf{Div}_{\mathcal{Q}}(\mathbf{Set}) \rightarrow \mathbf{Set}$  is a (Grothendieck) fibration, and the functor  $\bar{U} : \mathbf{Div}_{\mathcal{Q}}(\mathbb{C}) \rightarrow \mathbf{Div}_{\mathcal{Q}}(\mathbf{Set})$  defined by  $\bar{U}(I, d) \triangleq (UI, d)$  and  $\bar{U}(f) \triangleq f$  makes the following commutative square a pullback in **CAT** (the large category of categories and functors between them):

$$\begin{array}{ccc}
 \mathbf{Div}_{\mathcal{Q}}(\mathbb{C}) & \xrightarrow{\bar{U}} & \mathbf{Div}_{\mathcal{Q}}(\mathbf{Set}) \\
 V_{\mathcal{Q},\mathbb{C}} \downarrow \lrcorner & & \downarrow V_{\mathcal{Q},\mathbf{Set}} \\
 \mathbb{C} & \xrightarrow{U} & \mathbf{Set}
 \end{array}$$

Therefore, this pullback diagram asserts that  $V_{\mathcal{Q},\mathbb{C}} : \mathbf{Div}_{\mathcal{Q}}(\mathbb{C}) \rightarrow \mathbb{C}$  arises from the *change-of-base* (Jacobs, 1999, Lemma 1.5.1) of the fibration  $V_{\mathcal{Q},\mathbf{Set}}$  along the global section functor  $U : \mathbb{C} \rightarrow \mathbf{Set}$ .

### 4. Divergences on Monads

We introduce the concept of *divergence on monad* as a quantitative measure of difference between computational effects. This mildly generalizes Barthe and Olmedo’s composable divergences on probability distributions (Barthe and Olmedo, 2013). 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  $(M, \leq, 1, (\cdot))$ .

**Definition 5.** A basic endorelation is a functor  $E : \mathbb{C} \rightarrow \mathbf{BRel}(\mathbb{C})$  such that for each  $I \in \mathbb{C}$ ,  $EI$  is an endorelation.

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

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

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  $E_{\delta}$  parameterized by  $\delta \in \mathcal{Q}$ . It collects all pairs of global elements whose divergence is bound by  $\delta$ . That is,  $E_{\delta}(I, d) \triangleq ((I, d), (I, d), \{(x_1, x_2) \mid d(x_1, x_2) \leq \delta\})$ .
- The category **Pre** of preorders and monotone functions has the basic endorelation  $E_{eq}$  collecting equivalent global elements:  $E_{eq}(I, \leq) \triangleq ((I, \leq), (I, \leq), \{(x, y) \mid x \leq y \wedge y \leq x\})$ .

**Definition 6.** Let  $(\mathbb{C}, 1, (\times), T, \eta, \mu, \theta)$  be a CC-SM,  $\mathcal{Q} = (Q, \leq, 0, (+))$  be a divergence domain,  $(M, \leq, 1, (\cdot))$  be a grading monoid and  $E : \mathbb{C} \rightarrow \mathbf{BRel}(\mathbb{C})$  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  $\Delta = \{\Delta_I^m : (U(TI))^2 \rightarrow \mathcal{Q}\}_{m \in M, I \in \mathbb{C}}$  satisfying the following conditions:

- (Monotonicity) For any  $m \leq m'$  in  $M$  and  $I \in \mathbb{C}$ ,

$$\Delta_I^m \leq_{TI} \Delta_I^{m'} \quad (\text{see Definition 2 for } \leq).$$

- (*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  $f_1, f_2: I \rightarrow TJ$ ,

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

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  $\leq$  (reusing the notation for the partial order between divergences in Definition 2) on  $\mathbf{Div}(T, E, M, \mathcal{Q})$  by:

$$\Delta_1 \leq \Delta_2 \iff \forall m \in M, I \in \mathbb{C}. (\Delta_1)_I^m \leq_{TI} (\Delta_2)_I^m.$$

The *E*-composability condition is a generalization of the composability of differential privacy stated as Theorem 1 in Barthe and Olmedo (2013). What is new in the present 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 Barthe and Olmedo (2013) only considers the case where  $E = \text{Eq}$ . 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).

We end this section by stating an interaction between the strength of *T* and divergences on *T*.

**Proposition 7.** *Let  $(\mathbb{C}, T)$  be a CC-SM,  $E: \mathbb{C} \rightarrow \mathbf{BRel}(\mathbb{C})$  be a basic endorelation,  $(M, \leq, 1, (\cdot))$  be a grading monoid, and  $\mathcal{Q}$  be a divergence domain. Suppose also that  $EI \dot{\times} EJ \leq E(I \times J)$  holds for all  $I, J \in \mathbb{C}$ . Then each divergence  $\Delta \in \mathbf{Div}(T, E, M, \mathcal{Q})$  satisfies: for all  $m \in M, (x_1, x_2) \in EI$  and  $c_1, c_2 \in U(TJ)$ ,*

$$\Delta_{I \times J}^m(\theta_{I,J} \bullet \langle x_1, c_1 \rangle, \theta_{I,J} \bullet \langle x_2, c_2 \rangle) \leq \Delta_J^m(c_1, c_2).$$

## 5. Examples of Divergences on Monads

### 5.1 Cost difference for deterministic computations

We introduce examples of divergence on the *cost count monad*  $T \triangleq \mathbb{N} \times -$  on **Set** (which is isomorphic to the writer monad over a single alphabet  $\{*\}$ ). The divergence is inspired by the work on relational cost analysis (Çiçek et al., 2017; Radiček et al., 2017) measuring the difference of costs between two programs by *subtraction*. Through these examples we also discuss the roles of the *E-unit reflexivity* and *E-composability* conditions.

The unit and Kleisli extension of the cost count monad *T* are defined by:

$$\eta_I(x) \triangleq (0, x) \quad f^\#(i, x) \triangleq (i + \pi_1(f(x)), \pi_2(f(x))) \quad (x \in I, i \in \mathbb{N}, f: I \rightarrow TJ).$$

This monad *T* can be used to record the cost incurred by deterministic computations. For instance, consider the quicksort 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  $\llbracket \text{qsort} \rrbracket, \llbracket \text{isort} \rrbracket: \mathbb{N}^* \rightarrow T(\mathbb{N}^*)$  so that the first component of  $\llbracket \text{qsort} \rrbracket(x)$  and that of  $\llbracket \text{isort} \rrbracket(x)$  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:

$$C_I((i, x), (j, y)) \triangleq |i - j|.$$

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

$\Delta \in \mathbf{Div}(T, \mathbf{Top}, 1, \mathcal{Q})$	$T$	$\mathcal{Q}$	Definition of $\Delta_I(c_1, c_2)$
C	$\mathbb{N} \times -$	$\mathcal{N}$	$C_I((i, x), (j, y)) =  i - j $
NCI	$P(\mathbb{N} \times -)$	$\mathcal{Z}$	$\text{NCI}_I(A, B) = \sup_{(i,x) \in A, (j,y) \in B} i - j$
NC	$P(\mathbb{N} \times -)$	$\mathcal{N}$	$\text{NC}_I(A, B) = \sup_{(i,x) \in A, (j,y) \in B}  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  $C = C_{II \in \mathbf{Set}}$  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:

$$C_I(\eta_I(x), \eta_I(y)) = C_I((0, x), (0, y)) = 0.$$

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  $f, g: I \rightarrow TJ$ . The latter is measured by taking the sup of cost difference of  $f(x)$  and  $g(y)$ , where  $(x, y)$  range over the basic correlation  $\mathbf{Top} I$ .

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

We remark that  $C$  is *not* an Eq-relative  $\mathcal{N}$ -divergence on  $T$  because the Eq-composability fails: If  $I = \{x, y, z\}, J = \{v, w\}$  and  $f: I \rightarrow CJ$  is defined by  $f(x) = (0, w), f(y) = (1, w)$  and  $f(z) = (0, v)$ , then we have  $C_I((0, x), (0, y)) = 0$  and  $\sup_{(x,y) \in \text{Eq} I} C_I(f(x), f(y)) = 0$ , but we have  $C_I(f^\sharp(0, x), f^\sharp(0, y)) = C_I((0, w), (1, w)) = 1$ .

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

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

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  $\infty$ . This divergence is an Eq-relative  $\mathcal{N}$ -divergence on  $T$ .

**5.2 Cost difference for nondeterministic computations**

We may model the cost counting effect and nondeterministic choice by the monad  $P(\mathbb{N} \times -)$  on  $\mathbf{Set}$ , where  $P$  is the powerset monad. We extend the divergence on the cost count monad in the previous section to this combined monad as follows. For two results of nondeterministic computations  $A, B \in P(\mathbb{N} \times I)$  with cost counting effects, the least upper bound of the difference  $i - j$  for all possible choices of  $(i, x) \in A$  and  $(j, y) \in B$  forms the divergence on  $P(\mathbb{N} \times -)$ :

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

If either  $A$  or  $B$  is empty, we fail to get an information of costs. We then have  $\text{NCI}_I(A, B) = -\infty$ . Similarly, we can define the divergence  $\text{NC}$  in Table 1 measuring the *absolute difference* of costs.



**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  $c: I \rightarrow GJ$ , 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* (Giry, 1982); see also Section 2.1).

**Definition 8.** Differential privacy, (Dwork et al., 2006). Let  $c: I \rightarrow GJ$  be a morphism in **Meas**, representing a randomized query. The query  $c$  satisfies  $(\epsilon, \delta)$ -differential privacy ( $\epsilon, \delta \geq 0$  are reals) if for any adjacent datasets  $(d_1, d_2) \in R_{adj}^2$ , the following holds

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

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

$$DP_J^\epsilon(\mu_1, \mu_2) \triangleq \sup_{S \in \Sigma_J} (\mu_1(S) - \exp(\epsilon)\mu_2(S)) \quad (\mu_1, \mu_2 \in GJ).$$

Then the query  $c: I \rightarrow GJ$  satisfies  $(\epsilon, \delta)$ -DP if and only if

$$\forall (d_1, d_2) \in R_{adj}. DP_J^\epsilon(c(d_1), c(d_2)) \leq \delta.$$

The pair  $(\epsilon, \delta)$  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  $\epsilon$  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  $\epsilon$  is large, attackers can distinguish the datasets  $d_1$  and  $d_2$  from the outputs of the query  $c$ . The parameter  $\delta$  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  $G$  (Sato et al., 2019, Lemma 6). This is proved by extending the composability of the divergence for DP on discrete probability distributions shown as Lemmas 3 and 6 in Barthe et al. (2012) and Proposition 5 in Barthe and Olmedo (2013), based on the *composition theorem* of DP (Dwork and Roth, 2013, 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  $(\epsilon, \delta)$ -DP and  $\epsilon \leq \epsilon'$  and  $\delta \leq \delta'$  holds, then  $c$  satisfies  $(\epsilon', \delta')$ -DP.
- (Eq-unit reflexivity) The Eq-unit reflexivity of DP implies  $DP_J^0(\eta_J \circ h(x), \eta_J \circ h(x)) = 0$  for any measurable function  $h: I \rightarrow J$  and  $x \in I$ . This, together with the composability below, ensures the *robustness* of DP of a query  $c: I \rightarrow GJ$  with respect to deterministic postprocessing:

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

In fact, the divergence DP is reflexive: we have  $DP_J^0(\mu, \mu) = 0$  for every  $\mu \in GJ$ . Therefore,  $h: J \rightarrow K$  and  $Gh$  in (2) can be replaced by  $h: J \rightarrow GK$  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* (Dwork and Roth, 2013). If  $c_1: I \rightarrow GJ'$  and  $c_2: J' \rightarrow GJ$  are  $(\epsilon_1, \delta_1)$ -DP and  $(\epsilon_2, \delta_2)$ -DP, respectively, then the sequential composition  $c_2^\sharp \circ c_1: I \rightarrow GJ$  of the queries  $c_1$  and  $c_2$  is  $(\epsilon_1 + \epsilon_2, \delta_1 + \delta_2)$ -DP.

**A non-example: Pointwise differential privacy.** We stated above that a parameter  $(\epsilon, \delta)$  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  $(\epsilon, \delta)$ -DP and this intuition of  $\epsilon$  and  $\delta$ . *Pointwise differential privacy* in Prasad and Smith (2014, Definition 3.2) and Hall (2012, Proposition 1.2.3) is a finer definition of DP that is faithful to this intuition.

**Definition 9.** A measurable function  $c: I \rightarrow GJ$  (regarded as a query) is pointwise  $(\epsilon, \delta)$ -differentially private if whenever  $d_1$  and  $d_2$  are adjacent, there exists a measurable subset  $A \in \Sigma_J$  satisfying  $\Pr [c(d_1) \notin A] \leq \delta$  and the following inequality:

$$\forall s \in A. \Pr [c(d_1) = s] \leq \exp(\epsilon) \Pr [c(d_2) = s],$$

which is equivalent to<sup>3</sup>

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

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

$$\text{pwDP}_J^\epsilon(\mu_1, \mu_2) \triangleq \inf \left\{ \mu_1(J \setminus A) \mid A \in \Sigma_X \wedge (\forall S \in \Sigma_J. S \subseteq A \implies \mu_1(S) \leq \exp(\epsilon) \mu_2(S)) \right\}.$$

Then,  $c: I \rightarrow GJ$  is pointwise  $(\epsilon, \delta)$ -differentially private if and only if

$$\forall (d_1, d_2) \in R_{\text{adj}}. \text{pwDP}_J^\epsilon(c(d_1), c(d_2)) \leq \delta.$$

The family  $\text{pwDP}$  is obviously reflexive:  $\text{pwDP}_J^\epsilon(\mu, \mu) = 0$  holds for any  $\mu \in GJ$  and  $\epsilon \geq 0$ . Hence, it is Eq-unit reflexive. However, it is *not* Eq-composable. We let  $I = \{0, 1, 2\}$  and  $J = \{0, 1\}$  be discrete spaces, and let  $\alpha = \exp(\epsilon)$ . We define two probability distributions  $\mu_1, \mu_2 \in GI$  by:

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

We then have  $\text{pwDP}_J^\epsilon(\mu_1, \mu_2) = 1/10$  with  $A = \{1, 2\}$  because

$$\begin{aligned} \mu_1(\{0\}) &= \frac{1}{10} > \exp(\epsilon) \cdot 0 &&= \exp(\epsilon) \mu_2(\{0\}), \\ \mu_1(\{1\}) &= \frac{9}{10} \leq \exp(\epsilon) \cdot \frac{9}{10\alpha} &&= \exp(\epsilon) \mu_2(\{1\}), \\ \mu_1(\{2\}) &= 0 \leq \exp(\epsilon) \cdot \left(1 - \frac{9}{10\alpha}\right) &&= \exp(\epsilon) \mu_2(\{2\}). \end{aligned}$$

Next, we define  $f: I \rightarrow GJ$  by:

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

We then have

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

Hence, we obtain  $\text{pwDP}_J^\epsilon(f^\sharp(\mu_1), f^\sharp(\mu_2)) = 82/100$  with  $A = \{1\}$  because

$$\begin{aligned} f^\sharp(\mu_1)(\{0\}) &= \frac{82}{100} > \exp(\epsilon) \cdot \frac{81}{100\alpha} &&= \exp(\epsilon) f^\sharp(\mu_2)(\{0\}), \\ f^\sharp(\mu_1)(\{1\}) &= \frac{18}{100} \leq \exp(\epsilon) \cdot \left(\frac{100\alpha - 90 + 9}{100\alpha}\right) &&= \exp(\epsilon) f^\sharp(\mu_2)(\{1\}). \end{aligned}$$

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

$\Delta$	$M$	$\mathcal{Q}$	$\mathcal{Q}_s$	Definition of $\Delta_I^m(\mu_1, \mu_2)$	Composability proof
DP	$\mathcal{R}^+$	$\mathcal{R}^+$	$\mathcal{R}^+$	$\sup_{S \in \Sigma_I} (\mu_1(S) - \exp(\varepsilon)\mu_2(S))$	Barthe and Olmedo (2013)
$\alpha$ Re	1	$\mathcal{R}^+$	$\mathcal{R}$	$\frac{1}{\alpha-1} \log \int_I \left(\frac{\mu_1(x)}{\mu_2(x)}\right)^\alpha \mu_2(x) dx$ .	Mironov (2017)
zCDP	$\mathcal{R}^+$	$\mathcal{R}^+$	$\mathcal{R}$	$\sup_{1 < \alpha} \frac{1}{\alpha} (\alpha \text{Re}_I(\mu_1, \mu_2) - m)$	Bun and Steinke (2016)
$w$ tCDP	1	$\mathcal{R}^+$	$\mathcal{R}$	$\sup_{1 < \alpha < w} \frac{1}{\alpha} (\alpha \text{Re}_I(\mu_1, \mu_2))$	Bun et al. (2018)

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

Name	$\Delta$	$\mathcal{Q}$	$\mathcal{Q}_s$	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	$\mathcal{R}^+$	?	$\frac{1}{2} \int_I \left(\sqrt{\mu_1(x)} - \sqrt{\mu_2(x)}\right)^2 dx$
$\chi^2$ -divergence	Chi	$\mathcal{R}_1^+$	?	$\int_I \frac{(\mu_1(x) - \mu_2(x))^2}{\mu_2(x)} dx$

By the reflexivity of pwDP, we have  $\sup_{(x,y) \in \text{Eq} I} \text{pwDP}_I^0(f(x), f(y)) = 0$ . Therefore, we have obtained a counterexample to the Eq-composability of pwDP:

$$\text{pwDP}_I^\varepsilon(f^\sharp(\mu_1), f^\sharp(\mu_2)) = \frac{82}{100} > \frac{1}{10} = \text{pwDP}_I^\varepsilon(\mu_1, \mu_2) + \sup_{(x,y) \in \text{Eq} I} \text{pwDP}_I^0(f(x), f(y)).$$

**Various relaxations of differential privacy.** Since the seminal work on DP by Dwork et al. (2006), various relaxations of differential privacy have been proposed: Rényi DP (Mironov, 2017), zero-concentrated DP (Bun and Steinke, 2016), and truncated zero-concentrated DP (Bun et al., 2018). 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. Therefore,  $\alpha, w \in (1, \infty)$  are nongrading parameters for Re and tCDP. Each row of the table represents that  $\Delta$  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 composability 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  *$\chi^2$ -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 Barthe and Olmedo (2013) and Olmedo (2014). Later, Sato et al. (2019) 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  ${}^f\text{Div}$  (Csiszár, 1963, 1967; Morimoto, 1963):

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

Here,  $f$  is a parameter called *weight function* and has to be a convex function  $f: [0, \infty) \rightarrow \mathbb{R}$ , continuous at 0 and satisfying  $\lim_{x \rightarrow +0} xf(x) = 0$ . To support general  $\mu_1, \mu_2 \in G_s I$ , we suppose  $af(0/a) = af^*(0)$  for  $a \in [0, \infty)$  where  $f^*(0) \triangleq \lim_{x \rightarrow \infty} f(x)/x$  (see also Liese and Vajda (2006, Definition 2)). Weight functions for four divergences TV, KL, HD, Chi are in Table 4. In fact,  $\text{DP}^\varepsilon$

Table 4. Parameters for Proposition 10

${}^f\text{Div}$	Weight function $f$	$\gamma$	$\alpha$	$\beta$	$\beta'$
TV	$f(t) =  t - 1 /2$	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	$f(t) = (t - 1)^2$	1	-2	2	2

is also an  $f$ -divergence with weight function  $f(t) = \max(0, t - \exp(\varepsilon))$ ; see Barthe and Olmedo (2013, Proposition 2). We also remark that Rényi divergence  ${}^\alpha\text{Re}$  of order  $\alpha$  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 (Csiszár, 1967; Liese and Vajda, 2006). However, the Eq-composability of  $f$ -divergences is not guaranteed in general. Here, we provide a sufficient condition for the Eq-composability of  ${}^f\text{Div}$  over a specific form of divergence domain.

**Proposition 10.** *Let  $\gamma \geq 0$  be a nonnegative real number,  $\mathcal{R}_\gamma^+ = ([0, \infty], \leq, 0, \lambda(p, q) \cdot p + q + \gamma pq)$  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*

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

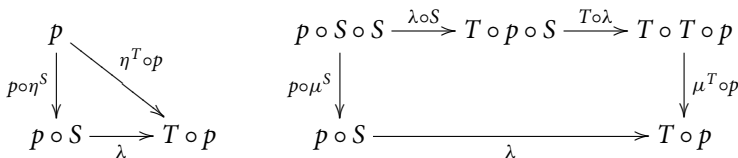
then  ${}^f\text{Div}$  is an Eq-relative  $\mathcal{R}_\gamma^+$ -divergence on the Giry monad  $G$ . When  $\alpha = 0$  and  $\beta, \beta' \in [0, 1]$ ,  $G$  can be replaced with the sub-Giry monad  $G_s$ .

The proof of this proposition generalizes and integrates the proofs given in Olmedo (2014, 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 2.1). 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 (Street, 1972, Section 4) is a functor  $p: \mathbb{C} \rightarrow \mathbb{D}$  together with a natural transformation  $\lambda: p \circ S \rightarrow T \circ p$  making the following diagrams commute:



**Proposition 11.** *Let  $(\mathbb{C}, S), (\mathbb{D}, T)$  be two CC-SMs,  $(p: \mathbb{C} \rightarrow \mathbb{D}, \lambda: p \circ S \rightarrow T \circ p)$  be a monad opfunctor and assume that  $U^{\mathbb{D}} \circ p = U^{\mathbb{C}}$  holds, and basic endorelations  $F: \mathbb{C} \rightarrow \mathbf{BRel}(\mathbb{C})$  and  $E: \mathbb{D} \rightarrow \mathbf{BRel}(\mathbb{D})$  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  $\langle p, \lambda \rangle^* \Delta = \{ \langle \langle p, \lambda \rangle^* \Delta \rangle_I^m \}_{m \in M, I \in \mathbb{C}}$  on  $S$  is an  $F$ -relative  $M$ -graded  $\mathcal{Q}$ -divergence on  $S$ :*

$$\langle \langle p, \lambda \rangle^* \Delta \rangle_I^m(v_1, v_2) \triangleq \Delta_{pI}^m(\lambda_I \bullet v_1, \lambda_I \bullet v_2) = \Delta_{pI}^m((U^{\mathbb{D}} \lambda_I)(v_1), (U^{\mathbb{D}} \lambda_I)(v_2)).$$

The left adjoint  $L: \mathbf{QBS} \rightarrow \mathbf{Meas}$  of the adjunction  $L \dashv K: \mathbf{Meas} \rightarrow \mathbf{QBS}$  and the natural transformation  $l: LP \Rightarrow GL$  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  $\mathbf{Meas}$  (Heunen et al., 2017, 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  $\Delta$ . 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  $l_{K\Omega}: LPK\Omega \cong GLK\Omega = G\Omega$  (Heunen et al., 2017, Prop. 22 (4)). A concrete description of its inverse is  $l_{K\Omega}^{-1} \bullet \mu = [\gamma', \mu(\gamma^{-1}(-))]_{\sim_{K\Omega}}$ , where  $\gamma': \mathbb{R} \rightarrow \Omega$  and  $\gamma: \Omega \rightarrow \mathbb{R}$  are a section-retraction pair (i.e.  $\gamma' \circ \gamma = \text{id}_\Omega$ ) that exists for any standard Borel  $\Omega$ .

**Theorem 12.** For any  $\Delta \in \mathbf{Div}(G, \text{Eq}, \mathcal{Q}, M)$  and standard Borel  $\Omega \in \mathbf{Meas}$ ,

$$(\langle p, \lambda \rangle^* \Delta)_m^{L,l} K\Omega (l_{K\Omega}^{-1} \bullet \mu_1, l_{K\Omega}^{-1} \bullet \mu_2) = \Delta_\Omega^m(\mu_1, \mu_2) \quad (\mu_1, \mu_2 \in U(G\Omega)).$$

### 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  $d_S: S^2 \rightarrow [0, \infty]$  satisfying  $d_S(s, s) = 0$ . The following  $\mathcal{R}^\times$ -divergence  $\Delta_I^{\text{lip}, d_S}(f_1, f_2)$  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,  $\Delta^{\text{lip}, d_S}$  measures the Lipschitz constant on state transformers.

**Proposition 13.** The family  $\Delta^{\text{lip}, d_S} = \{\Delta_I^{\text{lip}, d_S}\}_{I \in \mathbf{Set}}$  of  $\mathcal{R}^\times$ -divergences on  $T_S I$  defined by:

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

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  $f_1 = f_2 = g$ ,  $\Delta_I^{\text{lip}, d_S}(g, g)$  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  $\Delta_I^{\text{met}, d_S}(f_1, f_2)$  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 14.** Suppose that the function  $d_S$  also satisfy the triangle inequality. The family  $\Delta^{\text{met}, d_S} = \{\Delta_I^{\text{met}, d_S}\}_{I \in \mathbf{Set}}$  of  $\mathcal{R}^+$ -divergences on  $T_S I$  defined by:

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

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<sup>4</sup> and nonexpansive functions is Cartesian closed (Rutten, 1996, 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 15.** *The family  $\{d_{T_S I} : (T_S(I, d_I))^2 \rightarrow [0, 1]\}_{(I, d_I) \in \mathbf{Gum}}$  consisting of the metric part of the spaces  $T_S(I, d_I)$ , given by:*

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

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

In the category **Gum**, instead of Eq, there is another basic endorelation  $\text{Dist}_0$ :

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

By modifying the divergence  $d_{T_S(-)}$ , we obtain a  $\text{Dist}_0$ -relative  $([0, 1], \leq, \max, 0)$ -divergence as below:

**Proposition 16.** *The following forms a  $\text{Dist}_0$ -relative  $([0, 1], \leq, \max, 0)$ -divergence on  $T_S$ :*

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

5.7 **Combining divergence with cost**

In Section 5.2, we have introduced divergences on the monad  $P(\mathbb{N} \times -)$  modeling nondeterministic choice and cost counting. These divergences are based on the distance/subtractions of costs represented by natural numbers. In this section, we provide an alternative divergences on the combination of a general computational effect  $T$  and cost counting. The basic idea is the following: given two computations  $c_1, c_2 \in T(N \times I)$ , we discard the value part of  $c_i$  by  $T\pi_1 : T(N \times I) \rightarrow T(N)$  and measure their difference by the divergence assumed on  $T$ .

Let  $(\mathbb{C}, T)$  be a CC-SM and  $\Delta \in \mathbf{Div}(T, \text{Eq}, 1, \mathcal{Q})$  be a divergence and  $(N, 1_N : 1 \rightarrow N, (\star) : N \times N \rightarrow N)$  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  $C(\Delta, N) = \{C(\Delta, N)_I : (U(T(N \times I)))^2 \rightarrow \mathcal{Q}\}_{I \in \mathbb{C}}$  of  $\mathcal{Q}$ -divergences by:

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

**Proposition 17.** *The family  $C(\Delta, N)$  is an Eq-relative  $\mathcal{Q}$ -divergence on  $T(N \times -)$ .*

For example, the divergence  $C(\text{KL}, \mathbb{R})$  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  $\text{KL}_{\mathbb{R} \times I}(\mu_1, \mu_2) \leq \text{KL}_{\mathbb{R}}(G\pi_1 \bullet \mu_1, G\pi_1 \bullet \mu_2)$  in the definition of  $C(\text{KL}, \mathbb{R})$  means that the difference between  $\mu_1$  and  $\mu_2$  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  $\Delta = \{\Delta_I : (U(TI))^2 \rightarrow \mathcal{B}\}_{I \in \mathbb{C}}$  of  $\mathcal{B}$ -divergences and a family of adjacency relations  $\tilde{\Delta}(1)I \triangleq \{(c_1, c_2) \mid \Delta_I(c_1, c_2) \leq 1\}_{I \in \mathbf{Set}}$ .

We point out a connection between Eq-relative  $\mathcal{B}$ -divergences and *preorders on monads* studied in Katsumata and Sato (2013) and Sato (2014). A preorder on a monad  $T$  on  $\mathbf{Set}$  assigns a preorder  $\sqsubseteq_I$  on  $TI$  for each  $I \in \mathbf{Set}$ , and this assignment satisfies:

- (Substitutivity) For any function  $f : I \rightarrow TJ$  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  $f_1, f_2 : I \rightarrow TJ$ , 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 18.** *A preorder on a monad  $T$  on  $\mathbf{Set}$  bijectively corresponds to an Eq-relative  $\mathcal{B}$ -divergence  $\Delta$  on  $T$  such that each  $\tilde{\Delta}(1)I$  is a preorder.*

For a preorder  $\sqsubseteq$  on a monad  $T$  on  $\mathbf{Set}$ , by  $\Delta^\sqsubseteq$  we mean the divergence corresponding to  $\sqsubseteq$  by Proposition 18 (in fact, we have  $\tilde{\Delta}^\sqsubseteq(1)I = \sqsubseteq_I$  and  $\tilde{\Delta}^\sqsubseteq(0)I = TI \times TI$  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 the 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:

$$\mathbf{I} \triangleq (1, \lambda(x_1, x_2) \cdot 0),$$

$$(I, d) \otimes (J, e) \triangleq (I \times J, \lambda((x_1, y_1), (x_2, y_2)) \cdot d(x_1, x_2) + e(y_1, y_2)).$$

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

**6.1.1 Enrichments of Kleisli categories induced by divergences**

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

**Definition 19.** *A  $\mathbf{Div}_{\mathcal{Q}}(\mathbf{Set})$ -enrichment of a category  $\mathbb{D}$  is a family  $\{d_{I,J} : \mathbb{D}(I, J)^2 \rightarrow \mathcal{Q}\}_{I, J \in \mathbb{D}}$  of  $\mathcal{Q}$ -divergences on the homset  $\mathbb{D}(I, J)$  such that the following inequalities hold*

$$d_{I,I}(\text{id}_I, \text{id}_I) \leq 0, \tag{3}$$

$$d_{I,K}(g_1 \circ f_1, g_2 \circ f_2) \leq d_{J,K}(g_1, g_2) + d_{I,J}(f_1, f_2). \tag{4}$$

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  $\mathbb{D}^d$ :

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

are inherited from  $\mathbb{D}$ ; they are guaranteed to be nonexpansive by conditions (3) and (4).

We characterize  $\mathbf{Div}_{\mathcal{Q}}(\mathbf{Set})$ -enrichments of  $\mathbb{D}$  as  $\mathbf{Div}_{\mathcal{Q}}(\mathbf{Set})$ -enriched categories whose change-of-enriching-category along  $V_{\mathcal{Q},\mathbf{Set}}$  coincides with  $\mathbb{D}$ .

**Proposition 20.** *Let  $\mathbb{D}$  be a category. There is a bijective correspondence between 1) a  $\mathbf{Div}_{\mathcal{Q}}(\mathbf{Set})$ -enrichment  $\{d_{I,J}\}_{I,J \in \mathbb{D}}$  of  $\mathbb{D}$  and 2) a  $\mathbf{Div}_{\mathcal{Q}}(\mathbf{Set})$ -enriched category  $\mathbb{E}$  such that the change-of-enriching-category of  $\mathbb{E}$  by  $V_{\mathcal{Q},\mathbf{Set}} : \mathbf{Div}_{\mathcal{Q}}(\mathbf{Set}) \rightarrow \mathbf{Set}$  is equal to  $\mathbb{D}$ .*

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

**Theorem 21.** *Let  $(\mathbb{C}, T)$  be a CC-SM,  $E : \mathbb{C} \rightarrow \mathbf{BRel}(\mathbb{C})$  be a basic endorelation such that  $R_{E1} \neq \emptyset$ ,  $\mathcal{Q}$  be a divergence domain and  $\Delta = \{\Delta_I : (U(TI))^2 \rightarrow \mathcal{Q}\}_{I \in \mathbb{C}}$  be a family of  $\mathcal{Q}$ -divergences on  $TI$ . Define a family  $d^\Delta = \{d_{I,J}^\Delta : \mathbb{C}_T(I, J)^2 \rightarrow \mathcal{Q}\}_{I, J \in \mathbb{C}}$  of  $\mathcal{Q}$ -divergences on the homset  $\mathbb{C}_T(I, J)$  of the Kleisli category  $\mathbb{C}_T$  by:*

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

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

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

We seek a further characterization of the  $\mathcal{Q}$ -divergence (5) given to each homset of  $\mathbb{C}_T$ . Under a strengthened assumption, we relate it with the *closed structure* with respect to the tensor 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  $\mathcal{Q} = (Q, \leq, 0, (+))$  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 22.** *Let  $X \in \mathbf{Div}_{\mathcal{Q}}(\mathbb{C})$  be an object such that its  $\mathcal{Q}$ -divergence  $d_X$  takes values in  $\{0, \top\} \subseteq \mathcal{Q}$ . Define a functor  $X \multimap (-) : \mathbf{Div}_{\mathcal{Q}}(\mathbb{C}) \rightarrow \mathbf{Div}_{\mathcal{Q}}(\mathbb{C})$*

$$\begin{aligned} X \multimap Y &\triangleq (V_{\mathcal{Q},\mathbb{C}}X \Rightarrow V_{\mathcal{Q},\mathbb{C}}Y, d_{X \multimap Y}) \\ d_{X \multimap Y}(f_1, f_2) &\triangleq \sup_{x_1, x_2 \in U(V_{\mathcal{Q},\mathbb{C}}X), d_X(x_1, x_2)=0} d_Y([f_1] \bullet x_1, [f_2] \bullet x_2); \end{aligned} \tag{6}$$

here  $[-] : U(I \Rightarrow J) \rightarrow \mathbb{C}(I, J)$  is the bijection given in Section 2. Then it is a left adjoint to the functor  $(-) \otimes X : \mathbf{Div}_{\mathcal{Q}}(\mathbb{C}) \rightarrow \mathbf{Div}_{\mathcal{Q}}(\mathbb{C})$  tensoring with  $X$ . Moreover,  $V_{\mathcal{Q},\mathbb{C}} : \mathbf{Div}_{\mathcal{Q}}(\mathbb{C}) \rightarrow \mathbb{C}$  is a map of adjunction (Mac Lane, 1998, Section IV.7) in the following sense:

$$\begin{array}{ccc} \mathbf{Div}_{\mathcal{Q}}(\mathbb{C}) & \begin{array}{c} \xrightarrow{(-) \otimes X} \\ \xleftarrow{X \multimap (-)} \end{array} & \mathbf{Div}_{\mathcal{Q}}(\mathbb{C}) \\ \downarrow V_{\mathcal{Q},\mathbb{C}} & & \downarrow V_{\mathcal{Q},\mathbb{C}} \\ \mathbb{C} & \begin{array}{c} \xrightarrow{(-) \times V_{\mathcal{Q},\mathbb{C}}X} \\ \xleftarrow{V_{\mathcal{Q},\mathbb{C}}X \Rightarrow (-)} \end{array} & \mathbb{C} \end{array}$$

The  $\mathcal{Q}$ -divergence  $d_{X \multimap Y}$  in (6) is similar to the sup part of the composability condition in Definition 6. We exploit this similarity to express the unit reflexivity and composability conditions of divergence on monad (Definition 6) using the internal hom functor  $X \multimap (-)$ . First, we define



the uncurried bind morphism  $\text{ub}_{I,J} : TI \times (I \Rightarrow TJ) \rightarrow TJ$  by:

$$\text{ub}_{I,J} \triangleq TI \times (I \Rightarrow TJ) \xrightarrow{(\pi_2, \pi_1)} (I \Rightarrow TJ) \times TI \xrightarrow{\theta_{I \Rightarrow TJ, I}} T((I \Rightarrow TJ) \times I) \xrightarrow{\text{ev}^\#} TJ. \tag{7}$$

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

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

**Theorem 23.** Let  $(\mathbb{C}, T)$  be a CCC-SM,  $(M, \leq, 1, (\cdot))$  be a grading monoid,  $\mathcal{Q} = (Q, \leq, 0, (+))$  be a divergence domain whose monoid operation  $(+)$  satisfies  $x + \top = \top$ , and  $E : \mathbb{C} \rightarrow \mathbf{BRel}(\mathbb{C})$  be a basic endorelation. Let  $\Delta = \{\Delta_I^m : U(TI)^2 \rightarrow \mathcal{Q}\}_{m \in M, I \in \mathbb{C}}$  be a doubly indexed family of  $\mathcal{Q}$ -divergences on  $TI$ , regarded as a family of  $\mathbf{Div}_{\mathcal{Q}}(\mathbb{C})$ -objects. Then

- (1)  $\Delta$  satisfies the E-unit reflexivity condition if and only if for any  $I \in \mathbb{C}$ , the following non-expansivity holds on the global element  $[\eta_I] : 1 \rightarrow I \Rightarrow TI$  corresponding to the monad unit:

$$[\eta_I] \in \mathbf{Div}_{\mathcal{Q}}(\mathbb{C})(I, E^{\mathcal{Q}}I \multimap \Delta_I^1).$$

- (2)  $\Delta$  satisfies the E-composability condition if and only if for any  $I, J \in \mathbb{C}$  and  $m, n \in M$ , the following nonexpansivity holds on the uncurried bind morphism  $\text{ub}_{I,J} : TI \times (I \Rightarrow TJ) \rightarrow TJ$ :

$$\text{ub}_{I,J} \in \mathbf{Div}_{\mathcal{Q}}(\mathbb{C})(\Delta_I^m \otimes (E^{\mathcal{Q}}I \multimap \Delta_J^n), \Delta_J^{m \cdot n}).$$

Azevedo de Amorim et al. (2019) formalized families of composable divergences as *parameterized assignments* in *weakly closed monoidal refinements*. Roughly speaking, they adopted the equivalence (2) of Theorem 23 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 assignments are: 1) divergences on monads are defined with respect to basic endorelations, and 2) the underlying category of divergences on monads is any CCs, while parameterized assignments requires a closed structure on their underlying category.

### 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  $\mathbf{Div}_{\mathcal{Q}}(\mathbb{C})$  (Katsumata, 2014; Smirnov, 2008) whose unit and multiplication are inherited from a monad on  $\mathbb{C}$ .

**Definition 24.** Let  $(\mathbb{C}, T)$  be a CC-SM,  $(M, \leq, 1, (\cdot))$  be a grading monoid and  $\mathcal{Q}$  be a divergence domain. An  $M$ -graded  $\mathcal{Q}$ -divergence lifting of  $T$  is a mapping  $\dot{T} : M \times \mathbf{Obj}(\mathbf{Div}_{\mathcal{Q}}(\mathbb{C})) \rightarrow \mathbf{Obj}(\mathbf{Div}_{\mathcal{Q}}(\mathbb{C}))$  such that (below  $V$  stands for the forgetful functor  $V_{\mathcal{Q}, \mathbb{C}} : \mathbf{Div}_{\mathcal{Q}}(\mathbb{C}) \rightarrow \mathbb{C}$ )

- (1)  $V(\dot{T}mX) = T(VX)$
- (2)  $m \leq n$  implies  $\dot{T}mX \preceq_{T(VX)} \dot{T}nX$
- (3)  $\eta_{VX} \in \mathbf{Div}_{\mathcal{Q}}(\mathbb{C})(X, \dot{T}1X)$
- (4)  $\mu_{VX} \in \mathbf{Div}_{\mathcal{Q}}(\mathbb{C})(\dot{T}m(\dot{T}nX), \dot{T}(m \cdot n)X)$ .

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

$$\theta_{VX,J} \in \mathbf{Div}_{\mathcal{Q}}(\mathbb{C})(X \otimes \dot{T}m(E^{\mathcal{Q}}J), \dot{T}m(X \otimes E^{\mathcal{Q}}J)).$$

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  $\preceq$  (reusing the notation for the partial order between divergences in Definition 2) on  $\mathbf{SGDLift}(T, E, M, \mathcal{Q})$  by:

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

We will later see a similar concept of *strong graded relational lifting* of monad in Definition 37. Divergence liftings and relational liftings are actually instances of a common general definition of *strong graded lifting of monad* (Katsumata, 2014), but in this paper we omit this general definition.

In the following theorem, we show that every divergence on a monad can be expressed as the composite of a graded divergence lifting and the divergence corresponding to a basic endorelation.

**Theorem 25.** Let  $(\mathbb{C}, T)$  be a CC-SM,  $(M, \leq, 1, (\cdot))$  be a grading monoid,  $\mathcal{Q}$  be a divergence domain, and  $E : \mathbb{C} \rightarrow \mathbf{BRel}(\mathbb{C})$  be a basic endorelation. For any  $\Delta \in \mathbf{Div}(T, E, M, \mathcal{Q})$ , define a mapping  $[\Delta] : M \times \mathbf{Obj}(\mathbf{Div}_{\mathcal{Q}}(\mathbb{C})) \rightarrow \mathbf{Obj}(\mathbf{Div}_{\mathcal{Q}}(\mathbb{C}))$  by:

$$[\Delta]mX \triangleq (TI, d_{[\Delta]mX}) \quad (X = (I, d))$$

where

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

Then  $[\Delta]$  is an  $M$ -graded  $\mathcal{Q}$ -divergence lifting such that  $\Delta_I^m = [\Delta]m(E^{\mathcal{Q}}I)$ .

When  $M = 1$ , Theorem 25 implies that the assignment  $I \mapsto \Delta_I$  extends to the  $E^{\mathcal{Q}}$ -relative monad  $[\Delta] \circ : E^{\mathcal{Q}}\mathbb{C} \rightarrow \mathbf{Div}_{\mathcal{Q}}(\mathbb{C})$  in the sense of Altenkirch et al. (2015).

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 26.** 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  $E : \mathbb{C} \rightarrow \mathbf{BRel}(\mathbb{C})$  be a basic endorelation. Then there exists an adjunction between partial orders:

$$(\mathbf{SGDLift}(T, E, M, \mathcal{Q}), \preceq) \begin{matrix} \xleftarrow{\langle - \rangle} \\ \perp \\ \xrightarrow{[-]} \end{matrix} (\mathbf{Div}(T, E, M, \mathcal{Q}), \preceq)$$

where  $\langle \dot{T} \rangle mI \triangleq \dot{T}m(E^{\mathcal{Q}}I)$ .

### 6.2 Generation of divergences

It has been shown that DP can be interpreted as hypothesis testing (Kairouz et al., 2015; Wasserman and Zhou, 2010). Given a query  $c : I \rightarrow GJ$  and adjacent datasets  $(d_1, d_2) \in R_{\text{adj}} \subseteq I^2$ , 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. Kairouz et al. (2015) showed that  $c$  is  $(\epsilon, \delta)$ -DP if and only if for any adjacent datasets  $(d_1, d_2) \in R_{\text{adj}} \subseteq I^2$ , the pair of Type I error and Type II error lands in the privacy region  $R(\epsilon, \delta)$ :

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

They also showed that this is equivalent to the testing using probabilistic decision rules (Kairouz et al., 2015, Corollary 2.3):

$$\forall k: J \rightarrow G\{\text{Acc}, \text{Rej}\} . (\Pr [k^\sharp c(d_1) = \text{Acc}], \Pr [k^\sharp c(d_2) = \text{Rej}]) \in R(\epsilon, \delta) .$$

Later Balle et al. (2020) 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  $\Omega$ -generatedness of divergences on monads.

**Definition 27.** Let  $\Omega \in \mathbb{C}$ . A divergence  $\Delta \in \mathbf{Div}(T, E, M, \mathcal{Q})$  is  $\Omega$ -generated if for any  $m \in M$ ,  $I \in \mathbb{C}$  and  $c_1, c_2 \in U(TI)$ ,

$$\Delta_I^m(c_1, c_2) = \sup_{k: I \rightarrow T\Omega} \Delta_\Omega^m(k^\sharp \bullet c_1, k^\sharp \bullet c_2) .$$

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

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

Here,  $\tilde{\Delta}(m, v)\Omega$  is the binary relation  $\{(c_1, c_2) \mid \Delta_\Omega^m(c_1, c_2) \leq v\}$ ; see also (14). For an  $\Omega$ -generated divergence  $\Delta$ , its component  $\Delta_\Omega^m$  at  $\Omega$  is an essential part that determines all components  $\Delta_I^m$  of  $\Delta$ . When a divergence is shown to be  $\Omega$ -generated, the calculation of the codensity lifting  $T^{[\Delta]}$  given in Section 7 will be simplified (Section 7.1).

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

- Divergence DP is generated over the two-point discrete space 2 (Balle et al., 2020, Section B.7). The binary relation  $(\tilde{\text{DP}}(\epsilon, \delta)2)$  coincides with the privacy region  $R(\epsilon, \delta)$ .
- Divergence TV is also generated over 2 (Balle et al., 2020, Section C.1).
- Divergences  $\text{Re}^\alpha$ , 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  $N$  (Balle et al., 2020, 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 28.** The divergence  $\text{DP} \in \mathbf{Div}(G_s, \text{Eq}, \mathcal{R}^+, \mathcal{R}^+)$  is 1-generated.

**Proposition 29.** The divergence  $\text{TV} \in \mathbf{Div}(G_s, \text{Eq}, 1, \mathcal{R}^+)$  is not 1-generated but 2-generated.

**$\Omega$ -Generatedness of preorders on monads.** We relate  $\Omega$ -generatedness of divergences and preorders on monads studied in Katsumata and Sato (2013). Let  $T$  be a monad on  $\mathbf{Set}$  and  $\Omega$  be a set. Katsumata and Sato (2013) introduced the concept of congruent and substitutive preorders on  $T\Omega$  as those satisfying:

- (Substitutivity) For any function  $f: \Omega \rightarrow T\Omega$  and  $c_1, c_2 \in T\Omega$ ,  $c_1 \leq c_2$  implies  $f^\sharp(c_1) \leq f^\sharp(c_2)$ .
- (Congruence) For any function  $f_1, f_2: J \rightarrow T\Omega$ , if  $f_1(x) \leq f_2(x)$  holds for any  $x \in J$ , then  $f_1^\sharp(c) \leq f_2^\sharp(c)$  holds for any  $c \in T\Omega$ .

For instance, any component of a preorder on  $T$  at  $\Omega$  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  $T$ . Katsumata and Sato (2013) gave a construction  $[-]^\Omega: \mathbf{CSPre}(T, \Omega) \rightarrow \mathbf{Pre}(T)$  of preorders on  $T$  from congruent and substitutive preorders on  $T\Omega$ :

$$c_1 [ \leq ]_J^\Omega c_2 \iff \forall g: J \rightarrow T\Omega . g^\sharp(c_1) \leq g^\sharp(c_2)$$

The constructed preorders on  $T$  are  $\Omega$ -generated in the following sense:

**Proposition 30.** *For any  $\leq \in \mathbf{CSPre}(T, \Omega)$ , the  $\mathcal{B}$ -divergence  $\Delta^{[\leq]^\Omega}$  corresponding to the preorder  $[\leq]^\Omega$  on  $T$  is  $\Omega$ -generated (see Proposition 18 for the correspondence).*

Applying this proposition, we can determine  $\Omega$ -generatedness of preorders on monads:

- If the monad  $T$  has a rank  $\alpha$ , the construction  $[-]^\alpha$  is bijective (Katsumata and Sato, 2013, Theorem 7). Hence for such a monad, each preorder on  $T$  corresponds to an  $\alpha$ -generated  $\mathcal{B}$ -divergence.
- For the subprobability distribution monad  $D_s$  on  $\mathbf{Set}$ , Sato (2014) 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 (Sato, 2014, Proposition 6.3).

### 6.3 An adjunction between quantitative equational theories and divergences

Mardare et al. (2016) introduced a concept of *quantitative equational theory* as an algebraic presentation of monads on the category of pseudometric 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  $\Sigma$ -terms.

Consider a set  $\Sigma$  of function symbols of finite arity. If  $n$  is the arity of a function  $f \in \Sigma$ , we write  $f: n \in \Sigma$ . Let  $\Omega$  be a set of variables, and let  $T_\Sigma\Omega$  be the  $\Sigma$ -term algebra over  $\Omega$ . For  $f: n \in \Sigma$  and  $t_1, \dots, t_n \in T_\Sigma\Omega$ , we write  $f(t_1, \dots, t_n)$  for the term obtained by applying  $f$  to  $t_1, \dots, t_n$ . The construction  $\Omega \mapsto T_\Sigma\Omega$  forms a (strong) monad on  $\mathbf{Set}$  whose unit is given by  $\eta_\Omega(x) = x$ , and whose Kleisli extension  $h^\sharp: T_\Sigma I \rightarrow T_\Sigma\Omega$  of function  $h: I \rightarrow T_\Sigma\Omega$  is given 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  $\Sigma$ -terms over  $\Omega$  is a function  $\sigma: \Omega \rightarrow T_\Sigma\Omega$ . For  $t \in T_\Sigma\Omega$ , we call  $\sigma^\sharp(t)$  the substitution of  $\sigma$  to  $t$ . We define the set of *indexed equations* of terms by:

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

Here, the index  $\varepsilon$  runs over nonnegative rational numbers. A *conditional quantitative equation* is a judgment of the following form:

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

the left-hand side of turnstile ( $\vdash$ ) is called hypothesis and the right-hand side conclusion. By  $\mathbb{E}(T_\Sigma\Omega)$ , we mean the set of conditional quantitative equations. For any countable subset  $\Gamma$  of  $\mathbb{V}(T_\Sigma\Omega)$  and any substitution  $\sigma: \Omega \rightarrow T_\Sigma\Omega$ , we define  $\sigma(\Gamma) \triangleq \{\sigma^\sharp(t) =_\varepsilon \sigma^\sharp(u) \mid t =_\varepsilon u \in \Gamma\}$ .

$$\begin{aligned}
 \emptyset \vdash t =_0 t \in U & \quad \text{(Ref)} \\
 \{t =_\varepsilon u\} \vdash u =_\varepsilon t \in U & \quad \text{(Sym)} \\
 \{t =_\varepsilon u, u =_{\varepsilon'} v\} \vdash t =_{\varepsilon+\varepsilon'} v \in U & \quad \text{(Tri)} \\
 \forall \varepsilon' \in \mathbb{Q}^+ . \{t =_\varepsilon u\} \vdash t =_{\varepsilon+\varepsilon'} u \in U & \quad \text{(Max)} \\
 \forall \varepsilon \in \mathbb{Q}^+ . \{t =_{\varepsilon'} u \mid \varepsilon < \varepsilon'\} \vdash t =_\varepsilon u \in U & \quad \text{(Arch)} \\
 \forall f : n \in \Sigma . \{t_i =_\varepsilon u_i \mid 1 \leq i \leq n\} \vdash f(t_1, \dots, t_n) =_\varepsilon f(u_1, \dots, u_n) \in U & \quad \text{(Nonexp)} \\
 \forall \sigma : \Omega \rightarrow T_\Sigma \Omega . \Gamma \vdash t =_\varepsilon u \in U \implies \sigma(\Gamma) \vdash \sigma^\sharp(t) =_\varepsilon \sigma^\sharp(u) \in U & \quad \text{(Subst)} \\
 \Gamma' \vdash t =_\varepsilon u \in U \wedge \forall \psi \in \Gamma' . \Gamma \vdash \psi \in U \implies \Gamma \vdash t =_\varepsilon u \in U & \quad \text{(Cut)} \\
 t =_\varepsilon u \in \Gamma \implies \Gamma \vdash t =_\varepsilon u \in U & \quad \text{(Assumpt)}
 \end{aligned}$$

Figure 1. Quantitative equational theory rules.

**Definition 31.** Quantitative Equational Theory (Mardare et al., 2016, Definition 2.1). A quantitative equational theory (QET for short) of type  $\Sigma$  over  $\Omega$  is a set  $U \subseteq \mathbb{E}(T_\Sigma \Omega)$  closed under the rules summarized as Figure 1. We write  $\mathbf{QET}(\Sigma, \Omega)$  for the set of QETs of type  $\Sigma$  over  $\Omega$ . We regard it as a poset  $(\mathbf{QET}(\Sigma, \Omega), \subseteq)$  by the set inclusion order. Given a set  $U_0$  of conditional quantitative equations of type  $\Sigma$  over  $\Omega$ , by  $\overline{U_0}^{\mathbf{QET}(\Sigma, \Omega)}$  we mean the least QET of type  $\Sigma$  over  $\Omega$  including  $U_0$ .

The goal of this section is to establish an adjunction between the poset of quantitative equational theories and the poset of pseudometrics<sup>6</sup> on free-algebra monads on **Set**. More specifically, we construct the following adjunction and isomorphism between posets:

$$(\mathbf{QET}(\Sigma, \Omega), \subseteq) \xrightleftharpoons[\perp]{D[-]} (\mathbf{CSPMet}(T_\Sigma, \Omega), \leq) \xrightleftharpoons[\cong]{\text{Gen}} (\mathbf{PMet}(T_\Sigma, \Omega), \leq), \quad (8)$$

which are subsequently defined (Definition 32 for **CSPMet**, Definition 33 for **PMet** and equations (9)–(12) for morphisms).

The poset in the middle is that of congruent and substitutive pseudometrics, which are a quantitative analog of congruent and substitutive preorders appeared in Section 6.2.

**Definition 32.** Let  $T$  be a monad on **Set** and  $\Omega \in \mathbf{Set}$ . A congruent and substitutive pseudometric (CS-PMet for short) on  $T\Omega$  is a pseudometric  $d : (T\Omega)^2 \rightarrow \mathcal{R}^+$  on  $T\Omega$  satisfying

- (Substitutivity) For all functions  $f : \Omega \rightarrow T\Omega$  and  $c_1, c_2 \in T\Omega$ ,  $d(f^\sharp(c_1), f^\sharp(c_2)) \leq d(c_1, c_2)$ .
- (Congruence) For all sets  $I$ , functions  $f_1, f_2 : I \rightarrow T\Omega$  and  $c \in TI$ ,  $d(f_1^\sharp(c), f_2^\sharp(c)) \leq \sup_{i \in I} d(f_1(i), f_2(i))$ .

By  $\mathbf{CSPMet}(T, \Omega)$ , we mean the set of CS-PMets on  $T\Omega$ . We then make it into a poset  $(\mathbf{CSPMet}(T, \Omega), \leq)$  where  $\leq$  is the restriction of the partial order  $\leq_{T\Omega}$  in Definition 2 to CS-PMets.

**Definition 33.** Let  $T$  be a monad on **Set** and  $\Omega \in \mathbf{Set}$ . By  $\mathbf{PMet}(T, \Omega)$  we mean the collection of  $\Omega$ -generated Eq-relative  $\mathcal{R}^+$ -divergences  $\Delta$  on  $T$  such that each component  $\Delta_I$  is a pseudometric. We restrict the partial order  $\leq$  on  $\mathbf{Div}(T, \text{Eq}, 1, \mathcal{R}^+)$  to  $\mathbf{PMet}(T, \Omega)$ .

We next introduce monotone functions appearing in (8):

$$D[U](t, u) \triangleq \inf \{ \varepsilon \in \mathbb{Q}^+ \mid \emptyset \vdash t =_\varepsilon u \in U \} \quad (9)$$

$$U[d] \triangleq \overline{\{\emptyset \vdash t =_\varepsilon u \mid \varepsilon \in \mathbb{Q}^+, d(t, u) \leq \varepsilon\}}^{\text{QET}(\Sigma, \Omega)} \quad (10)$$

$$\text{Gen}(d)_I(c_1, c_2) \triangleq \sup_{k: I \rightarrow T\Omega} d(k^\sharp(c_1), k^\sharp(c_2)) \quad (11)$$

$$(\Delta)_\Omega \triangleq \Delta_\Omega \quad (12)$$

**Proposition 34.** *The functions  $D[-]$ ,  $U[-]$ ,  $\text{Gen}$ ,  $(-)_\Omega$  defined as (9)-(12) are all well-defined monotone functions having types given in (8).*

That  $D[U]$  is a pseudometric is shown in the beginning of Mardare et al. (2016, Section 5). Here, we additionally show that it enjoys congruence and substitutivity of Definition 32. The function  $\text{Gen}$  is taken from the right-hand side of the definition of  $\Omega$ -generatedness (Definition 27). The function  $(-)_\Omega$  simply extracts the  $\Omega$  component of a given divergence.

**Theorem 35.** *For any set  $\Sigma$  of function symbols with finite arity and set  $\Omega$ , the following holds for the monotone functions in (8):*

- (1) *Gen is the inverse of  $(-)_\Omega$ .*
- (2) *We have an adjunction  $U[-] \dashv D[-]$  satisfying  $D[U[-]] = \text{id}$ :*

$$(\text{QET}(\Sigma, \Omega), \subseteq) \xrightleftharpoons[U[-]]{D[-]} (\text{CSPMet}(T_\Sigma, \Omega), \leq). \quad (13)$$

Intuitively, the right adjoint  $D[-]$  extracts the pseudometric on  $T_\Sigma \Omega$  from a given QET. The left adjoint  $U[-]$  constructs the least QET containing all information of a given pseudometric on  $T_\Sigma \Omega$ . The range of  $U[-]$  is characterized as the set of *unconditional QETs* of type  $\Sigma$  over  $\Omega$  defined below (see also Mardare et al. (2017, Section 3)):

$$\text{UQET}(\Sigma, \Omega) \triangleq \left\{ S^{\text{QET}(\Sigma, \Omega)} \mid S \subseteq \{\emptyset \vdash t =_\varepsilon u \mid t, u \in T_\Sigma \Omega, \varepsilon \in \mathbb{Q}^+\} \right\}.$$

Therefore, the adjunction (13) cuts down to the following isomorphism between posets, stating that unconditional QETs of type  $\Sigma$  over  $\Omega$  are *equivalent* to  $\Omega$ -generated pseudometrics on  $T_\Sigma$ :

**Theorem 36.**  $(\text{UQET}(\Sigma, \Omega), \subseteq) \cong (\text{CSPMet}(T_\Sigma, \Omega), \leq) \cong (\text{PMet}(T_\Sigma, \Omega), \leq).$

### 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 (Barthe et al., 2012), then later used in various program logics (Barthe et al., 2014, 2015; Barthe and Olmedo, 2013; Sato, 2016; Sato et al., 2019). Independently, it is also introduced as a semantic structure for effect system (Katsumata, 2014). Liftings introduced in the study of differential privacy are designed to satisfy a special property called *fundamental property* (Barthe et al., 2012, 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 (Gabboardi et al., 2021; Katsumata, 2014).

**Definition 37.** 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  $\dot{T} : M \times \mathbf{Obj}(\mathbf{BRel}(\mathbb{C})) \rightarrow \mathbf{Obj}(\mathbf{BRel}(\mathbb{C}))$  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)  $(\eta_{X_1}, \eta_{X_2}) : X \dot{\rightarrow} \dot{T}1(X)$ .
- (3)  $(f_1, f_2) : X \dot{\rightarrow} \dot{T}m(Y)$  implies  $(f_1^\sharp, f_2^\sharp) : \dot{T}m'X \dot{\rightarrow} \dot{T}(m \cdot m')Y$ .
- (4)  $(\theta_{X_1, Y_1}, \theta_{X_2, Y_2}) : X \dot{\times} \dot{T}mY \dot{\rightarrow} \dot{T}m(X \dot{\times} Y)$ .

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 formulate such liftings by the following *fundamental property*. First, we define the *adjacency relation* of  $\Delta$  by:

$$\tilde{\Delta}(m, v)I \triangleq (TI, TI, \{(c_1, c_2) \mid \Delta_I^m(c_1, c_2) \leq v\}) \quad (m \in M, v \in \mathcal{Q}, I \in \mathbb{C}). \tag{14}$$

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

**Definition 38.** 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 39.** Let  $(\mathbb{C}, T)$  be a CC-SM,  $(M, \leq, 1, (\cdot))$  be a grading monoid,  $\mathcal{Q} = (Q, \leq, 0, (+))$  be a divergence domain and  $\Delta = \{\Delta_I^m : (U(TI))^2 \rightarrow \mathcal{Q}\}_{m \in M, I \in \mathbb{C}}$  be a doubly indexed family of  $\mathcal{Q}$ -divergences on  $TI$  satisfying monotonicity on  $m$  (Definition 6). Define the following mapping  $T^{[\Delta]} : (M \times \mathcal{Q}) \times \mathbf{Obj}(\mathbf{BRel}(\mathbb{C})) \rightarrow \mathbf{Obj}(\mathbf{BRel}(\mathbb{C}))$ :

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

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

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

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

Intuitively,  $T^{[\Delta]}$  is a graded version of the *codensity lifting* (Katsumata et al., 2018) of  $T$  along the specific fibration  $p_{\mathbb{C}} : \mathbf{BRel}(\mathbb{C}) \rightarrow \mathbb{C}^2$ . We extend the codensity lifting with the grading mechanism in the same way as the graded  $\top$ -lifting in Katsumata (2014). The graded codensity lifting is also a generalization of the graded relational lifting for DP given in Sato (2016).

*Proof.* (Proof of (1)) Proving conditions 1–3 of graded strong relational lifting (Definition 37) are routine generalization of Katsumata et al. (2018) and Katsumata (2014, Section 5); thus omitted here (see Lemma 50 in appendix).

Condition 4 of Definition 37 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  $f_i \bullet j = f \bullet \langle i, j \rangle$  holds for any  $j \in UJ$ , we have the equivalence:

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

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

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

(Proof of (2)-(S)) We show the equivalence of  $\Delta$  being  $E$ -unit-reflexive and the implication:

$$\begin{aligned}
 &\forall I \in \mathbb{C}, m \in M, v \in \mathcal{Q}, c, c' \in U(TI) . \\
 &(\forall J \in \mathbb{C}, m' \in M, v' \in \mathcal{Q}, (k, l) : EI \rightarrow \tilde{\Delta}(m', v')J . \Delta_J^{m \cdot m'}(k^\sharp \bullet c, l^\sharp \bullet c') \leq v + v') \quad (15) \\
 &\implies \Delta_I^m(c, c') \leq v.
 \end{aligned}$$

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 (15) becomes

$$\forall J \in \mathbb{C}, m' \in M, v' \in \mathcal{Q}, (k, l) : EI \rightarrow \tilde{\Delta}(m', v')J . \Delta_J^{m' \cdot 1}(k \bullet i, l \bullet j) \leq v',$$

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

Conversely, we suppose that  $\Delta$  satisfies the unit reflexivity. We take  $I, m, v, c, c'$  of appropriate type and assume the middle part of (15). By instantiating it with  $J = I, m' = 1, v' = 0, k = l = \eta_I$ , we conclude  $\Delta_I^m(c, c') \leq v$ .

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

$$\begin{aligned}
 &\forall I \in \mathbb{C}, m \in M, v \in \mathcal{Q} . \tilde{\Delta}(m, v)I \leq T^{[\Delta]}(m, v)(EI) \\
 &\iff \left( \begin{array}{l} \forall I \in \mathbb{C}, m \in M, v \in \mathcal{Q}, c, c' \in U(TI) . \\ \Delta_I^m(c, c') \leq v \implies \\ \forall J \in \mathbb{C}, m' \in M, v' \in \mathcal{Q}, (k, l) : EI \rightarrow \tilde{\Delta}(m', v')J . \\ (k^\sharp \bullet c, l^\sharp \bullet c') \in \tilde{\Delta}(m \cdot m', v + v')J \end{array} \right) \\
 &\iff \left( \begin{array}{l} \forall I, J \in \mathbb{C}, m \in M, v \in \mathcal{Q}, c, c' \in U(TI), m' \in M, v' \in \mathcal{Q}, k, l \in \mathbb{C}(I, TJ) . \\ \Delta_I^m(c, c') \leq v \implies \\ (\forall (i, j) \in EI . (k \bullet i, l \bullet j) \in \tilde{\Delta}(m', v')J) \implies \Delta_J^{m \cdot m'}(k^\sharp \bullet c, l^\sharp \bullet c') \leq v + v' \end{array} \right) \\
 &\iff \left( \begin{array}{l} \forall I, J \in \mathbb{C}, m \in M, v \in \mathcal{Q}, c, c' \in U(TI), m' \in M, v' \in \mathcal{Q}, k, l \in \mathbb{C}(I, TJ) . \\ \Delta_I^m(c, c') \leq v \implies \\ \sup_{(i, j) \in EI} \Delta_J^{m'}(k \bullet i, l \bullet j) \leq v' \implies \Delta_J^{m \cdot m'}(k^\sharp \bullet c, l^\sharp \bullet c') \leq v + v' \end{array} \right) \\
 &\iff \left( \begin{array}{l} \forall I, J \in \mathbb{C}, m \in M, c, c' \in U(TI), m' \in M, k, l \in \mathbb{C}(I, TJ) . \\ \Delta_I^{m \cdot m'}(k^\sharp \bullet c, l^\sharp \bullet c') \leq \Delta_{I(c, c')}^m + \sup_{(i, j) \in EI} \Delta_J^{m'}(k \bullet i, l \bullet j). \end{array} \right)
 \end{aligned}$$



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

The construction of  $T^{[\Delta]}$  gives the greatest relational lifting of  $T$  with the fundamental property.

**Proposition 40.** *Let  $(\mathbb{C}, T)$  be a CC-SM,  $E : \mathbb{C} \rightarrow \mathbf{BRel}(\mathbb{C})$  be a basic endorelation,  $(M, \leq, 1, (\cdot))$  be a grading monoid,  $\mathcal{Q}$  be a divergence domain,  $\Delta \in \mathbf{Div}(T, E, M, \mathcal{Q})$  be a divergence, and  $\dot{T}$  be an  $M \times \mathcal{Q}$ -graded relational lifting satisfying the fundamental property with respect to  $\Delta$ . Then, for all  $m \in M$ ,  $q \in \mathcal{Q}$ , and  $X \in \mathbf{BRel}(\mathbb{C})$ , the following inequality holds*

$$\dot{T}(m, v)X \leq T^{[\Delta]}(m, v)X.$$

**7.1 Simplifying codensity liftings by  $\Omega$ -generatedness of divergences**

We show that the calculation of the codensity lifting  $T^{[\Delta]}$  can be simplified when  $\Delta$  is  $\Omega$ -generated. For an object  $I \in \mathbb{C}$ , we define  $T^{[\Delta],I}$  by:

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

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, while if  $\Delta$  is  $\Omega$ -generated, the parameter  $I$  can be fixed to  $\Omega$ .

**Proposition 41.** *For any  $\Omega$ -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  $X \in \mathbf{BRel}(\mathbb{C})$ .

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

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

$$(c'_1, c'_2) \in \tilde{\Delta}(m', v')I \iff \forall k : I \rightarrow T\Omega \cdot (k^\# \bullet c'_1, k^\# \bullet c'_2) \in \tilde{\Delta}(m', v')\Omega$$

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

$$\begin{aligned} (c_1, c_2) \in T^{[\Delta],\Omega}X \\ \iff \forall n \in M, w \in \mathcal{Q}, (k_1, k_2) : X \dot{\rightarrow} \tilde{\Delta}(n, w)\Omega \cdot (k_1^\# \bullet c_1, k_2^\# \bullet c_2) \in \tilde{\Delta}(m \cdot n, v + w)\Omega \\ \implies \left( \forall I \in \mathbb{C}, n \in M, w \in \mathcal{Q}, (l_1, l_2) : X \dot{\rightarrow} \tilde{\Delta}(n, w)I, k : I \rightarrow T\Omega \cdot \right. \\ \left. (k^\# \circ l_1^\# \bullet c_1, k^\# \circ l_2^\# \bullet c_2) \in \tilde{\Delta}(m \cdot n, v + w)\Omega \right) \\ \iff \forall I \in \mathbb{C}, n \in M, w \in \mathcal{Q}, (l_1, l_2) : X \dot{\rightarrow} \tilde{\Delta}(n, w)I \cdot (l_1^\# \bullet c_1, l_2^\# \bullet c_2) \in \tilde{\Delta}(m \cdot n, v + w)I \\ \iff (c_1, c_2) \in T^{[\Delta]}X. \end{aligned}$$

This completes the proof.  $\square$

For example, the generatedness of DP shown in Section 6.2 implies that  $G^{[DP]} = G^{[DP],2}$  and  $G_s^{[DP]} = G_s^{[DP],1}$ . In fact, the simplification  $G_s^{[DP],1}$  is equal to the  $(\mathcal{R}^+)^2$ -graded relational lifting  $G_s^{\top\top}$  for DP given in Sato (2016, Section 2.2), which is defined by, for each  $(X_1, X_2, R_X) \in \mathbf{BRel}(\mathbf{Meas})$ :

$$\begin{aligned} G_s^{\top\top}(\varepsilon, \delta)(X_1, X_2, R_X) \\ \triangleq (G_s(X_1), G_s(X_2), \{(v_1, v_2) \mid \forall A \in \Sigma_{X_1}, B \in \Sigma_{X_2} \cdot R_X(A) \subseteq B \implies v_1(A) \leq \exp(\varepsilon)v_2(B) + \delta\}). \end{aligned}$$

For detail, see the proof of equalities ( $\dagger$ ) and ( $\ddagger$ ) in the proof of Theorem 2.2 (iv) in Sato (2016).

**7.2 Two lifting approaches: Codensity and coupling**

We briefly compare two lifting approaches: graded codensity lifting and coupling-based lifting, and the latter of which is employed in Barthe et al. (2012), Barthe and Olmedo (2013), Barthe et al. (2014), Barthe et al. (2015), and Sato et al. (2019).

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

$$\dot{D}^\Delta(m, \nu)X \triangleq (DX_1, DX_2, \{(Dp_1 \bullet \mu_1, Dp_2 \bullet \mu_2) \mid (\mu_1, \mu_2) \in (DR_X)^2, \Delta_{R_X}^m(\mu_1, \mu_2) \leq \nu\}) \quad (16)$$

where  $p_i: R_X \rightarrow X_i$  is the projection ( $i = 1, 2$ ) from the binary relation. The pair  $(\mu_1, \mu_2)$  of probability distributions collected in the right-hand side of (16) is called a *coupling*.

The fundamental property  $\dot{D}^\Delta(\text{Eq } I) = \tilde{\Delta}(m, \nu)I$  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 (Barthe and Olmedo, 2013, 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 (16) 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 (C, T). As the prior study by Sato et al. (2019) pointed out, there is already a difficulty in extending it to the CC-SM (Meas, G), where G is the Giry monad.

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  $p_i: H_X \rightarrow X_i$  for measurable projections ( $i = 1, 2$ ). We then define a candidate  $M \times \mathcal{Q}$ -graded lifting of G by:

$$\dot{G}(m, \nu)X = (GX_1, GX_2, \{(Gp_1 \bullet \mu_1, Gp_2 \bullet \mu_2) \mid (\mu_1, \mu_2) \in (U(GH_X))^2, \Delta_{H_X}^m(\mu_1, \mu_2) \leq \nu\}).$$

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

$$(f, g): Y \rightarrow \dot{G}(m', \nu')X \implies (f^\sharp, g^\sharp): \dot{G}(m, \nu)Y \rightarrow \dot{G}(mm', \nu + \nu')X.$$

Let  $(f, g): Y \rightarrow \dot{G}(m', \nu')X$  be pair of measurable functions. Then for each  $(x, y) \in R_Y$ , we have  $(f \bullet x, g \bullet y) \in R_{\dot{G}(m, \nu)X}$ . 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  $\mu_1, \mu_2: R_Y \rightarrow UGH_X$ . If they were measurable functions of type  $H_Y \rightarrow GH_X$ , then from the composability of Δ, we would have  $\Delta_{H_X}^{mm'}(\mu_1^\sharp \bullet w_1, \mu_2^\sharp \bullet w_2) \leq \nu + \nu'$  for  $w_1, w_2 \in U(GH_Y)$  such that  $\Delta_{H_Y}^{m'}(w_1, w_2) \leq \nu'$ . This gives  $(f^\sharp, g^\sharp): \dot{G}(m, \nu)Y \rightarrow \dot{G}(mm', \nu + \nu')X$ . However, in general, ensuring the measurability of  $\mu_1, \mu_2$  is not possible, especially because they are picked up by the axiom of choice. A solution given in Sato et al. (2019) is to use the category **Span(Meas)** of spans that guarantees the existence of good measurable functions  $h_1, h_2: H_Y \rightarrow GH_X$ .

**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 (Barthe et al., 2015). 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

$$\begin{aligned}
 \mathbf{Typ}(B) \ni \tau ::= & b \mid 1 \mid \tau \times \tau \mid 0 \mid \tau + \tau \mid \tau \Rightarrow \tau \mid T\tau \quad (b \in B) \\
 M ::= & x \mid o(M) \mid c(M) \mid () \mid (M, M) \mid \pi_1(M) \mid \pi_2(M) \quad (o \in O_v, c \in O_e) \\
 & \mid \iota_1(M) \mid \iota_2(M) \mid M \text{ with } \iota_1(x : \tau).M \mid \iota_2(x : \tau).M \\
 & \mid (\lambda x : \tau. M) \mid (MM) \mid \text{ret}(M) \mid \text{let } x : \tau = M \text{ in } M
 \end{aligned}$$

Figure 2. Syntax of types and raw terms of the computational metalanguage.

relational refinement type system HOARE2 (Barthe et al., 2015), 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 nontermination.

The relational logic acRL adopts the *extensional approach* (Nielson and Nielson, 2007, Chapter 9.2):

- Relational assertions between contexts  $\Gamma$  and  $\Delta$  are defined as binary relations between  $U[\Gamma]$  and  $U[\Delta]$ , or equivalently  $\mathbf{BRel}(\mathbb{C})$ -objects  $\phi$  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,  $\phi$  be a relational assertion between  $\Gamma, \Delta$ , and  $\psi$  be an assertion between  $\tau, \sigma$ . The main concern of acRL is the statement “ $\forall (\gamma, \delta) \in \phi. ([M] \bullet \gamma, [N] \bullet \delta) \in \psi$ ” (equivalently  $([M], [N]) : \phi \dot{\rightarrow} \psi$ ). In this section, this statement is denoted 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* (Moggi, 1991). 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  $(B, \Sigma_v, \Sigma_e)$  where  $B$  is a set of base types, and  $\Sigma_v$  and  $\Sigma_e$  are functions whose range is  $\mathbf{Typ}_1(B)^2$ . The domains of  $\Sigma_v, \Sigma_e$  are sets of *value operation* symbols and *effectful operation* symbols and are denoted by  $O_v$  and  $O_e$ , respectively. These functions assign input and output types to these operations.

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  $\Gamma, \Delta$ . For contexts  $\Gamma, \Delta$  such that  $\text{dom}(\Gamma) \cap \text{dom}(\Delta) = \emptyset$ , by  $\Gamma, \Delta$  we mean the join of  $\Gamma$  and  $\Delta$ .

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  $\Gamma$  is a context,  $M$  is a raw term, and  $\tau$  is a type. It adopts the standard rules for products, coproducts, implications, and monadic types; see for example, Moggi (1991). The typing rules for value operations and effectful operations are given by:

$$\frac{o \in O_v \quad \Sigma_v(o) = (b, b') \quad \Gamma \vdash M : b}{\Gamma \vdash o(M) : b'} \quad \frac{c \in O_e \quad \Sigma_e(c) = (b, b') \quad \Gamma \vdash M : b}{\Gamma \vdash c(M) : Tb'}$$

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

Figure 3. Data for the categorical semantics of metalanguage.

A *simultaneous substitution*  $\theta$  from  $\Gamma$  to  $\Gamma'$  (written  $\Gamma \vdash \theta : \Gamma'$ ) is a function  $\theta$  from the set  $\text{dom}(\Gamma')$  to raw terms that assigns to each variable  $x \in \text{dom}(\Gamma')$  a well-typed raw term  $\Gamma \vdash \theta(x) : \Gamma'(x)$ . The application of  $\theta$  to a term  $\Gamma' \vdash M : \tau$  is denoted by  $M\theta$ , which has a typing  $\Gamma \vdash M\theta : \tau$ . For disjoint contexts  $\Gamma_i$  ( $i = 1, 2$ ), 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  $(B, \Sigma_v, \Sigma_e)$  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  $\Gamma$ , we fix a product diagram  $(\llbracket \Gamma \rrbracket, \{\pi_x : \llbracket \Gamma \rrbracket \rightarrow \llbracket \Gamma(x) \rrbracket\}_{x \in \text{dom}(\Gamma)})$ ; when  $\text{dom}(\Gamma) = \{x\}$ , we assume that  $\llbracket \Gamma \rrbracket = \llbracket \Gamma(x) \rrbracket$  with  $\pi_x = \text{id}$ . Lastly, we interpret a typing derivation of  $\Gamma \vdash M : \tau$  as a morphism  $\llbracket M \rrbracket : \llbracket \Gamma \rrbracket \rightarrow \llbracket \tau \rrbracket$  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 morphism  $\llbracket \theta \rrbracket : \llbracket \Gamma \rrbracket \rightarrow \llbracket \Gamma' \rrbracket$ .

## 8.2 Approximate relational computational logic

### 8.2.1 Relational logic in external form

A *relational assertion*  $\phi$  between disjoint contexts  $\Gamma$  and  $\Delta$  is a binary relation between  $U\llbracket \Gamma \rrbracket$  and  $U\llbracket \Delta \rrbracket$ . Such a relational assertion is denoted by  $\Gamma_{\Delta} \vdash \phi$ . We identify it as a **BRel**( $\mathbb{C}$ )-object  $\phi$  such that  $p_{\mathbb{C}}(\phi) = (\llbracket \Gamma \rrbracket, \llbracket \Delta \rrbracket)$ . Similarly, a relational assertion between types  $\tau$  and  $\sigma$  is defined to be a relational assertion  $\frac{u:\tau}{d:\sigma} \vdash \phi$ ; here  $u$  and  $d$  are reserved and fixed variables, respectively.

Relational assertions between contexts  $\Gamma$  and  $\Delta$  carry a *boolean algebra structure*  $\wedge, \vee, \neg$  given by the set-intersection, set-union, and set-complement (see the boolean algebra **BRel**( $\mathbb{C}$ )( $\llbracket \Gamma \rrbracket, \llbracket \Delta \rrbracket$ ) in Section 2.2). The pseudo-complement  $\phi \Rightarrow \psi$  is defined to be  $\neg\phi \vee \psi$ . For  $\frac{\Gamma, x:\tau}{\Delta, y:\sigma} \vdash \phi$ , by  $\Gamma_{\Delta} \vdash \forall_y^x . \phi$  and  $\Gamma_{\Delta} \vdash \exists_y^x . \phi$  we mean the relational assertions defined by the following equivalence:

$$\begin{aligned} (\gamma, \delta) \in \forall_y^x . \phi &\iff \forall \gamma' \in U\llbracket \Gamma, x : \tau \rrbracket, \delta' \in U\llbracket \Delta, y : \sigma \rrbracket . \\ &\quad (\llbracket \pi_1^{\Gamma, x:\tau} \rrbracket \bullet \gamma' = \gamma \wedge (\llbracket \pi_1^{\Delta, y:\sigma} \rrbracket \bullet \delta' = \delta) \Rightarrow (\gamma', \delta') \in \phi \\ (\gamma, \delta) \in \exists_y^x . \phi &\iff \exists \gamma' \in U\llbracket \Gamma, x : \tau \rrbracket, \delta' \in U\llbracket \Delta, y : \sigma \rrbracket . \\ &\quad (\llbracket \pi_1^{\Gamma, x:\tau} \rrbracket \bullet \gamma' = \gamma \wedge (\llbracket \pi_1^{\Delta, y:\sigma} \rrbracket \bullet \delta' = \delta) \wedge (\gamma', \delta') \in \phi \end{aligned}$$

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  $E : \mathbb{C} \rightarrow \mathbf{BRel}(\mathbb{C})$ , grading monoid  $M$ , divergence domain  $\mathcal{Q}$  and divergence  $\Delta \in \mathbf{Div}(T, E, M, \mathcal{Q})$ , we obtain a relational assertion  $\frac{u:T\tau}{d:\sigma} \vdash T^{[\Delta]}(m, \nu)\phi$  from any  $\frac{u:\tau}{d:\sigma} \vdash \phi, m \in M$  and  $\nu \in \mathcal{Q}$ .

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

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  $\Gamma \vdash \phi$  and  $\frac{u:\tau}{d:\sigma} \vdash \psi$ , by the judgment:

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

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

- Proposition 42.** (1)  $\phi \vdash (M, N) : \psi$  and  $[[M]] = [[M']]$  and  $[[N]] = [[N']]$  implies  $\phi \vdash (M', N') : \psi$ .  
 (2)  $\phi \vdash (M, N) : \psi$  and  $\phi' \subseteq \phi$  and  $\psi \subseteq \psi'$  implies  $\phi' \vdash (M, N) : \psi'$ .  
 (3)  $\phi \vdash (M, N) : T^{[\Delta]}(m, v)\psi$  and  $m \leq n$  and  $v \leq w$  and  $\psi \leq \psi'$  implies  $\phi \vdash (M, N) : T^{[\Delta]}(n, w)\psi'$ .  
 (4)  $\phi \vdash (M, N) : \psi$  implies  $\phi \vdash (\text{ret } (M), \text{ret } (N)) : T^{[\Delta]}(1, 0)\psi$ .  
 (5)  $\phi \vdash (M, N) : T^{[\Delta]}(m, v)\psi$  and  $\phi, \psi[[x/u]; [x'/d]] \vdash (M', N') : T^{[\Delta]}(n, w)\rho$  implies  $\phi \vdash (\text{let } x = M \text{ in } M', \text{let } x' = N \text{ in } N') : T^{[\Delta]}(m \cdot n, v \cdot w)\rho$ .

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

*Proof.* Take an arbitrary pair  $(x, y) \in \phi$ . We have  $\Delta_{[[b']]^m}([\![c]\!] \bullet x, [\![c]\!] \bullet y) \leq v$  by definition of  $v$ . Thanks to the fundamental property of  $T^{[\Delta]}$  (Theorem 39), it is equivalent to  $([\![c]\!] \bullet x, [\![c]\!] \bullet y) \in T^{[\Delta]}(m, v)(E[[b']])$ .  $\square$

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 distributions and Laplace distributions. This computational metalanguage is specified by the computational signature:

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

where  $\Sigma_v$  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 2.1),
- (2) for the interpretation  $[[\mathbb{R}]]$  of  $\mathbb{R}$ , we take the quasi-Borel space  $K\mathbb{R}$  associated with the standard Borel space  $\mathbb{R}$ , where  $K : \mathbf{Meas} \rightarrow \mathbf{QBS}$  is defined in Section 2.1,

- (3) the interpretation of value operations is given as expected (we omit it here); for example when  $\Sigma_V$  contains the real number addition operator  $+$  as type  $(\mathbb{R} \times \mathbb{R}, \mathbb{R})$ , its interpretation is the QBS morphism  $[[+]](x, y) = x + y : [[\mathbb{R} \times \mathbb{R}]] \rightarrow [[\mathbb{R}]]$ ,
- (4) for the interpretation of effectful operations, we put

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

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

**9.1 A relational logic for differential privacy**

To formulate differential privacy and its relaxations in the quasi-Borel setting, we convert statistical divergences  $\Delta$  on the Giry monad  $G$  in Table 2 to Eq-relative divergences  $\langle L, l \rangle^* \Delta$  on the probability monad  $P$  on QBS by the construction in Section 5.5. Then, we construct the graded relational lifting  $P^{\langle L, l \rangle^* \Delta}$  by Theorem 39. Using this, as an instantiation of acRL, we build a 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 42.

For effectful operations, we import basic proof rules on noise-adding mechanisms given in prior studies (Bun et al., 2018; Dwork et al., 2006; Dwork and Roth, 2013; Mironov, 2017) via Theorem 12 and Proposition 43. For example, consider the Eq-relative  $\mathcal{R}^+$ -graded  $\mathcal{R}^+$ -divergence  $\Delta = \langle L, l \rangle^* \text{DP}$  on  $P$ . Proposition 43 with an effectful operation  $c = \text{lap}$  and a relational assertion (below we identify global elements in  $K\mathbb{R}$  and real numbers):

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

together with Theorem 12 and the prior result (Dwork et al., 2006, Example 1) yields the following judgment:

$$\phi \vdash (\text{lap}(u), \text{lap}): (d)P^{\langle L, l \rangle^* \text{DP}}(0, \varepsilon)(\text{Eq } K\mathbb{R}).$$

By letting  $\text{diff}_r$  be the relational assertion  $\frac{u: \mathbb{R}}{d: \mathbb{R}} \vdash \{(x, y) \mid |x - y| \leq r\}$ , the above judgment is equivalent to:

$$\text{diff}_1 \vdash (\text{lap}(u, 1/\varepsilon), \text{lap}(d, 1/\varepsilon)): P^{\langle L, l \rangle^* \text{DP}}(0, \varepsilon)(\text{Eq } K\mathbb{R}). \tag{17}$$

This rule corresponds to the rule [LapGen] of the program logic apRHL+ (Barthe et al., 2017) for differential privacy. For another example, by the reflexivity of DP,  $\langle L, l \rangle^* \text{DP}$  is also reflexive. Hence, by letting  $\text{succ}_r$  be the relational assertion  $\frac{u: \mathbb{R}}{d: \mathbb{R}} \vdash \{(x, y) \mid y = x + r\}$ , we obtain the following judgments:

$$\text{succ}_1 \vdash (\text{lap}(u, \lambda), \text{lap}(d, \lambda)): P^{\langle L, l \rangle^* \text{DP}}(0, 0)(\text{succ}_1) \tag{18}$$

$$\text{succ}_1 \vdash (\text{norm}(u, \sigma), : \text{norm}(d, \sigma))P^{\langle L, l \rangle^* \text{DP}}(0, 0)(\text{succ}_1). \tag{19}$$

The judgment (18) corresponds to [LapNull] of apRHL+. Similarly, the following judgments about the DP, Rényi-DP, and zero-concentrated DP of the Gaussian mechanism can be derived as (20)–(22):

$$\text{diff}_1 \vdash (\text{norm}(u, \sigma), \text{norm}(d, \sigma)): P^{\langle L, l \rangle^* \text{DP}}(\varepsilon, \delta)(\text{Eq } K\mathbb{R}) \tag{20}$$

$$\text{diff}_r \vdash (\text{norm}(u, \sigma), \text{norm}(d, \sigma)): P^{\langle L, l \rangle^* \alpha \text{Re}}(\alpha r^2 / 2\sigma^2)(\text{Eq } K\mathbb{R}) \tag{21}$$

$$\text{diff}_r \vdash (\text{norm}(u, \sigma), \text{norm}(d, \sigma)): P^{\langle L, l \rangle^* \text{zCDP}}(0, r^2 / 2\sigma^2)(\text{Eq } K\mathbb{R}) \tag{22}$$

In (20), we require  $\sigma \geq \max((1 + \sqrt{3})/2, \sqrt{2 \log(0.66/\delta)}/\epsilon)$ . The derivation is done via Proposition 43, Theorem 12, and prior studies (Bun and Steinke, 2016; Mironov, 2017; Sato, 2016).

### 10. Case Study II: Probabilistic Programs with Costs

We further extend the computational signature  $\mathcal{C}$  in the previous section with an effectful operation `tick` such that  $\Sigma_c(\text{tick}) = (\mathbb{R}, 1)$ . The intention of `tick`( $r$ ) is to increase cost counter by  $r$  during execution<sup>8</sup>. 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,
- (4) for the interpretation of effectful operations, put

$$\begin{aligned} \llbracket \text{norm} \rrbracket(x, \sigma) &= [(0, \text{id}), \mathcal{N}(x, \sigma^2)]_{\sim_{K\mathbb{R} \times K\mathbb{R}}}, \\ \llbracket \text{lap} \rrbracket(x, \lambda) &= [(0, \text{id}), \text{Lap}(x, \lambda)]_{\sim_{K\mathbb{R} \times K\mathbb{R}}}, \\ \llbracket \text{tick} \rrbracket(r) &= \eta_{K\mathbb{R} \times \llbracket 1 \rrbracket}^P(r, *) = [\text{const}(r, *), \mu]_{\sim_{K\mathbb{R} \times 1}}. \end{aligned}$$

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

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

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

#### 10.1 Relational reasoning on probabilistic costs

We convert the total valuation distance  $\text{TV} \in \mathbf{Div}(G, \text{Eq}, 1, \mathcal{R}^+)$  to the divergence  $\Delta_c \triangleq C(\langle L, l \rangle^* \text{TV}, K\mathbb{R}) \in \mathbf{Div}(P_c, \text{Eq}, 1, \mathcal{R}^+)$  on  $P_c$  by Propositions 11 and 17. We also prove basic facts on effectful operations. First, the following relational judgments on `tick` can be easily given:

$$\begin{aligned} \top \vdash (\text{tick}(u), \text{tick}(d)) &: T^{[\Delta_c]}(1)(\top) \\ u = d \vdash (\text{tick}(u), \text{tick}(d)) &: T^{[\Delta_c]}(0)(\top) \end{aligned} \tag{23}$$

Remark that  $\text{Eq } 1 = \top$  and  $\llbracket \text{tick}(0) \rrbracket = \llbracket \text{ret } (*) \rrbracket$  holds. Next, in the similar way as (18), by the reflexivity of TV, we have the reflexivity of  $\langle L, l \rangle^* \text{TV}$ , and we obtain, for each real number constant  $\sigma$  and  $\lambda$ :

$$\begin{aligned} \text{succ}_r \vdash (\text{norm}(u, \sigma), \text{norm}) &: (d, \sigma) T^{[\Delta_c]}(0)(\text{succ}_r) \\ \text{succ}_r \vdash (\text{lap}(u, \lambda), \text{lap}(d, \lambda)) &: T^{[\Delta_c]}(0)(\text{succ}_r) \end{aligned} \tag{24}$$

We also directly verify the following judgment on `ntick` using Theorem 12 and Proposition 43:

$$\text{diff}_1 \vdash (\text{ntick } \sigma \ u, \text{ntick } \sigma \ d) : T^{[\Delta_c]}(\text{Pr}_{r \sim \mathcal{N}(0, \sigma^2)}[|r| < 0.5])(\top). \tag{25}$$

##### 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:

$$M \triangleq \lambda r: \mathbb{R}. \lambda t: \mathbb{R} \rightarrow T1. \text{let } x = \text{in lap}(r, 5) \text{ let } \_ = \text{in } t(r) \text{ ret } (x - r).$$

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:

$$\vdash (M\ 0\ (\lambda x.\text{tick}(x)), M\ 1\ (\lambda x.\text{tick}(x))): T^{[\Delta_c]}(1)(\text{Eq } \llbracket \mathbb{R} \rrbracket), \tag{A}$$

$$\vdash (M\ 0\ (\text{ntick}(2)), M\ 1\ (\text{ntick}(2))): T^{[\Delta_c]}(0.20)(\text{Eq } \llbracket \mathbb{R} \rrbracket) \tag{B}$$

In judgment (A), we pass the tick operation  $t = \lambda x.\text{tick}(x)$  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  $t' = \text{ntick}(2)$  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 (23) and 2 of Proposition 42, we have

$$\text{succ}_1 \vdash (\text{tick}(u), \text{tick}(d)): T^{[\Delta_c]}(1)(\top). \tag{26}$$

By (26), and 4, 5 of Proposition 42, we obtain

$$\begin{aligned} \text{succ}_1 \vdash & (\text{let } \_ = \text{in tick}(u)\text{ret } (u), \\ & \text{let } \_ = \text{in tick}(d)\text{ret } (d - 1)): T^{[\Delta_c]}(1)(\text{Eq } \llbracket \mathbb{R} \rrbracket). \end{aligned} \tag{27}$$

By (24), (27), and 1 and 5 of Proposition 42 again, we conclude (A).

To show (B), it suffices to replace (26) by the following judgment proved by (25), the inequality  $\Pr_{r \sim \mathcal{N}(0,4)} [|r| < 0.5] \leq 0.20$  and 2 of Proposition 42:

$$\text{succ}_1 \vdash (\text{ntick } 2\ u, \text{ntick } 2\ d): T^{[\Delta_c]}(0.20)(\top).$$

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

### 11. Related Work

This work is inspired by relational Hoare logics for verifying differential privacy of probabilistic programs, as summarized in Table 5. Composable divergences employed in these logics include the one for the standard DP, plus its recent relaxations, such as, Rényi DP, zero-concentrated DP, and truncated-concentrated DP (Bun et al., 2018; Bun and Steinke, 2016; Mironov, 2017).

The key semantic structure in these logics is *graded relational liftings* of the probability distribution monad. Barthe et al. gave a graded relational lifting of the probability distribution monad based on *couplings* (Barthe et al., 2012). Since then, coupling-based liftings have been refined and used in several works (Barthe et al., 2014, 2016; Barthe and Olmedo, 2013; Sato et al., 2019). They can be systematically constructed from *composable* divergences on the probability distribution monad (Barthe and Olmedo, 2013). 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 (Albarghouthi and Hsu, 2018a,b). These coupling-based

Table 5. Relational Hoare logics for verifying divergences

Work	Monad	Relation	Lifting method	Supported divergences
Barthe et al. (2014, 2016, 2012)	Dist	<b>BRel(Set)</b>	Coupling	DP
Barthe and Olmedo (2013)	Dist	<b>BRel(Set)</b>	Coupling	<i>f</i> -Divergences
Sato (2016)	Giry	<b>BRel(Meas)</b>	Codensity	DP
Sato et al. (2019)	Giry	<b>Span(Meas)</b>	Coupling	Composable ones
This work	Generic	<b>BRel(C)</b>	Codensity	Composable ones



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 by Sato et al. (2019).

The relational Hoare logic for differential privacy that supports sampling from continuous probability measures is given in the study by Sato (2016). In his work, the graded relational lifting for  $(\epsilon, \delta)$ -DP is constructed by a technique called *codensity lifting* (Katsumata et al., 2018), which does not rely on the existence of coupling. It has been an open question (Sato et al., 2019, Section VIII) how to extend the codensity lifting technique to support various relaxations of differential privacy. Theorem 39 of this paper answers to this question. Later, coupling-based liftings has also been extended to support sampling from continuous probability measures (Sato et al., 2019). This extension is achieved by redefining binary relations as *spans* of measurable functions. Comparison of these approaches is in Section 7.2.

Verification of differential privacy in functional programming languages has also been pursued (Azevedo de Amorim et al., 2019; Barthe et al., 2015; Gaboardi et al., 2013; Reed and Pierce, 2010). Reed and Pierce (2010) introduced a linear functional programming language with graded comonadic types that supports reasoning about  $\epsilon$ -differential privacy of probabilistic programs. Later, Gaboardi et al. (2013) strengthened the Reed–Pierce type system with dependent types. A category-theoretic account of the Reed–Pierce type system is given by Azevedo de Amorim et al. (2019), where general  $(\epsilon, \delta)$ -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. A coupling-based lifting is also employed in a relational model of a higher-order probabilistic programming language (Barthe et al., 2015).

The study by Azevedo de Amorim et al. (2019) gives a categorical definition of composable divergences in a general framework called *weakly closed refinements of symmetric monoidal closed categories* (Azevedo de Amorim et al., 2019, Definition 1). A comparison is given in Section 6.1.2.

Mardare et al. (2016) introduced a quantitative refinement of algebraic theory called *quantitative equational theory* and studied variety theorem for quantitative algebras. Bacci et al. (2021) discussed tensor products of quantitative equational theories. QETs and divergences on monads share the common interest of measuring quantitative difference between computational effects. Divergences on monads are derived as a generalization of the composability condition of statistical divergences studied by Barthe and Olmedo (2013). To make a precise connection between these two concepts, in Section 6.3, we have given an adjunction between QETs of type  $\Sigma$  over  $\Omega$  and  $\Omega$ -generated divergences on the free monad  $T_\Sigma$ . The adjunction cuts down to the isomorphism between *unconditional* QETs of type  $\Sigma$  over  $\Omega$  and  $\Omega$ -generated divergences on  $T_\Sigma$ .

In this paper, we have constructed strong graded relational liftings of monads from divergences on monads. The concept of relational lifting and its fibrational generalization are also key technical concepts in the categorical studies of bisimulation (Balan et al., 2019; Baldan et al., 2018; Hermida and Jacobs, 1995; Katsumata et al., 2018; Kurz and Velebil, 2016; Sprunger et al., 2021). It remains to be seen if the relational liftings obtained in this paper, as well as the divergence liftings in Section 6.1.3, have applications in the coalgebraic study of bisimulations.

Metric-like spaces are used in several recent papers on semantics of programming languages and systems. Gavazzo (2018) studied a quantitative refinement of Abramsky's applicative bisimilarity for the Reed–Pierce type system. He introduced a monadic operational semantics of the type system and formalized the concept of quantitative applicative bisimilarity using monads that are lifted to the category of quantale-valued relations. Bonchi et al. (2018) also used metric-like spaces to study bisimulations and up-to techniques in the category of quantale-valued relations. In this paper, we are interested in relational program verification of effectful programs, and we carry it 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  $\Delta$  is represented by the binary relation  $\tilde{\Delta}(v)$  that relates two computational effects whose distance is bound by  $v$ .

The RelCost system by Çiçek et al. (2017) is a formal system for reasoning about relational properties of higher-order functional programs and measuring cost difference of programs. It

consists of two subsystems: the relational refinement type system that can measure cost difference of programs, and the unary logic that can estimate lower and upper bounds of cost (i.e. cost intervals) of programs. We expect a connection between the semantics of the former system and the graded relational lifting constructed from the divergence NCI on  $P(\mathbb{N} \times -)$  (or its variant) in Section 5.2. On the other hand, identifying a semantic structure behind the latter system is not the scope of this paper, and we leave it to future work to identify it and relate it with divergences on monads.

**Acknowledgements.** 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 and Kazuki Matsuoka for helpful comments on the manuscript.

## Notes

- 1 In particular, there is no  $\sigma$ -algebra  $\Sigma$  over  $\mathbf{Meas}(\mathbb{R}, \mathbb{R})$  that makes the evaluation mapping  $(x, f) \mapsto f(x)$  a measurable function of type  $\mathbb{R} \times (\mathbf{Meas}(\mathbb{R}, \mathbb{R}), \Sigma) \rightarrow \mathbb{R}$ .
- 2 Strictly speaking, differential privacy depends on the definition of adjacency of datasets. The adjacency relation  $R_{\text{adj}}$  is usually defined as  $\{(d_1, d_2) \mid \rho(d_1, d_2) \leq 1\}$  with a metric  $\rho$  over  $I$ .
- 3 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  $\nu$  such that  $c(d_1), c(d_2) \ll \nu$ . [  $\implies$  ] Obvious. [  $\impliedby$  ] 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  $\nu$ .
- 4 Recall that an ultrametric space  $(I, d_I)$  is a set  $I$  together with a function  $d_I: I^2 \rightarrow [0, 1]$  such that  $d_I(x, x) = 0$  and  $d_I(x, z) \leq \max(d_I(x, y), d_I(y, z))$ .
- 5  $R_{EI} = \emptyset$  happens if and only if  $R_{EI} = \emptyset$  for any  $I \in \mathbb{C}$ . Therefore, nontrivial basic endorelations always satisfy  $R_{EI} \neq \emptyset$ .
- 6 A *pseudometric* is a function  $d: A^2 \rightarrow \mathcal{R}^+$  such that  $d(a, a) = 0$  (reflexivity),  $d(b, a) = d(a, b)$  (symmetry), and  $d(a, c) \leq d(a, b) + d(b, c)$  (triangle inequality) hold for all  $a, b, c \in A$ . Since  $d$  may return the positive infinity  $\infty$ , it is sometimes called an *extended pseudometric*.
- 7 If  $\sigma = 0$  (or  $\lambda \leq 0$ ),  $\mathcal{N}(x, \sigma^2)$  (resp.  $\text{Lap}(x, \lambda)$ ) is not defined; thus, we replace it by the Dirac distribution  $\mathbf{d}_x$  at  $x$  instead.
- 8 To make examples simpler, we allow negative costs.
- 9 For a measurable subset  $A \in \Sigma_I$ , an indicator function  $\chi_A: I \rightarrow [0, 1]$  of  $A$  is defined by  $\chi_A(x) = 1$  if  $x \in A$  and  $\chi_A = 0$  otherwise. A simple function is a linear combination of finite number of indicator functions.

## References

- Albarghouthi, A. and Hsu, J. (2018a). Constraint-based synthesis of coupling proofs. In *Computer Aided Verification – 30th International Conference, CAV 2018, Proceedings, Part I*, vol. 10981. LNCS. Springer, 327–346.
- Albarghouthi, A. and Hsu, J. (2018b). Synthesizing coupling proofs of differential privacy. *PACMPL* 2 (POPL) 58:1–58:30.
- Altenkirch, T., Chapman, J., and Uustalu, T. (2015). Monads need not be endofunctors. *Logical Methods in Computer Science* 11 (1).
- Aumann, R. J. (1961). Borel structures for function spaces. *Illinois Journal of Mathematics* 5 (4) 614–630.
- Azevedo de Amorim, A., Gaboardi, M., Hsu, J., and Katsumata, S. (2019). Probabilistic relational reasoning via metrics. In *34th Annual ACM/IEEE Symposium on Logic in Computer Science, LICS 2019*. IEEE, 1–19.
- Bacci, G., Mardare, R., Panangaden, P., and Plotkin, G. (2021). Tensor of quantitative equational theories. In Gadducci, F. and Silva, A. (eds.), *9th Conference on Algebra and Coalgebra in Computer Science (CALCO 2021)*, volume 211 of *Leibniz International Proceedings in Informatics (LIPIcs)*, Dagstuhl, Germany: Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 7:1–7:17.
- Balan, A., Kurz, A., and Velebil, J. (2019). Extending set functors to generalised metric spaces. *Logical Methods in Computer Science* 15 (1).
- Baldan, P., Bonchi, F., Kerstan, H., and König, B. (2018). Coalgebraic behavioral metrics. *Logical Methods in Computer Science* 14 (3).
- Balle, B., Barthe, G., Gaboardi, M., Hsu, J., and Sato, T. (2020). Hypothesis testing interpretations and renyi differential privacy. In Chiappa, S. and Calandra, R. (eds.), *Proceedings of the Twenty Third International Conference on Artificial Intelligence and Statistics (AISTATS 2020)*, vol. 108. *Proceedings of Machine Learning Research*, Online. PMLR, 2496–2506.

- Barthe, G., Crespo, J. M., and Kunz, C. (2011). Relational verification using product programs. In Butler, M. and Schulte, W., editors, *FM 2011: Formal Methods*, Berlin, Heidelberg: Springer Berlin Heidelberg, 200–214.
- Barthe, G., D'Argenio, P. R., and Rezk, T. (2004). Secure information flow by self-composition. In *Proceedings of the 17th IEEE Workshop on Computer Security Foundations, CSFW '04*, vol. 100, USA: IEEE Computer Society.
- Barthe, G., Gaboardi, M., Arias, E. J. G., Hsu, J., Kunz, C., and Strub, P. (2014). Proving differential privacy in Hoare logic. In *IEEE 27th Computer Security Foundations Symposium, CSF 2014*. IEEE Computer Society, 411–424.
- Barthe, G., Gaboardi, M., Arias, E. J. G., Hsu, J., Roth, A., and Strub, P. (2015). 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*. ACM, 55–68.
- Barthe, G., Gaboardi, M., Grégoire, B., Hsu, J., and Strub, P. (2016). Proving differential privacy via probabilistic couplings. In *Proceedings of the 31st Annual ACM/IEEE Symposium on Logic in Computer Science, LICS '16*. ACM, 749–758.
- Barthe, G., Grégoire, B., Hsu, J., and Strub, P.-Y. (2017). Coupling proofs are probabilistic product programs. In *Proceedings of the 44th ACM SIGPLAN Symposium on Principles of Programming Languages, POPL 2017*, New York, NY, USA: Association for Computing Machinery, 161–174.
- Barthe, G., Köpf, B., Olmedo, F., and Béguelin, S. Z. (2012). Probabilistic relational reasoning for differential privacy. In *Proceedings of the 39th ACM SIGPLAN-SIGACT Symposium on Principles of Programming Languages, POPL 2012*. ACM, 97–110.
- Barthe, G. and Olmedo, F. (2013). 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*, vol. 7966. LNCS. Springer, 49–60.
- Benton, N. (2004). Simple relational correctness proofs for static analyses and program transformations. *SIGPLAN Notices* 39 (1) 14–25.
- Bonchi, F., König, B., and Petrisan, D. (2018). Up-To Techniques for Behavioural Metrics via Fibrations. In *29th International Conference on Concurrency Theory (CONCUR 2018)*, vol. 118. *Leibniz International Proceedings in Informatics (LIPIcs)*, Dagstuhl, Germany: Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, 17:1–17:17.
- Bun, M., Dwork, C., Rothblum, G. N., and Steinke, T. (2018). Composable and versatile privacy via truncated CDP. In *Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing, STOC 2018*, New York, NY, USA: Association for Computing Machinery, 74–86.
- Bun, M. and Steinke, T. (2016). Concentrated differential privacy: Simplifications, extensions, and lower bounds. In *Theory of Cryptography*, Berlin, Heidelberg: Springer Berlin Heidelberg, 635–658.
- Çiçek, E., Barthe, G., Gaboardi, M., Garg, D., and Hoffmann, J. (2017). Relational cost analysis. *SIGPLAN Notices* 52 (1) 316–329.
- Csiszár, I. (1963). Eine informationstheoretische Ungleichung und ihre Anwendung auf den beweis der ergodizität von markoffschen ketten. *Magyar. Tud. Akad. Mat. Kutato Int. Kozl.* 8 85–108.
- Csiszár, I. (1967). Information-type measures of difference of probability distributions and indirect observations. *Studia Scientiarum Mathematicarum Hungarica* 2 299–318.
- Dwork, C., McSherry, F., Nissim, K., and Smith, A. (2006). Calibrating noise to sensitivity in private data analysis. In *Theory of Cryptography*, vol. 3876. LNCS. Springer Berlin Heidelberg, 265–284.
- Dwork, C. and Roth, A. (2013). The algorithmic foundations of differential privacy. *Foundations and Trends® in Theoretical Computer Science* 9 (3-4) 211–407.
- Gaboardi, M., Haeberlen, A., Hsu, J., Narayan, A., and Pierce, B. C. (2013). Linear dependent types for differential privacy. In *The 40th Annual ACM SIGPLAN-SIGACT Symposium on Principles of Programming Languages, POPL '13*. ACM, 357–370.
- Gaboardi, M., Katsumata, S., Orchard, D., and Sato, T. (2021). Graded hoare logic and its categorical semantics. In Yoshida, N., 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*, vol. 12648. *Lecture Notes in Computer Science*. Springer, 234–263.
- Gavazzo, F. (2018). 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*, New York, NY, USA: Association for Computing Machinery, 452–461.
- Giry, M. (1982). A categorical approach to probability theory. In Banaschewski, B., editor, *Categorical Aspects of Topology and Analysis*, vol. 915. LNM. Springer, 68–85.
- Hall, R. (2012). *New Statistical Applications for Differential Privacy*. PhD thesis, Machine Learning Department School of Computer Science Carnegie Mellon University.
- Hermida, C. and Jacobs, B. (1995). An algebraic view of structural induction. In *Proceedings of CSL '94*, vol. 933. LNCS. Springer-Verlag, 412–426.
- Heunen, C., Kammar, O., Staton, S., and Yang, H. (2017). A convenient category for higher-order probability theory. In *32nd Annual ACM/IEEE Symposium on Logic in Computer Science, LICS 2017*, 1–12.
- Jacobs, B. (1999). *Categorical Logic and Type Theory*. Elsevier.
- Kairouz, P., Oh, S., and Viswanath, P. (2015). The composition theorem for differential privacy. In *Proceedings of the 32nd International Conference on Machine Learning, ICML 2015, Lille, France, 6-11 July 2015*, 1376–1385.

- Katsumata, S. (2014). Parametric effect monads and semantics of effect systems. In *The 41st Annual ACM SIGPLAN-SIGACT Symposium on Principles of Programming Languages, POPL '14*. ACM, 633–646.
- Katsumata, S. and Sato, T. (2013). Preorders on monads and coalgebraic simulations. In Pfenning, F., editor, *Foundations of Software Science and Computation Structures*, Berlin, Heidelberg: Springer Berlin Heidelberg, 145–160.
- Katsumata, S., Sato, T., and Uustalu, T. (2018). Codensity lifting of monads and its dual. *Logical Methods in Computer Science* **14** (4).
- Kurz, A. and Velebil, J. (2016). Relation lifting, a survey. *Journal of Logical and Algebraic Methods in Programming* **85** (4) 475–499. Relational and algebraic methods in computer science.
- Liese, F. and Vajda, I. (2006). On divergences and informations in statistics and information theory. *IEEE Transactions on Information Theory* **52** (10) 4394–4412.
- Lucassen, J. M. and Gifford, D. K. (1988). Polymorphic effect systems. In *Conference Record of the Fifteenth Annual ACM Symposium on Principles of Programming Languages*. ACM Press, 47–57.
- Mac Lane, S. (1998). *Categories for the Working Mathematician (Second Edition)*, vol. 5. *Graduate Texts in Mathematics*. Springer.
- Mardare, R., Panangaden, P., and Plotkin, G. (2016). Quantitative algebraic reasoning. In *Proceedings of the 31st Annual ACM/IEEE Symposium on Logic in Computer Science, LICS '16*, New York, NY, USA: Association for Computing Machinery, 700–709.
- Mardare, R., Panangaden, P., and Plotkin, G. (2017). On the axiomatizability of quantitative algebras. In *2017 32nd Annual ACM/IEEE Symposium on Logic in Computer Science (LICS)*, Los Alamitos, CA, USA: IEEE Computer Society, 1–12.
- Mironov, I. (2017). Rényi differential privacy. In *2017 IEEE 30th Computer Security Foundations Symposium (CSF)*, 263–275.
- Mitchell, J. C. and Scedrov, A. (1992). Notes on scoping and relators. In *Computer Science Logic, 6th Workshop, CSL '92*, vol. 702. LNCS. Springer, 352–378.
- Moggi, E. (1991). Notions of computation and monads. *Information and Computation* **93** (1) 55–92.
- Morimoto, T. (1963). Markov processes and the H-theorem. *Journal of the Physical Society of Japan* **18** (3) 328–331.
- Nielson, H. R. and Nielson, F. (2007). *Semantics with Applications: An Appetizer*. Springer-Verlag, Berlin, Heidelberg.
- Olmedo, F. (2014). *Approximate Relational Reasoning for Probabilistic Programs*. PhD thesis, Technical University of Madrid.
- Prasad, S. and Smith, K. A. (2014). A note on differential privacy: Defining resistance to arbitrary side information. *Journal of Privacy and Confidentiality* **6** (1).
- Radiček, I., Barthe, G., Gaboardi, M., Garg, D., and Zuleger, F. (2017). Monadic refinements for relational cost analysis. *Proceedings of the ACM on Programming Languages* **2** (POPL) 36:1–36:32.
- Reed, J. and Pierce, B. C. (2010). 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*. ACM, 157–168.
- Rutten, J. J. M. M. (1996). Elements of generalized ultrametric domain theory. *Theoretical Computer Science* **170** (1–2) 349–381.
- Sato, T. (2014). Identifying all preorders on the subdistribution monad. In Jacobs, B., Silva, A., and Staton, S., 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*, pp. 309–327. Elsevier.
- Sato, T. (2016). Approximate relational hoare logic for continuous random samplings. In *The Thirty-second Conference on the Mathematical Foundations of Programming Semantics, MFPS 2016*, vol. 325. *Electronic Notes in Theoretical Computer Science*. Elsevier, 277–298.
- Sato, T., Barthe, G., Gaboardi, M., Hsu, J., and Katsumata, S. (2019). Approximate span liftings: Compositional semantics for relaxations of differential privacy. In *34th Annual ACM/IEEE Symposium on Logic in Computer Science, LICS 2019*, 1–14. IEEE.
- Smirnov, A. (2008). Graded monads and rings of polynomials. *Journal of Mathematical Sciences*, 151:3032–3051.
- Sprunger, D., Katsumata, S., Dubut, J., and Hasuo, I. (2021). Fibrational bisimulations and quantitative reasoning: Extended version. *Journal of Logic and Computation* **31** (6) 1526–1559.
- Street, R. (1972). The formal theory of monads. *Journal of Pure and Applied Algebra* **2** (2) 149–168.
- Wasserman, L. and Zhou, S. (2010). A statistical framework for differential privacy. *Journal of the American Statistical Association* **105** (489) 375–389.
- Zaks, A. and Pnueli, A. (2008). Covac: Compiler validation by program analysis of the cross-product. In Cuellar, J., Maibaum, T., and Sere, K. (eds.), *FM 2008: Formal Methods*, Berlin, Heidelberg: Springer Berlin Heidelberg, 35–51.

## Appendix A. Proofs for Section 4 (Divergences on Monads)

*Proof.* (Proof of Proposition 7) Let  $m \in M$ ,  $(x_1, x_2) \in EI$ ,  $c_1, c_2 \in U(TJ)$ . Below  $\eta$  stands for the  $I \times J$ -component  $\eta_{I \times J} : I \times J \rightarrow T(I \times J)$  of the unit of  $T$ . From the composability of  $\Delta$  and  $\eta_x \bullet y = \eta \bullet \langle x, y \rangle$ , we have

$$\begin{aligned} \Delta_{I \times J}^m((\eta_{x_1})^\sharp \bullet c_1, (\eta_{x_2})^\sharp \bullet c_2) &\leq \Delta_I^m(c_1, c_2) + \sup_{(y_1, y_2) \in EJ} \Delta_{I \times J}^1(\eta_{x_1} \bullet y_1, \eta_{x_2} \bullet y_2) \\ &= \Delta_I^m(c_1, c_2) + \sup_{(y_1, y_2) \in EJ} \Delta_{I \times J}^1(\eta \bullet \langle x_1, y_1 \rangle, \eta \bullet \langle x_2, y_2 \rangle). \end{aligned}$$

Since  $EI \dot{\times} EJ \leq E(I \times J)$ , we have  $(\langle x_1, y_1 \rangle, \langle x_2, y_2 \rangle) \in E(I \times J)$  for any  $(y_1, y_2) \in EJ$ . Therefore,

$$\forall (y_1, y_2) \in EJ. \Delta_{I \times J}^1(\eta \bullet \langle x_1, y_1 \rangle, \eta \bullet \langle x_2, y_2 \rangle) \leq 0$$

holds by the E-unit reflexivity. From  $(\eta_x)^\sharp \bullet c = \theta \bullet \langle x, c \rangle$  by (1), we obtain

$$\Delta_{I \times J}^m(\theta_{I,J} \bullet \langle x_1, c_1 \rangle, \theta_{I,J} \bullet \langle x_2, c_2 \rangle) = \Delta_{I \times J}^m((\eta_{x_1})^\sharp \bullet c_1, (\eta_{x_2})^\sharp \bullet c_2) \leq \Delta_I^m(c_1, c_2).$$

□

**Appendix B. Proofs for Section 5 (Examples of Divergences on Monads)**

**Proposition 44.** *The family  $C' = \{C'_I : (\mathbb{N} \times I)^2 \rightarrow \mathcal{N}\}_{I \in \text{Set}}$  of  $\mathcal{N}$ -divergences defined by:*

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

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  $(x, y) \in \text{Eq } I$  (that is,  $x = y \in I$ ), we have

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

We show the Eq-composability of  $C'$ . Let  $(i, x), (j, y) \in \mathbb{N} \times I$  and  $f, g : I \rightarrow \mathbb{N} \times J$ . 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

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

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

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

This completes the proof. □

**Proposition 45.** *The family  $\text{NC} = \{\text{NC}_I : (P(\mathbb{N} \times I))^2 \rightarrow \mathcal{N}\}_{I \in \text{Set}}$  of  $\mathcal{N}$ -divergences defined by:*

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

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

*Proof.* The monotonicity of  $\text{NC}$  is obvious.

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

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

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

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

This completes the proof. □

**Proposition 46.** The family  $\text{NCl} = \{\text{NCl}_I: (P(\mathbb{N} \times I))^2 \rightarrow \mathcal{N}\}_{I \in \text{Set}}$  of  $\mathcal{L}$ -divergences defined by:

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

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

*Proof.* The monotonicity of NCl is obvious.

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

$$\text{NCl}_I(\eta_I(x), \eta_I(y)) = \text{NCl}_I(\{(0, x)\}, \{(0, y)\}) = 0 - 0 = 0.$$

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

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

This completes the proof. □

*Proof.* (Proof of Proposition 10) We recall the continuity of  $f$ -divergence  ${}^f\text{Div}$  in Liese and Vajda (2006, Theorem 16) and Sato et al. (2019, Theorem 3):

$${}^f\text{Div}_I(\mu_1, \mu_2) = \sup \left\{ \sum_{i=0}^n \mu_2(B_i) f \left( \frac{\mu_1(B_i)}{\mu_2(B_i)} \right) \mid \{B_j\}_{j=0}^n : \text{measurable partition of } I \right\}.$$

Here, a measurable partition of  $I$  is a finite family  $\{B_i\}_{i=0}^n$  of measurable subsets  $B_i \in \Sigma_I$  satisfying  $i \neq j \implies B_i \cap B_j = \emptyset$  and  $\bigcup_{i=0}^n B_i = I$ .

We have the Eq-unit reflexivity because the reflexivity  ${}^f\text{Div}_I(\mu, \mu) = 0$  is obtained from  $f(1) = 0$ . We show the Eq-composability. To show this, we prove a bit stronger statement.

Consider three positive weight functions  $f, f_1, f_2 \geq 0$  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)

$$\begin{aligned} &\text{for all } x, y, z, w \in [0, 1], 0 \leq (\beta'z + (1 - \beta')x) + \gamma xf_1(z/x) \text{ and} \\ &xyf(zw/xy) \leq (\beta w + (1 - \beta)y)xf_1(z/x) + (\beta'z + (1 - \beta')x)yf_2(w/y) \\ &\quad + \gamma xyf_1(z/x)f_2(w/y) + \alpha(x - z)(w - y). \end{aligned}$$

Let  $\mu_1, \mu_2 \in G_s I$ , and  $h, k: I \rightarrow G_s J$ . We suppose that at least one of the following two conditions holds

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

We then prove the composability in the sense of Olmedo (2014, Definition 5.2):

$$\begin{aligned} &f \text{Div}_J(h^\sharp \mu_1, k^\sharp \mu_2) \\ &\leq \int^f \text{Div}_I(\mu_1, \mu_2) + \sup_{x \in I} \int^f \text{Div}_J(h(x), k(x)) + \gamma \int^f \text{Div}_I(\mu_1, \mu_2) \cdot \sup_{x \in I} \int^f \text{Div}_J(h(x), k(x)). \end{aligned} \tag{28}$$

We fix a measurable partition  $\{A_i\}_{i=0}^n$  of  $J$ . For each  $0 \leq i \leq n$ , by definition of the  $\sigma$ -algebra  $\Sigma_{GJ}$ , the functions  $h(-)(A_i): I \rightarrow [0, 1]$  and  $k(-)(A_i): I \rightarrow [0, 1]$  are measurable, and hence we have two monotone increasing sequences  $\{h_l^i\}_{l=0}^\infty$  and  $\{k_l^i\}_{l=0}^\infty$  of (nonnegative) simple functions<sup>9</sup> that converge uniformly to  $h(-)(A_i)$  and  $k(-)(A_i)$ , respectively. Since the sequences  $\{h_l^i\}_{l=0}^\infty$  and  $\{k_l^i\}_{l=0}^\infty$  converge uniformly and are bounded above, we obtain

$$\begin{aligned} \sum_{i=0}^n k^\sharp(\mu_2)(A_i) f \left( \frac{h^\sharp(\mu_1)(A_i)}{k^\sharp(\mu_2)(A_i)} \right) &= \sum_{i=0}^n \left( \int_X k(x)(A_i) d\mu_2(x) \right) f \left( \frac{\int_X h(x)(A_i) d\mu_1(x)}{\int_X k(x)(A_i) d\mu_2(x)} \right) \\ &= \sum_{i=0}^n \left( \int_X \lim_{l \rightarrow \infty} k_l^i d\mu_2 \right) f \left( \frac{\int_X \lim_{l \rightarrow \infty} h_l^i d\mu_1}{\int_X \lim_{l \rightarrow \infty} k_l^i d\mu_2} \right) \\ &= \sum_{i=0}^n \left( \lim_{l \rightarrow \infty} \int_X k_l^i d\mu_2 \right) f \left( \frac{\lim_{l \rightarrow \infty} \int_X h_l^i d\mu_1}{\lim_{l \rightarrow \infty} \int_X k_l^i d\mu_2} \right) \\ &= \lim_{l \rightarrow \infty} \sum_{i=0}^n \left( \int_X k_l^i d\mu_2 \right) f \left( \frac{\int_X h_l^i d\mu_1}{\int_X k_l^i d\mu_2} \right). \end{aligned}$$

We remark that the above computation is consistent even if  $k^\sharp(\mu_2)(A_i) = 0$  for some  $0 \leq i \leq n$ . We here recall  $0f(a/0) = af^*(0)$  for any  $a \in [0, \infty)$ . If  $k^\sharp(\mu_2)(A_i) = 0$ , then  $\int_X k_l^i d\mu_2 = 0$  for all  $l \in \mathbb{N}$  because it is a nonnegative monotone increasing sequence whose limit is 0. Then  $(\int_X k_l^i d\mu_2) f((\int_X h_l^i d\mu_1)/(\int_X k_l^i d\mu_2)) = (\int_X h_l^i d\mu_1) f^*(0)$  holds for all  $l \in \mathbb{N}$ . It is monotonically increasing, and converges to  $(h^\sharp(\mu_1)(A_i)) f^*(0)$ .

We thus show the following inequality:

$$\begin{aligned} &\lim_{l \rightarrow \infty} \sum_{i=0}^n \left( \int_X k_l^i d\mu_2 \right) f \left( \frac{\int_X h_l^i d\mu_1}{\int_X k_l^i d\mu_2} \right) \\ &\leq \int^f \text{Div}_I(\mu_1, \mu_2) + \sup_{x \in I} \int^f \text{Div}_J(h(x), k(x)) + \gamma \int^f \text{Div}_I(\mu_1, \mu_2) \sup_{x \in I} \int^f \text{Div}_J(h(x), k(x)). \end{aligned} \tag{29}$$

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

By Jensen’s inequality and the second inequality of (A’), we calculate as follows:

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

We will evaluate the three expressions  $V_1, V_2, V_3$ .

First, we obtain  $\lim_{l \rightarrow \infty} V_1 \leq \int_1 \text{Div}_l(\mu_1, \mu_2)$  as follows:

$$\begin{aligned}
 V_1 &\leq \left( \sup_{0 \leq j \leq m} \sum_{i=0}^n (\beta \alpha_j^i + (1 - \beta) \beta_j^i) \right) \cdot \sum_{j=0}^m \mu_2(B_j) f_1 \left( \frac{\mu_1(B_j)}{\mu_2(B_j)} \right) \\
 &= \sup_{x \in I} \left( \beta \sum_{i=0}^n h_i^j(x) + (1 - \beta) \sum_{i=0}^n k_i^j(x) \right) \cdot \sum_{j=0}^m \mu_2(B_j) f_1 \left( \frac{\mu_1(B_j)}{\mu_2(B_j)} \right) \\
 &\leq \sup_{x \in I} \left( \beta \sum_{i=0}^n h_i^j(x) + (1 - \beta) \sum_{i=0}^n k_i^j(x) \right) \cdot \int_1 \text{Div}_l(\mu_1, \mu_2) \\
 &\xrightarrow{l \rightarrow \infty} \sup_{x \in I} (\beta h(x)(J) + (1 - \beta) k(x)(J)) \cdot \int_1 \text{Div}_l(\mu_1, \mu_2) \\
 &\leq \int_1 \text{Div}_l(\mu_1, \mu_2)
 \end{aligned}$$

Here, the first inequality is given from the nonnegativity of each  $\mu_2(B_j) f_1 \left( \frac{\mu_1(B_j)}{\mu_2(B_j)} \right)$  derived from  $0 \leq f_1$ ; the equality is given by definition of  $\alpha_j^i$  and  $\beta_j^i$ ; the second inequality can be given by the continuity of  $\int_1 \text{Div}$ ; 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.

Second, we obtain  $\lim_{l \rightarrow \infty} V_2 \leq \gamma \cdot \int_1 \text{Div}_l(\mu_1, \mu_2) \cdot \sup_{x \in I} \int_2 \text{Div}_l(h(x), k(x))$  as follows:

$$V_2 \leq \left( \sup_{0 \leq j \leq m} \sum_{i=0}^n \beta_j^i f_2 \left( \frac{\alpha_j^i}{\beta_j^i} \right) \right) \sum_{j=0}^m \left( \beta' \mu_1(B_j) + (1 - \beta') \mu_2(B_j) \right) + \gamma \sum_{j=0}^m \mu_2(B_j) f_1 \left( \frac{\mu_1(B_j)}{\mu_2(B_j)} \right)$$



$$\begin{aligned}
 &= \left( \sup_{x \in I} \sum_{i=0}^n k_i^i(x) f_2 \left( \frac{h_i^i(x)}{k_i^i(x)} \right) \right) \left( \beta' \mu_1(I) + (1 - \beta') \mu_2(I) + \gamma \sum_{j=0}^m \mu_2(B_j) f_1 \left( \frac{\mu_1(B_j)}{\mu_2(B_j)} \right) \right) \\
 &\leq \left( \sup_{x \in I} \sum_{i=0}^n k_i^i(x) f_2 \left( \frac{h_i^i(x)}{k_i^i(x)} \right) \right) \left( \beta' \mu_1(I) + (1 - \beta') \mu_2(I) + \gamma \cdot f_1 \operatorname{Div}_I(\mu_1, \mu_2) \right) \\
 &\xrightarrow{l \rightarrow \infty} \left( \sup_{x \in I} \sum_{i=0}^n k(x)(A_i) f_2 \left( \frac{h(x)(A_i)}{k(x)(A_i)} \right) \right) \left( \beta' \mu_1(I) + (1 - \beta') \mu_2(I) + \gamma \cdot f_1 \operatorname{Div}_I(\mu_1, \mu_2) \right) \\
 &\leq \sup_{x \in I} f_2 \operatorname{Div}_J(h(x), k(x)) \left( \beta' \mu_1(I) + (1 - \beta') \mu_2(I) + \gamma \cdot f_1 \operatorname{Div}_I(\mu_1, \mu_2) \right) \\
 &\leq \sup_{x \in I} f_2 \operatorname{Div}_J(h(x), k(x)) \cdot \gamma \cdot f_1 \operatorname{Div}_I(\mu_1, \mu_2) \\
 &= \gamma \cdot f_1 \operatorname{Div}_I(\mu_1, \mu_2) \cdot \sup_{x \in I} f_2 \operatorname{Div}_J(h(x), k(x)).
 \end{aligned}$$

Here, the first inequality is proved by the nonnegativity of each

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

which is derived from the first inequality of  $(A')$ ; the first equality is given by definition of  $\alpha_j^i$  and  $\beta_j^i$  and the countable additivity of  $\mu_1$  and  $\mu_2$ ; the second inequality is given by the continuity of  $f_1 \operatorname{Div}$  and  $0 \leq \gamma$ ; 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. We recall that the sequences  $\{h_i^i(x)\}_{i=0}^\infty$  and  $\{k_i^i(x)\}_{i=0}^\infty$  are monotone increasing at each  $x \in I$ . Since  $f_2$  is convex, by Jensen's inequality, the sequence  $\{\sum_{i=0}^n k_i^i(x) f_2(h_i^i(x)/k_i^i(x))\}_{l=0}^\infty$  is monotone increasing at each  $x \in I$ . Then, the sequence  $\{\sup_{x \in I} \sum_{i=0}^n k_i^i(x) f_2(h_i^i(x)/k_i^i(x))\}_{l=0}^\infty$  is monotone increasing, because  $\sum_{i=0}^n k_{l+1}^i(x) f_2(h_{l+1}^i(x)/k_{l+1}^i(x))$  is always greater than  $\sum_{i=0}^n k_l^i(x) f_2(h_l^i(x)/k_l^i(x))$  for all  $l \in \mathbb{N}$  and  $x \in I$ . Hence, from the continuity of  $f_2 \operatorname{Div}$ , we obtain

$$\begin{aligned}
 \lim_{l \rightarrow \infty} \sup_{x \in I} \sum_{i=0}^n k_i^i(x) f_2 \left( \frac{h_i^i(x)}{k_i^i(x)} \right) &= \sup_{l \in \mathbb{N}} \sup_{x \in I} \sum_{i=0}^n k_i^i(x) f_2 \left( \frac{h_i^i(x)}{k_i^i(x)} \right) \\
 &= \sup_{x \in I} \sup_{l \in \mathbb{N}} \sum_{i=0}^n k_i^i(x) f_2 \left( \frac{h_i^i(x)}{k_i^i(x)} \right) \\
 &= \sup_{x \in I} \sum_{i=0}^n k(x)(A_i) f_2 \left( \frac{h(x)(A_i)}{k(x)(A_i)} \right) \\
 &\leq \sup_{x \in I} f_2 \operatorname{Div}(h(x), k(x)).
 \end{aligned}$$

Finally, we obtain  $\lim_{l \rightarrow \infty} V_3 = 0$  as follows:

$$\begin{aligned}
 V_3 &= \sum_{j=0}^m \alpha(\mu_2(B_j) - \mu_1(B_j)) \left( \sum_{i=0}^n \alpha_j^i - \beta_j^i \right) \\
 &= \alpha \left( \int_I h_1^i d\mu_2 - \int_I k_1^i d\mu_2 + \int_I k_1^i d\mu_1 - \int_I h_1^i d\mu_1 \right)
 \end{aligned}$$

$$\begin{aligned} &\xrightarrow{l \rightarrow \infty} \alpha \left( \int_I h(-)(J) d\mu_2 - \int_I k(-)(J) d\mu_2 + \int_I k(-)(J) d\mu_1 - \int_I h(-)(J) d\mu_1 \right) \\ &= 0. \end{aligned}$$

Here, the last equality is derived from the assumption that either  $\alpha = 0$  or  $h(x)(J) = k(x)(J)$  for any  $x \in I$  holds.

From the above evaluations of  $V_1, V_2,$  and  $V_3,$  we obtain the inequality (29). Since the measurable partition  $\{A_i\}_{i=0}^n$  of  $J$  is arbitrary, we conclude (28) by the continuity of  $f \text{Div}$ . This completes the proof.  $\square$

Parameters for Proposition 10 for for weight functions of TV, KL, HD, and Chi are shown in Table 4. Below, we check the conditions in Proposition 10.

- 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

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

- 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

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

- 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  $x, y, z, w \in [0, 1],$

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

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

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

*Proof.* (Proof of Proposition 11)

We first show the monotonicity of  $\langle p, \lambda \rangle^* \Delta$ . Assume  $m \leq m'$ . From the monotonicity of the original  $\Delta$ , we obtain for each  $v_1, v_2 \in U^{\mathbb{C}}(SI),$

$$\begin{aligned} \langle p, \lambda \rangle^* \Delta_I^m(v_1, v_2) &= \Delta_{pI}^m((U^{\mathbb{D}}\lambda_I)(v_1), (U^{\mathbb{D}}\lambda_I)(v_2)) \\ &\geq \Delta_{pI}^{m'}((U^{\mathbb{D}}\lambda_I)(v_1), (U^{\mathbb{D}}\lambda_I)(v_2)) \\ &= \langle p, \lambda \rangle^* \Delta_I^{m'}(v_1, v_2). \end{aligned}$$

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

$$\begin{aligned} (\langle p, \lambda \rangle^* \Delta)_I^{1M} (\eta_I^S \bullet x, \eta_I^S \bullet y) &= \Delta_{pI}^{1M} ((U^{\mathbb{D}} \lambda_I)(U^{\mathbb{C}} \eta_I^S \circ x), (U^{\mathbb{D}} \lambda_I)(U^{\mathbb{C}} \eta_I^S \circ y)) \\ &= \Delta_{pI}^{1M} (U^{\mathbb{D}} \lambda_I \circ U^{\mathbb{D}} p \eta_I^S \circ x, U^{\mathbb{D}} \lambda_I \circ U^{\mathbb{D}} p \eta_I^S \circ y) \\ &= \Delta_{pI}^{1M} ((\lambda_I \circ p \eta_I^S) \bullet x, (\lambda_I \circ p \eta_I^S) \bullet y) \\ &= \Delta_{pI}^{1M} (\eta_{pI}^T \bullet x, \eta_{pI}^T \bullet y) \leq 0. \end{aligned}$$

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

$$\begin{aligned} (\langle p, \lambda \rangle^* \Delta)_J^{mn} (f_1^\sharp \bullet c_1, f_2^\sharp \bullet c_2) &= \Delta_{pJ}^{mn} ((U^{\mathbb{D}} \lambda_J)(U^{\mathbb{C}} (f_1^\sharp) \circ c_1), (U^{\mathbb{D}} \lambda_J)(U^{\mathbb{C}} (f_2^\sharp) \circ c_2)) \\ &= \Delta_{pJ}^{mn} (U^{\mathbb{D}} \lambda_J \circ U^{\mathbb{D}} p (f_1^\sharp) \circ c_1, U^{\mathbb{D}} \lambda_J \circ U^{\mathbb{D}} p (f_2^\sharp) \circ c_2) \\ (*) &= \Delta_{pJ}^{mn} (U^{\mathbb{D}} ((\lambda_J \circ p f_1)^\sharp) \circ U^{\mathbb{D}} \lambda_I \circ c_1, U^{\mathbb{D}} ((\lambda_J \circ p f_2)^\sharp) \circ U^{\mathbb{D}} \lambda_I \circ c_2) \\ &= \Delta_{pJ}^{mn} ((\lambda_J \circ p f_1)^\sharp \bullet (\lambda_I \bullet c_1), (\lambda_J \circ p f_2)^\sharp \bullet (\lambda_I \bullet c_2)) \\ &\leq \Delta_{pI}^m (\lambda_I \bullet c_1, \lambda_I \bullet c_2) + \sup_{(x,y) \in EpI} \Delta_{pJ}^n ((\lambda_J \circ p f_1) \bullet x, (\lambda_J \circ p f_2) \bullet y) \\ &= (\langle p, \lambda \rangle^* \Delta)_I^m (c_1, c_2) + \sup_{(x,y) \in FI} (\langle p, \lambda \rangle^* \Delta)_J^n (f_1 \bullet x, f_2 \bullet y). \end{aligned}$$

To prove the equality (\*), we calculate

$$\begin{aligned} U^{\mathbb{D}} \lambda_J \circ U^{\mathbb{D}} p (f_1^\sharp) &= U^{\mathbb{D}} (\lambda_J \circ p \mu_J^S \circ p S f_1) = U^{\mathbb{D}} (\mu_{pJ}^T \circ T \lambda_J \circ \lambda_{SJ} \circ p S f_1) \\ &= U^{\mathbb{D}} (\mu_{pJ}^T \circ T \lambda_J \circ T p f_1 \circ \lambda_I) = U^{\mathbb{D}} ((\lambda_J \circ p f_1)^\sharp \bullet \lambda_I). \end{aligned}$$

This completes the proof. □

**Proof.** (Proof of Proposition 13)

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

$$\begin{aligned} \Delta_I^{\text{lip}, d_S} (\eta_I(x), \eta_I(y)) &= \sup_{s', s \in S} \frac{d_S(\pi_2(x, s), \pi_2(y, s'))}{d_S(s, s')} = \frac{d_S(s, s')}{d_S(s, s')} = 1, \\ \Delta_J^{\text{lip}, d_S} (F_1^\sharp(f_1), F_2^\sharp(f_2)) &= \sup_{s', s \in S} \frac{d_S(\pi_2(F_1(\pi_1 f_1(s)))(\pi_2 f_1(s)), \pi_2(F_2(\pi_1 f_2(s')))(\pi_2 f_2(s'))))}{d_S(s, s')} \\ &= \sup_{s', s \in S} \frac{d_S(\pi_2 f_1(s), \pi_2 f_2(s'))}{d_S(s, s')} \cdot \frac{d_S(\pi_2(F_1(\pi_1 f_1(s)))(\pi_2 f_1(s)), \pi_2(F_2(\pi_1 f_2(s')))(\pi_2 f_2(s'))))}{d_S(\pi_2 f_1(s), \pi_2 f_2(s'))} \\ &\leq \sup_{s', s \in S} \frac{d_S(\pi_2 f_1(s), \pi_2 f_2(s'))}{d_S(s, s')} \cdot \sup_{t', t \in S} \frac{d_S(\pi_2(F_1(\pi_1 f_1(s)))(t), \pi_2(F_2(\pi_1 f_2(s')))(t'))}{d_S(t, t')} \\ &\leq \Delta_I^{\text{lip}, d_S} (f_1, f_2) \cdot \sup_{x, y \in I} \Delta_J^{\text{lip}, d_S} (F_1(x), F_2(y)) \end{aligned}$$

Here,  $F_1, F_2 : I \rightarrow T_S J$  and  $f_1, f_2 \in T_S I$ . □

*Proof.* (Proof of Proposition 14)

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

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

Here,  $F_1, F_2 : I \rightarrow T_S J$  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 15) We first show the Eq-unit reflexivity of  $d^{T_S(-)}$ . For any  $s \in S$ , we calculate

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

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  $F_1, F_2 : (I, d_I) \rightarrow T_S(J, d_J)$ , we compute

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

We note here that the nonexpansivity of  $F_2: (I, d_I) \rightarrow (S, d_S) \Rightarrow (S, d_S) \times (J, d_J)$  is equivalent to the one of its uncurrying  $\overline{F}_2: (S, d_S) \times (I, d_I) \rightarrow (S, d_S) \times (J, d_J)$ .  $\square$

*Proof.* (Proof of Proposition 16) We first show the  $\text{Dist}_0$ -unit reflexivity of  $\Delta^{\text{Dist}_0}$ . For  $(x_1, x_2) \in \text{Dist}_0(I, d_I)$  (i.e.  $d_I(x_1, x_2) = 0$ ), we calculate

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

Next, we show the  $\text{Dist}_0$ -composability of  $\Delta^{\text{Dist}_0}$ . For any  $f_1, f_2 \in T_S(I, d_I)$  and nonexpansive functions  $F_1, F_2: (I, d_I) \rightarrow T_S(J, d_J)$ , we compute

$$\begin{aligned} &\Delta_J^{\text{Dist}_0}(F_1^\sharp(f_1), F_2^\sharp(f_2)) \\ &= \sup_{d_S(s_1, s_2)=0} \max \left( 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))), \right. \\ &\quad \left. 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))) \right) \\ &\leq \sup_{d_S(s_1, s_2)=0} \max \left( 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))), \right. \\ &\quad \left. 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))), \right. \\ &\quad \left. 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))), \right. \\ &\quad \left. 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))) \right) \\ &= \sup_{d_S(s_1, s_2)=0} \max \left( 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))), \right. \\ &\quad \left. 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))), \right. \\ &\quad \left. 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))), \right. \\ &\quad \left. 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))) \right) \\ &\leq \sup_{d_S(s_1, s_2)=0} \max \left( \begin{array}{l} 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 \text{Dist}_0(I, d_I)} \sup_{d_S(s'_1, s'_2)=0} \max \left( d_J(\pi_1(F_1(x)(s'_1)), \pi_1(F_2(x)(s'_2))), \right. \right. \\ \left. \left. d_S(\pi_2(F_1(x)(s'_1)), \pi_2(F_2(x)(s'_2))) \right) \end{array} \right) \\ &= \max \left( \begin{array}{l} \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 \text{Dist}_0(I, d_I)} \sup_{d_S(s'_1, s'_2)=0} \max \left( d_J(\pi_1(F_1(x_1)(s'_1)), \pi_1(F_2(x_2)(s'_2))), \right. \right. \\ \left. \left. d_S(\pi_2(F_1(x_1)(s'_1)), \pi_2(F_2(x_2)(s'_2))) \right) \end{array} \right) \\ &= \max(\Delta_I^{\text{Dist}_0}(f_1, f_2), \sup_{(x_1, x_2) \in \text{Dist}_0(I, d_I)} \Delta_J^{\text{Dist}_0}(F_1(x), F_2(x))). \end{aligned}$$

This completes the proof.  $\square$

*Proof.* (Proof of Proposition 17) Since  $M = 1$ , the monotonicity of  $C(\Delta, N)$  is obvious.

We first show the Eq-unit reflexivity of  $C(\Delta, N)$ . We recall that for all  $x \in UI$ , we have

$$\begin{aligned} T\pi_1 \bullet (\eta_I^{T(N \times -)} \bullet x) &= (T\pi_1 \circ \eta_I^{T(N \times -)}) \bullet x &= (T\pi_1 \circ \eta_{N \times I}^T \circ \eta_I^{(N \times -)}) \bullet x \\ &= (\eta_N^T \circ \pi_1 \circ \langle 1_N \circ !_I, \text{id}_I \rangle) \bullet x &= (\eta_N^T \circ 1_N \circ !_I) \bullet x \\ &= \eta^T \bullet ((1_N \circ !_I) \bullet x) &= \eta^T \bullet (1_N \bullet (!_I \bullet x)) \\ &= \eta^T \bullet (1_N \bullet \text{id}_I) &= \eta^T \bullet U1_N. \end{aligned}$$

Hence,

$$\begin{aligned} C(\Delta, N)_I(\eta^{T(N \times -)} \bullet x, \eta^{T(N \times -)} \bullet x) &= C(\Delta, N)_I(\eta^{T(N \times -)} \bullet x, \eta^{T(N \times -)} \bullet x) \\ &= \Delta_N(T\pi_1 \bullet (\eta^{T(N \times -)} \bullet x), T\pi_1 \bullet (\eta^{T(N \times -)} \bullet x)) \\ &= \Delta_N(\eta^T \bullet U1_N, \eta^T \bullet U1_N) \\ &\leq 0_{\varnothing}. \end{aligned}$$

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

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

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

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

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

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

This completes the proof. □

*Proof.* (Proof of Proposition 18)

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

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

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  $c \in TI$ ,

$$(\Delta_I^\sqsubseteq(c, c) \leq 1) \iff (c \sqsubseteq_I c).$$

Since  $\sqsubseteq$  is a preorder on  $T$ , for any  $I, J \in \mathbf{Set}$ ,  $c_1, c_2 \in TI$  and  $f, g: I \rightarrow TJ$ ,

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

Hence, we have the Eq-composability:

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

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

Each component  $\sqsubseteq_I^\Delta$  of  $\sqsubseteq^\Delta$  at set  $I$  is a preorder on the set  $TI$ . We here note that the divergence  $\Delta$  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  $\sqsubseteq_I^\Delta$ :

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

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

$$\forall c_1, c_2 \in TI, f: I \rightarrow TJ. \Delta_J(f^\sharp(c_1), f^\sharp(c_2)) \leq \Delta_I(c_1, c_2), \tag{30}$$

$$\forall c \in TI, f, g: I \rightarrow TJ. \Delta_J(f^\sharp(c), g^\sharp(c)) \leq \sup_{x \in I} \Delta_J(f(x), g(x)). \tag{31}$$

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

$$(30) \iff \forall c_1, c_2 \in TI, f: I \rightarrow TJ. (c_1 \sqsubseteq_I^\Delta c_2 \implies f^\sharp(c_1) \sqsubseteq_J^\Delta f^\sharp(c_2)),$$

$$(31) \iff \forall c \in TI, f, g: I \rightarrow TJ. (\forall x \in I. f(x) \sqsubseteq_J^\Delta g(x) \implies f^\sharp(c) \sqsubseteq_J^\Delta g^\sharp(c)).$$

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

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

This completes the proof. □

**Appendix C. Proofs for Section 5 (Properties of Divergences on Monads)**

*Proof.* (Proof of Theorem 21) First, it is easy to see that the inequality (3) is equivalent to  $\Delta$  satisfying  $E$ -unit reflexivity.

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

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

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

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

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

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

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

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

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

□

*Proof.* (Proof of Lemma 22) We write  $V$  for  $V_{\mathcal{Q}, \mathbb{C}}$ . From  $x + \top = \top$ , we obtain the following equivalence (below  $z_1, z_2 \in U(VZ)$  and  $x_1, x_2 \in U(VX)$ ):

$$\begin{aligned} f \in \mathbf{Div}_{\mathcal{Q}}(\mathbb{C})(Z \otimes X, Y) &\iff \forall z_1, z_2, x_1, x_2. d_Y(f \bullet \langle z_1, x_1 \rangle, f \bullet \langle z_2, x_2 \rangle) \leq d_Z(z_1, z_2) + d_X(x_1, x_2) \\ &\iff \forall z_1, z_2, x_1, x_2. d_X(x_1, x_2) = 0. d_Y(f \bullet \langle z_1, x_1 \rangle, f \bullet \langle z_2, x_2 \rangle) \leq d_Z(z_1, z_2) \\ &\iff \forall z_1, z_2. \sup_{x_1, x_2, d_X(x_1, x_2) = 0} d_Y(f \bullet \langle z_1, x_1 \rangle, f \bullet \langle z_2, x_2 \rangle) \leq d_Z(z_1, z_2) \\ &\iff \forall z_1, z_2. d_{X \rightarrow Y}(\lambda(f) \bullet z_1, \lambda(f) \bullet z_2) \leq d_Z(z_1, z_2) \\ &\iff \lambda(f) \in \mathbf{Div}_{\mathcal{Q}}(\mathbb{C})(Z, X \rightarrow Y). \end{aligned}$$

This shows that the currying bijection  $\lambda : \mathbb{C}(VZ \times VX, VY) \rightarrow \mathbb{C}(VZ, VX \Rightarrow VY)$  restricts to the one from  $\mathbf{Div}_{\mathcal{Q}}(\mathbb{C})(Z \otimes X, Y)$  to  $\mathbf{Div}_{\mathcal{Q}}(\mathbb{C})(Z, X \rightarrow Y)$ , showing that  $V_{\mathcal{Q}, \mathbb{C}}$  is a map of adjunction.

□

*Proof.* (Proof of Theorem 25)  $[\Delta]$  is a graded variant of the *codensity lifting* performed along the fibration  $V_{\mathcal{Q}, \mathbb{C}} : \mathbf{Div}_{\mathcal{Q}}(\mathbb{C}) \rightarrow \mathbb{C}$  (Katsumata et al., 2018, see also Definition 37). Proving that it is a graded lifting of  $T$  is routine. We show  $\Delta_I^m = [\Delta]m(E^{\mathcal{Q}}I)$ . The direction  $[\Delta]m(E^{\mathcal{Q}}I) \leq_{TI} \Delta_I^m$  is easy. We show the converse. From the composability of  $\Delta$ , for any  $c_1, c_2 \in U(TI)$ ,  $J \in \mathbb{C}$ ,  $n \in M$  and  $f \in \mathbf{Div}_{\mathcal{Q}}(\mathbb{C})(E^{\mathcal{Q}}I, \Delta_J^n)$ , we have

$$\Delta_J^{m \cdot n}(f^\# \bullet c_1, f^\# \bullet c_2) \leq \Delta_I^m(c_1, c_2) + \sup_{(x_1, x_2) \in EI} \Delta_J^n(f \bullet x_1, f \bullet x_2).$$

Next, the nonexpansivity of  $f$  is equivalent to

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

Therefore, we conclude  $\Delta_J^{m \cdot n}(f^\# \bullet c_1, f^\# \bullet c_2) \leq \Delta_I^m(c_1, c_2)$ . By discharging  $J, n, f$ , we conclude the inequality  $\Delta_I^m \leq_{TI} [\Delta]m(E^{\mathcal{Q}}I)$ .

□



*Proof.* (Proof of Theorem 26) 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:

$$\begin{aligned} & \theta \in \mathbf{Div}_{\mathcal{Q}}(\mathbb{C})(X \otimes [\Delta]m(E^{\mathcal{Q}}J), [\Delta]m(X \otimes E^{\mathcal{Q}}J)) \\ \iff & \left( \begin{array}{l} \forall x_1, x_2 \in UI, c_1, c_2 \in U(TJ) . \\ d_{[\Delta]m(X \otimes E^{\mathcal{Q}}J)}(\theta \bullet \langle x_1, c_1 \rangle, \theta \bullet \langle x_2, c_2 \rangle) \leq d(x_1, x_2) + d_{[\Delta]m(E^{\mathcal{Q}}J)}(c_1, c_2) \end{array} \right) \\ \iff & \left( \begin{array}{l} \forall x_1, x_2 \in UI, c_1, c_2 \in U(TJ), K \in \mathbb{C}, n \in M, f \in \mathbf{Div}_{\mathcal{Q}}(\mathbb{C})(X \otimes E^{\mathcal{Q}}J, \Delta_K^n) . \\ \Delta_K^{m \cdot n}(f^{\sharp} \bullet \theta \bullet \langle x_1, c_1 \rangle, f^{\sharp} \bullet \theta \bullet \langle x_2, c_2 \rangle) \leq d(x_1, x_2) + d_{[\Delta]m(E^{\mathcal{Q}}J)}(c_1, c_2) \end{array} \right) \\ \overset{\dagger}{\iff} & \left( \begin{array}{l} \forall x_1, x_2 \in UI, c_1, c_2 \in U(TJ), K \in \mathbb{C}, n \in M, f \in \mathbf{Div}_{\mathcal{Q}}(\mathbb{C})(X \otimes E^{\mathcal{Q}}J, \Delta_K^n) . \\ \Delta_K^{m \cdot n}((f_{x_1})^{\sharp} \bullet c_1, (f_{x_2})^{\sharp} \bullet c_2) \leq d(x_1, x_2) + d_{[\Delta]m(E^{\mathcal{Q}}J)}(c_1, c_2) \end{array} \right) . \end{aligned}$$

In the step  $\overset{\dagger}{\iff}$ , we use 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  $f \in \mathbf{Div}_{\mathcal{Q}}(\mathbb{C})(X \otimes E^{\mathcal{Q}}J, \Delta_K^n)$ . First, from the composability of  $\Delta$ , we obtain

$$\Delta_K^{m \cdot n}((f_{x_1})^{\sharp} \bullet c_1, (f_{x_2})^{\sharp} \bullet c_2) \leq \Delta_J^m(c_1, c_2) + \sup_{(y_1, y_2) \in EJ} \Delta_K^n(f_{x_1} \bullet y_1, f_{x_2} \bullet y_2).$$

We look at summands of the right-hand side. First, we have  $\Delta_J^m(c_1, c_2) \leq d_{[\Delta]m(E^{\mathcal{Q}}J)}(c_1, c_2)$  by Theorem 25. Next, from the nonexpansivity of  $f$ , for any  $x_1, x_2 \in UI, y_1, y_2 \in UJ$ , we have

$$\Delta_K^n(f_{x_1} \bullet y_1, f_{x_2} \bullet y_2) = \Delta_K^n(f \bullet \langle x_1, y_1 \rangle, f \bullet \langle x_2, y_2 \rangle) \leq d(x_1, x_2) + E^{\mathcal{Q}}J(y_1, y_2).$$

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

$$\forall x_1, x_2 \in UI . \sup_{(y_1, y_2) \in EJ} \Delta_K^n(f_{x_1} \bullet y_1, f_{x_2} \bullet y_2) \leq d(x_1, x_2).$$

Therefore, we obtain the goal:

$$\Delta_K^{m \cdot n}((f_{x_1})^{\sharp} \bullet c_1, (f_{x_2})^{\sharp} \bullet c_2) \leq d(x_1, x_2) + d_{[\Delta]m(E^{\mathcal{Q}}J)}(c_1, c_2).$$

Next, let  $\dot{T} \in \mathbf{SGDLift}(T, E, M, \mathcal{Q})$ . We show  $\langle \dot{T} \rangle \in \mathbf{Div}(T, E, M, \mathcal{Q})$ .

The unit law of  $\dot{T}$  immediately entails

$$\eta_I \in \mathbf{Div}_{\mathcal{Q}}(\mathbb{C})(E^{\mathcal{Q}}I, \dot{T}1(E^{\mathcal{Q}}I)).$$

Next, under the assumption on  $(\mathbb{C}, T)$  and  $\mathcal{Q}$ , in  $\mathbf{Div}_{\mathcal{Q}}(\mathbb{C})$  the functor  $(-)\otimes E^{\mathcal{Q}}I$  has a right adjoint  $E^{\mathcal{Q}}I \multimap (-)$  above the adjunction  $(-)\times I \dashv I \Rightarrow (-)$  (Lemma 22). Therefore, each component of the uncurried bind morphism  $\text{ub}$  given in (7) are nonexpansive morphisms in  $\mathbf{Div}_{\mathcal{Q}}(\mathbb{C})$ :

$$\begin{array}{c} \langle \dot{T} \rangle m(E^{\mathcal{Q}}I) \otimes (E^{\mathcal{Q}}I \multimap \langle \dot{T} \rangle n(E^{\mathcal{Q}}J)) \\ \downarrow \langle \pi_2, \pi_1 \rangle \\ (E^{\mathcal{Q}}I \multimap \langle \dot{T} \rangle n(E^{\mathcal{Q}}J)) \otimes \langle \dot{T} \rangle m(E^{\mathcal{Q}}I) \\ \downarrow \theta \\ \langle \dot{T} \rangle m((E^{\mathcal{Q}}I \multimap \langle \dot{T} \rangle n(E^{\mathcal{Q}}J)) \otimes E^{\mathcal{Q}}I) \\ \downarrow \text{ev}^{\sharp} \\ \langle \dot{T} \rangle (m \cdot n)(E^{\mathcal{Q}}J) \end{array}$$

Therefore, we conclude

$$\text{ub} \in \mathbf{Div}_{\mathcal{Q}}(\mathbb{C})(\langle \dot{T} \rangle m(E^{\mathcal{Q}}I) \otimes (E^{\mathcal{Q}}I \multimap \langle \dot{T} \rangle n(E^{\mathcal{Q}}J)), \langle \dot{T} \rangle (m \cdot n)(E^{\mathcal{Q}}J)).$$

We also easily have monotonicity:  $\langle \dot{T} \rangle m(E^{\mathcal{Q}}I) \leq \langle \dot{T} \rangle n(E^{\mathcal{Q}}I)$  for  $m \leq n$  by condition 1 of graded divergence lifting. We thus conclude that  $\langle \dot{T} \rangle mE^{\mathcal{Q}}I \in \mathbf{Div}(T, E, M, \mathcal{Q})$ .

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

$$\sup_{n \in M, J \in \mathbb{C}, f \in \mathbf{Div}_{\mathcal{Q}}(\mathbb{C})(X, \dot{T}n(E^{\mathcal{Q}}J))} d_{\langle \dot{T} \rangle (m \cdot n)(E^{\mathcal{Q}}J)}(f^{\sharp}(c_1), f^{\sharp}(c_2)) \leq d_{\dot{T}mX}(c_1, c_2). \tag{32}$$

Let  $n \in M, J \in \mathbb{C}, f \in \mathbf{Div}_{\mathcal{Q}}(\mathbb{C})(X, \dot{T}n(E^{\mathcal{Q}}J))$ . Since  $\dot{T}$  is an  $M$ -graded  $\mathcal{Q}$ -divergence lifting of  $T$ , we obtain

$$f^{\sharp} \in \mathbf{Div}_{\mathcal{Q}}(\mathbb{C})(\dot{T}mX, \dot{T}(m \cdot n)(E^{\mathcal{Q}}J)).$$

This implies the inequality  $d_{\langle \dot{T} \rangle (m \cdot n)(E^{\mathcal{Q}}J)}(f^{\sharp}(c_1), f^{\sharp}(c_2)) \leq d_{\dot{T}mX}(c_1, c_2)$  in  $\mathcal{Q}$ . By taking the sup for  $n, J, f$ , we obtain the inequality (32). □

*Proof.* (Proof of Proposition 28) We write  $|1| = \{*\}$ . We first check the measurable isomorphism  $G_s1 \cong [0, 1]$ . The measurable functions  $\text{ev}_{\{*\}}: G_s1 \rightarrow [0, 1]$  ( $v \mapsto v(*)$ ) and the function  $H: |[0, 1]| \rightarrow |G_s1|$  ( $r \mapsto r \cdot \mathbf{d}_*$ ) are mutually inverse. For any (Borel-)measurable  $U \in \Sigma_{[0,1]}$ , we have  $H^{-1}(\text{ev}_{\{*\}}^{-1}(U)) = U$  and  $H^{-1}(\text{ev}_{\emptyset}^{-1}(U)) = [0, 1]$  if  $0 \in U$  and  $H^{-1}(\text{ev}_{\emptyset}^{-1}(U)) = \emptyset$  otherwise. Since all generators of  $\Sigma_{G_s1}$  are  $\text{ev}_{\{*\}}^{-1}(U)$  and  $\text{ev}_{\emptyset}^{-1}(U)$  where  $U \in \Sigma_{[0,1]}$ , we conclude the measurability of  $H: [0, 1] \rightarrow G_s1$ . Thus,  $f: I \rightarrow [0, 1]$  corresponds bijectively to  $H \circ f: I \rightarrow G_s1$ , and

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

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

$$\begin{aligned} \text{DP}_I^{\varepsilon}(\nu_1, \nu_2) &= \sup_{S \in \Sigma_I} (\nu_1(S) - \exp(\varepsilon)\nu_2(S)) \\ &= \sup_{S \in \Sigma_I} \left( \int_I \chi_S d\nu_1 - \exp(\varepsilon) \int_I \chi_S d\nu_2 \right) \\ &\leq \sup_{f: I \rightarrow [0,1]} \left( \int_I f d\nu_1 - \exp(\varepsilon) \int_I f d\nu_2 \right) \\ &= \sup_{f: I \rightarrow [0,1]} \left( ((H \circ f)^{\sharp} \nu_1)(\{*\}) - \exp(\varepsilon)((H \circ f)^{\sharp} \nu_2)(\{*\}) \right) \\ &\leq \sup_{f: I \rightarrow [0,1]} \sup_{S' \in \Sigma_1 (\iff S' = \{*\}, \emptyset)} \left( ((H \circ f)^{\sharp} \nu_1)(S') - \exp(\varepsilon)((H \circ f)^{\sharp} \nu_2)(S') \right) \\ &= \sup_{f: I \rightarrow [0,1]} \text{DP}_1^{\varepsilon}((H \circ f)^{\sharp} \nu_1, (H \circ f)^{\sharp} \nu_2) \\ &= \sup_{g: I \rightarrow G_s1} \text{DP}_1^{\varepsilon}(g^{\sharp} \nu_1, g^{\sharp} \nu_2) \\ &\leq \text{DP}_I^{\varepsilon}(\nu_1, \nu_2). \end{aligned}$$

The first inequality is given by  $\nu(S) = \int_I \chi_S d\nu$  where  $\chi_S: I \rightarrow [0, 1]$  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 29) We first prove that TV is not 1-generated. We write  $|2| = \{0, 1\}$ . We define  $\nu_1, \nu_2 \in G_s 2$  by:

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

Then the total variation distance between them is calculated by:

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

On the other hand, for any  $f: 2 \rightarrow G_s 1$ , we have

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

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  $\nu_1, \nu_2 \in G_s I$ ,

$$\text{TV}_I(\nu_1, \nu_2) \geq \sup_{g: I \rightarrow G_s 2} \text{TV}_2(g^\# \nu_1, g^\# \nu_2).$$

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  $\mu$  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  $\nu_1, \nu_2$  with respect to  $\mu$ , 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  $g: I \rightarrow G_s 2$  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

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

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

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

We then conclude that  $\Delta^{\text{TV}}$  is 2-generated. □

*Proof.* (Proof of Proposition 30) For all set  $J$  and  $c_1, c_2 \in TJ$ , we have

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

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

**Lemma 47.** For any  $U \in \mathbf{QET}(\Sigma, \Omega)$ ,  $t, u \in T_\Sigma\Omega$  and  $\varepsilon \in \mathbb{Q}^+$ , we have

$$D[U](t, u) \leq \varepsilon \iff \emptyset \vdash t =_\varepsilon u \in U.$$

*Proof.* ( $\implies$ ) Assume  $D[U](t, u) \leq \varepsilon$ . We first fix an arbitrary  $\varepsilon' \in \mathbb{Q}^+$  such that  $\varepsilon < \varepsilon'$ . By the definition of  $D[U]$ , there is  $\varepsilon^* \in \mathbb{Q}^+$  such that  $D[U](t, u) \leq \varepsilon^* < \varepsilon'$  and  $\emptyset \vdash t =_{\varepsilon^*} u \in U$ . Here  $(\varepsilon - \varepsilon^*) \in \mathbb{Q}^+$  and  $\varepsilon' = \varepsilon^* + (\varepsilon' - \varepsilon^*)$ . Therefore, by (Max) and (Cut), we conclude  $\emptyset \vdash t =_{\varepsilon'} u \in U$ . Since  $\varepsilon'$  is arbitrary, we obtain  $\{\emptyset \vdash t =_{\varepsilon'} u \mid \varepsilon' \in \mathbb{Q}^+, \varepsilon < \varepsilon'\} \subseteq U$ . Hence, by (Arch) and (Cut), we have  $\emptyset \vdash t =_\varepsilon u \in U$ . ( $\impliedby$ ) Obvious. □

**Lemma 48.** For any  $U \in \mathbf{QET}(\Sigma, \Omega)$ , the function  $D[U]: (T_\Sigma\Omega)^2 \rightarrow \mathcal{R}^+$  defined by (9) is a CS-PMet on  $T_\Sigma\Omega$ .

*Proof.* That  $D[U]$  is a pseudometric is shown in the beginning of (Mardare et al., 2016, Section 5).

We check the substitutivity. Let  $t, u \in T_\Sigma\Omega$  and  $h: \Omega \rightarrow T_\Sigma\Omega$ . By (Subst), we have

$$\forall \varepsilon \in \mathbb{Q}^+. \emptyset \vdash t =_\varepsilon u \in U \implies \emptyset \vdash h^\sharp(t) =_\varepsilon h^\sharp(u) \in U.$$

Since  $\varepsilon$  is arbitrary, we conclude the substitutivity as follows:

$$\begin{aligned} D[U](h^\sharp(t), h^\sharp(u)) &= \inf \{ \varepsilon \in \mathbb{Q}^+ \mid \emptyset \vdash h^\sharp(t) =_\varepsilon h^\sharp(u) \in U \} \\ &\leq \inf \{ \varepsilon \in \mathbb{Q}^+ \mid \emptyset \vdash t =_\varepsilon u \in U \} \\ &= D[U](t, u). \end{aligned}$$

We check the congruence. Let  $t \in T_\Sigma I$  and  $h_1, h_2: I \rightarrow T_\Sigma\Omega$ . By unfolding the structure of  $t$  and applying (Nonexp) and (Cut) repeatedly, we have the implication:

$$\forall i \in I, \varepsilon \in \mathbb{Q}^+. \emptyset \vdash h_1(i) =_\varepsilon h_2(i) \in U \implies \emptyset \vdash h_1^\sharp(t) =_\varepsilon h_2^\sharp(t) \in U. \tag{33}$$

From (33) and Lemma 47, we conclude the congruence as follows:

$$\begin{aligned} D[U](h_1^\sharp(t), h_2^\sharp(t)) &= \inf \left\{ \varepsilon \in \mathbb{Q}^+ \mid \emptyset \vdash h_1^\sharp(t) =_\varepsilon h_2^\sharp(t) \in U \right\} \\ &\leq \inf \left\{ \varepsilon \in \mathbb{Q}^+ \mid \forall i \in I. \emptyset \vdash h_1(i) =_\varepsilon h_2(i) \in U \right\} \\ &= \inf \left\{ \varepsilon \in \mathbb{Q}^+ \mid \sup_{i \in I} D[U](h_1(i), h_2(i)) \leq \varepsilon \right\} \\ &= \sup_{i \in I} D[U](h_1(i), h_2(i)). \end{aligned}$$

Here, the last equality follows from the density of  $\mathbb{Q}^+$ .

The monotonicity of  $D[-]: (\mathbf{QET}(\Sigma, \Omega), \subseteq) \rightarrow (\mathbf{CSPMet}(T_\Sigma, \Omega), \preceq)$  is shown as:

$$\begin{aligned} U \subseteq V &\implies \forall t, u \in T_\Sigma \Omega . \inf \{ \varepsilon \in \mathbb{Q}^+ \mid \emptyset \vdash t =_\varepsilon u \in U \} \geq \inf \{ \varepsilon \in \mathbb{Q}^+ \mid \emptyset \vdash t =_\varepsilon u \in V \} \\ &\iff \forall t, u \in T_\Sigma \Omega . D[U](t, u) \geq D[V](t, u) \\ &\iff D[U] \preceq D[V]. \end{aligned}$$

□

**Lemma 49.** *Let  $T$  be a monad on  $\mathbf{Set}$  and  $\Omega \in \mathbf{Set}$ . For any  $d \in \mathbf{CSPMet}(T, \Omega)$ , the family  $\text{Gen}(d) = \{ \text{Gen}(d)_I: (TI)^2 \rightarrow \mathcal{R}^+ \}_{I \in \mathbf{Set}}$  defined by (11) is an  $\Omega$ -generated Eq-relative  $\mathcal{R}^+$ -divergence on  $T$  where each  $\text{Gen}(d)_I$  is a pseudometric.*

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

$$\text{Gen}(d)_I(c, c) = \sup_{k: I \rightarrow T\Omega} d(k^\sharp(c), k^\sharp(c)) = 0.$$

Hence, the Eq-unit reflexivity of  $\text{Gen}(d)$  follows. From the symmetry of  $d$ , we have the symmetry of  $\text{Gen}(d)_I$ : for each  $c_1, c_2 \in TI$ ,

$$\text{Gen}(d)_I(c_1, c_2) = \sup_{k: I \rightarrow T\Omega} d(k^\sharp(c_1), k^\sharp(c_2)) = \sup_{k: I \rightarrow T\Omega} d(k^\sharp(c_2), k^\sharp(c_1)) = \text{Gen}(d)_I(c_2, c_1).$$

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

$$\begin{aligned} \text{Gen}(d)_I(c_1, c_3) &= \sup_{k: I \rightarrow T\Omega} d(k^\sharp(c_1), k^\sharp(c_3)) \\ &\leq \sup_{k: I \rightarrow T\Omega} d(k^\sharp(c_1), k^\sharp(c_2)) + d(k^\sharp(c_2), k^\sharp(c_3)) \\ &\leq \sup_{k: I \rightarrow T\Omega} d(k^\sharp(c_1), k^\sharp(c_2)) + \sup_{k: I \rightarrow T\Omega} d(k^\sharp(c_2), k^\sharp(c_3)) \\ &= \text{Gen}(d)_I(c_1, c_2) + \text{Gen}(d)_I(c_2, c_3). \end{aligned}$$

Using the reflexivity, congruence, and substitutivity of  $d$  and the triangle inequality of  $\text{Gen}(d)_I$ , we next show the composability. Let  $c_1, c_2 \in TI$  and  $f_1, f_2: I \rightarrow TJ$ . We obtain

$$\begin{aligned} &\text{Gen}(d)_J(f_1^\sharp(c_1), f_2^\sharp(c_2)) \\ &\leq \text{Gen}(d)_J(f_1^\sharp(c_1), f_1^\sharp(c_2)) + \text{Gen}(d)_J(f_1^\sharp(c_2), f_2^\sharp(c_2)) \\ &= \sup_{k: J \rightarrow T\Omega} d((k^\sharp \circ f_1)^\sharp(c_1), (k^\sharp \circ f_1)^\sharp(c_2)) + \sup_{k: J \rightarrow T\Omega} d((k^\sharp \circ f_1)^\sharp(c_2), (k^\sharp \circ f_2)^\sharp(c_2)) \\ &\leq \sup_{l: I \rightarrow T\Omega} d(l^\sharp(c_1), l^\sharp(c_2)) + \sup_{k: J \rightarrow T\Omega} \sup_{i \in I} d(k^\sharp \circ f_1(i), k^\sharp \circ f_2(i)) \\ &= \sup_{l: I \rightarrow T\Omega} d(l^\sharp(c_1), l^\sharp(c_2)) + \sup_{i \in I} \sup_{k: J \rightarrow T\Omega} d(k^\sharp \circ f_1(i), k^\sharp \circ f_2(i)) \\ &= \text{Gen}(d)_I(c_1, c_2) + \sup_{i \in I} \text{Gen}(d)_J(f_1(i), f_2(i)). \end{aligned}$$

We show the  $\Omega$ -generatedness of  $\text{Gen}(d)$ :

$$\begin{aligned} \text{Gen}(d)_I(c_1, c_2) &= \sup_{l: I \rightarrow T\Omega} d(l^\sharp(c_1), l^\sharp(c_2)) \\ (*) &= \sup_{h: \Omega \rightarrow T\Omega, k: I \rightarrow T\Omega} d((h^\sharp \circ k)^\sharp(c_1), (h^\sharp \circ k)^\sharp(c_2)) \\ &= \sup_{k: I \rightarrow T\Omega} \sup_{h: \Omega \rightarrow T\Omega} d(h^\sharp(k^\sharp(c_1)), h^\sharp(k^\sharp(c_2))) \end{aligned}$$

$$= \sup_{k: I \rightarrow T\Omega} \text{Gen}(d)_\Omega(k^\sharp(c_1), k^\sharp(c_2)).$$

The step (\*) uses the equality  $\{l \mid l: I \rightarrow T\Omega\} = \{h^\sharp \circ k \mid h: \Omega \rightarrow T\Omega, k: I \rightarrow T\Omega\}$ .

Gen is indeed a monotone function of type  $(\mathbf{CSPMet}(T, \Omega), \leq) \rightarrow (\mathbf{PMet}(T, \Omega), \leq)$  because

$$\begin{aligned} d \leq d' &\implies \forall I \in \mathbf{Set}. \forall c_1, c_2 \in TI. \sup_{k: I \rightarrow T\Omega} d(k^\sharp(c_1), k^\sharp(c_2)) \geq \sup_{k: I \rightarrow T\Omega} d'(k^\sharp(c_1), k^\sharp(c_2)) \\ &\iff \forall I \in \mathbf{Set}. \forall c_1, c_2 \in TI. \text{Gen}(d)_I(c_1, c_2) \geq \text{Gen}(d')_I(c_1, c_2) \\ &\iff \text{Gen}(d) \leq \text{Gen}(d'). \end{aligned}$$

□

*Proof.* (Proof of Theorem 34) This is proved by Lemma 47, 48, 49. □

*Proof.* (Proof of Theorem 35) Let  $d \in \mathbf{CSPMet}(T_\Sigma, \Omega)$ . We show  $(\text{Gen}(d))_\Omega = d$ . Let  $t, u \in T_\Sigma\Omega$ . From the substitutivity of  $d$ , we have

$$\sup_{k: \Omega \rightarrow T_\Sigma\Omega} d(k^\sharp(t), k^\sharp(u)) \leq d(t, u).$$

On the other hand,  $d(t, u) = d(\eta_\Omega^\sharp(t), \eta_\Omega^\sharp(u))$ . Therefore, we conclude

$$(\text{Gen}(d))_\Omega(t, u) = \sup_{k: \Omega \rightarrow T_\Sigma\Omega} d(k^\sharp(t), k^\sharp(u)) = d(t, u).$$

Let  $\Delta \in \mathbf{PMet}(T_\Sigma, \Omega)$ . We show  $\text{Gen}(\Delta)_\Omega = \Delta$ . By the  $\Omega$ -generatedness of  $\Delta$ , we have, for all sets  $I$  and  $t, u \in T_\Sigma I$ ,

$$\text{Gen}((\Delta)_\Omega)_I(t, u) = \sup_{k: I \rightarrow T_\Sigma\Omega} \Delta_\Omega(k^\sharp(t), k^\sharp(u)) = \Delta_I(t, u).$$

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

$$\begin{aligned} U[d] \subseteq V &\iff \overline{\{\emptyset \vdash t =_\varepsilon u \mid \varepsilon \in \mathbb{Q}^+, d(t, u) \leq \varepsilon\}}^{\mathbf{QET}(\Sigma, \Omega)} \subseteq V \\ &\iff \forall t, u \in T_\Sigma\Omega, \varepsilon \in \mathbb{Q}^+. d(t, u) \leq \varepsilon \implies \emptyset \vdash t =_\varepsilon u \in V \\ (*) &\iff \forall t, u \in T_\Sigma\Omega, \varepsilon \in \mathbb{Q}^+. d(t, u) \leq \varepsilon \implies \inf \{\varepsilon' \in \mathbb{Q}^+ \mid \emptyset \vdash t =_{\varepsilon'} u \in V\} \leq \varepsilon \\ &\iff \forall t, u \in T_\Sigma\Omega. \inf \{\varepsilon' \in \mathbb{Q}^+ \mid \emptyset \vdash t =_{\varepsilon'} u \in V\} \leq d(t, u) \\ &\iff d \leq D[V] \end{aligned}$$

The step (\*) uses Lemma 47.

Let  $d \in \mathbf{CSPMet}(T_\Sigma, \Omega)$ . We finally show  $D[U[d]] = d$ . By the adjunction  $U[-] \dashv D[-]$ , we have  $d \leq D[U[d]]$ . We thus show  $D[U[d]] \leq d$ . We unfold this goal:

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

Since  $U[d]$  is the least QET including  $\{\emptyset \vdash t =_\varepsilon u \mid d(t, u) \leq \varepsilon\}$ , it suffices to have a QET  $V \in \mathbf{QET}(\Sigma, \Omega)$  such that

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

Inspired from the definition of models of QET (Bacci et al., 2021), we define  $V$  as follows:

$$\Gamma \vdash t =_\varepsilon u \in V$$

$$\iff \forall \sigma : \Omega \rightarrow T_{\Sigma} \Omega . ((\forall t' =_{\varepsilon'} u' \in \Gamma . d(\sigma^{\sharp}(t'), \sigma^{\sharp}(u')) \leq \varepsilon') \implies d(\sigma^{\sharp}(t), \sigma^{\sharp}(u)) \leq \varepsilon) .$$

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

$$\emptyset \vdash t =_{\varepsilon} u \in V \iff (\forall \sigma : \Omega \rightarrow T_{\Sigma} \Omega . d(\sigma^{\sharp}(t), \sigma^{\sharp}(u)) \leq \varepsilon) \iff d(t, u) \leq \varepsilon .$$

We check that  $V$  satisfies all rules of QET:

(Ref) From the reflexivity of  $d$ , we have  $\emptyset \vdash t =_0 t \in V$ :

$$\forall \sigma : \Omega \rightarrow T_{\Sigma} \Omega . (\top \implies d(\sigma^{\sharp}(t), \sigma^{\sharp}(t)) \leq 0) .$$

(Sym) From the symmetry of  $d$ , we have  $\{t =_{\varepsilon} u\} \vdash u =_{\varepsilon} t \in V$ :

$$\forall \sigma : \Omega \rightarrow T_{\Sigma} \Omega . ((d(\sigma^{\sharp}(t), \sigma^{\sharp}(u)) \leq \varepsilon) \implies d(\sigma^{\sharp}(u), \sigma^{\sharp}(t)) \leq \varepsilon) .$$

(Tri) From the triangle inequality of  $d$ , we have  $\{t =_{\varepsilon} u, u =_{\varepsilon'} v\} \vdash t =_{\varepsilon + \varepsilon'} v \in V$ :

$$\forall \sigma : \Omega \rightarrow T_{\Sigma} \Omega . ((d(\sigma^{\sharp}(t), \sigma^{\sharp}(u)) \leq \varepsilon \wedge d(\sigma^{\sharp}(u), \sigma^{\sharp}(v)) \leq \varepsilon') \implies d(\sigma^{\sharp}(t), \sigma^{\sharp}(v)) \leq \varepsilon + \varepsilon') .$$

(Max) Since  $(\mathbb{Q}^+, \leq, 0, +)$  is a preordered monoid where the unit  $0$  is the least element, we have  $\{t =_{\varepsilon} u\} \vdash t =_{\varepsilon + \varepsilon'} u \in V$ :

$$\forall \sigma : \Omega \rightarrow T_{\Sigma} \Omega . ((d(\sigma^{\sharp}(t), \sigma^{\sharp}(u)) \leq \varepsilon + 0) \implies d(\sigma^{\sharp}(t), \sigma^{\sharp}(u)) \leq \varepsilon + \varepsilon') .$$

(Arch) From the density of  $\mathbb{Q}^+$ , we have

$$\forall \sigma : \Omega \rightarrow T_{\Sigma} \Omega . (d(\sigma^{\sharp}(t), \sigma^{\sharp}(u)) > \varepsilon \implies (\exists \varepsilon' \in \mathbb{Q}^+ \text{ s.t. } \varepsilon < \varepsilon' . d(\sigma^{\sharp}(t), \sigma^{\sharp}(u)) > \varepsilon')) .$$

It is equivalent to  $\{t =_{\varepsilon'} u \mid \varepsilon < \varepsilon'\} \vdash t =_{\varepsilon} u \in V$ :

$$\forall \sigma : \Omega \rightarrow T_{\Sigma} \Omega . ((\forall \varepsilon' \in \mathbb{Q}^+ \text{ s.t. } \varepsilon < \varepsilon' . d(\sigma^{\sharp}(t), \sigma^{\sharp}(u)) \leq \varepsilon') \implies d(\sigma^{\sharp}(t), \sigma^{\sharp}(u)) \leq \varepsilon) .$$

(Nonexp) Let  $f : n \in \Sigma$  be a function and  $\{t_i =_{\varepsilon} u_i \mid 1 \leq i \leq n\} \subseteq \mathbb{V}(T_{\Sigma} \Omega)$  be a set of quantitative equations. Let  $I = \{1, \dots, n\}$ . We then take  $t_f = f(1, \dots, n) \in T_{\Sigma} I$ . We define  $t, s : I \rightarrow T_{\Sigma} \Omega$  by  $t(i) = t_i$  and  $s(i) = s_i$  ( $i \in I$ ). We now fix an arbitrary  $\sigma : \Omega \rightarrow T_{\Sigma} \Omega$ , and assume  $d(\sigma^{\sharp}(t_i), \sigma^{\sharp}(s_i)) \leq \varepsilon$  for all  $i \in I$ . Then this asserts  $\sup_{i \in I} d(\sigma^{\sharp}(t(i)), \sigma^{\sharp}(s(i))) \leq \varepsilon$ . From the congruence of  $d$ , we obtain

$$\begin{aligned} d(\sigma^{\sharp}(f(t_1, \dots, t_n)), \sigma^{\sharp}(f(s_1, \dots, s_n))) &= d(\sigma^{\sharp}(t^{\sharp}(t_f)), \sigma^{\sharp}(s^{\sharp}(t_f))) \\ &\leq \sup_{i \in I} d(\sigma^{\sharp}(t(i)), \sigma^{\sharp}(s(i))) \leq \varepsilon . \end{aligned}$$

Since  $\sigma$  is arbitrary, we conclude  $\{t_i =_{\varepsilon} u_i \mid 1 \leq i \leq n\} \vdash f(t_1, \dots, t_n) =_{\varepsilon} f(s_1, \dots, s_n) \in V$ .

(Subst) Immediate by definition of  $V$ :

$$\Gamma \vdash t =_{\varepsilon} u \in V$$

$$\iff \forall \sigma : \Omega \rightarrow T_{\Sigma} \Omega . ((\forall t' =_{\varepsilon'} u' \in \Gamma . d(\sigma^{\sharp}(t'), \sigma^{\sharp}(u')) \leq \varepsilon') \implies d(\sigma^{\sharp}(t), \sigma^{\sharp}(u)) \leq \varepsilon)$$

$$\implies \forall \sigma', \sigma : \Omega \rightarrow T_{\Sigma} \Omega . \left( \begin{aligned} &(\forall t' =_{\varepsilon'} u' \in \Gamma . d((\sigma^{\sharp} \circ \sigma')^{\sharp}(t'), (\sigma^{\sharp} \circ \sigma')^{\sharp}(u')) \leq \varepsilon') \\ &\implies d((\sigma^{\sharp} \circ \sigma')^{\sharp}(t), (\sigma^{\sharp} \circ \sigma')^{\sharp}(u)) \leq \varepsilon \end{aligned} \right)$$

$$\iff \forall \sigma' : \Omega \rightarrow T_{\Sigma} \Omega . \forall \sigma : \Omega \rightarrow T_{\Sigma} \Omega . \left( \begin{aligned} &(\forall t' =_{\varepsilon'} u' \in \Gamma . d(\sigma^{\sharp}(\sigma'^{\sharp}(t')), \sigma^{\sharp}(\sigma'^{\sharp}(u')))) \leq \varepsilon') \\ &\implies d(\sigma^{\sharp}(\sigma'^{\sharp}(t)), \sigma^{\sharp}(\sigma'^{\sharp}(u))) \leq \varepsilon \end{aligned} \right)$$

$$\implies \forall \sigma' : \Omega \rightarrow T_{\Sigma} \Omega . \forall \sigma : \Omega \rightarrow T_{\Sigma} \Omega . \left( \begin{aligned} &(\forall t'' =_{\varepsilon'} u'' \in \sigma'(\Gamma) . d(\sigma^{\sharp}(t''), \sigma^{\sharp}(u'')) \leq \varepsilon') \\ &\implies d(\sigma^{\sharp}(\sigma'^{\sharp}(t)), \sigma^{\sharp}(\sigma'^{\sharp}(u))) \leq \varepsilon \end{aligned} \right)$$

$$\iff \forall \sigma' : \Omega \rightarrow T_{\Sigma} \Omega . \sigma'(\Gamma) \vdash \sigma'^{\sharp}(t) =_{\varepsilon} \sigma'^{\sharp}(u) \in V .$$

(Cut) Assume  $\Gamma' \vdash t =_\varepsilon u$  and  $\Gamma \vdash \psi \in V$  holds for all  $\psi \in \Gamma'$ . Fix an arbitrary  $\sigma : \Omega \rightarrow T_\Sigma \Omega$ , and assume  $(\forall t' =_{\varepsilon''} u'' \in \Gamma . d(\sigma^\sharp(t'), \sigma^\sharp(u'')) \leq \varepsilon'')$ . By definition of  $V$ , for any  $t' =_{\varepsilon'} u' \in \Gamma'$ , we obtain  $d(\sigma^\sharp(t'), \sigma^\sharp(u')) \leq \varepsilon'$  from  $\Gamma \vdash t' =_{\varepsilon'} u' \in V$ . Hence, by definition of  $V$  again, we obtain  $d(\sigma^\sharp(t), \sigma^\sharp(u)) \leq \varepsilon$  from  $\Gamma \vdash t =_\varepsilon u \in V$ . Since  $\sigma$  is arbitrary, we conclude  $\Gamma \vdash t =_\varepsilon u \in V$ .

(Assumpt) Assume  $t =_\varepsilon u \in \Gamma$ . Fix an arbitrary  $\sigma : \Omega \rightarrow T_\Sigma \Omega$ . Regardless of the value of  $d(\sigma^\sharp(t), \sigma^\sharp(u))$ , the following predicare is true:

$$(\forall t' =_{\varepsilon'} u' \in \Gamma . d(\sigma^\sharp(t'), \sigma^\sharp(u')) \leq \varepsilon') \implies d(\sigma^\sharp(t), \sigma^\sharp(u)) \leq \varepsilon$$

because the premise  $(\forall t' =_{\varepsilon'} u' \in \Gamma . d(\sigma^\sharp(t'), \sigma^\sharp(u')) \leq \varepsilon')$  is false whenever  $d(\sigma^\sharp(t), \sigma^\sharp(u)) > \varepsilon$ . Since  $\sigma$  is arbitrary, we conclude  $\Gamma \vdash t =_\varepsilon u \in V$ . □

*Proof.* (Proof of Theorem 36) It is clear that  $U[d]$  is an unconditional QET from its definition. Therefore, we take arbitrary  $V \in \mathbf{UQET}(\Sigma, \Omega)$  and show  $U[D[V]] = V$ . We assume  $V = \overline{S}^{\mathbf{QET}(\Sigma, \Omega)}$  for some  $S \subseteq \{\emptyset \vdash t =_\varepsilon u \mid t, u \in T_\Sigma \Omega, \varepsilon \in \mathbb{Q}^+\}$ . The adjunction  $U[-] \dashv D[-]$  implies  $U[D[V]] \subseteq V$ . We thus show  $V \subseteq U[D[V]]$ . For any  $t, u \in T_\Sigma \Omega$  and  $\varepsilon \in \mathbb{Q}^+$ , we have

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

Here, (\*) uses Lemma 47. From the monotonicity of the closure  $\overline{(-)}^{\mathbf{QET}(\Sigma, \Omega)}$ , we conclude

$$V = \overline{S}^{\mathbf{QET}(\Sigma, \Omega)} \subseteq \overline{\{\emptyset \vdash t =_\varepsilon u \mid D[V](t, u) \leq \varepsilon\}}^{\mathbf{QET}(\Sigma, \Omega)} = U[D[V]].$$

□

**Appendix D. Proofs for Section 7 (Graded Strong Relational Liftings for Divergences)**

**Lemma 50.** *Let  $(\mathbb{C}, T)$  be a CC-SM and  $\Delta = \{\Delta_I^m : (U(TI))^2 \rightarrow \mathcal{Q}\}_{m \in M, I \in \mathbb{C}}$  be a doubly indexed family of  $\mathcal{Q}$ -divergences on  $TI$  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 37).*

*Proof.* (Condition 1) We first show that  $(\text{id}_{TX_1}, \text{id}_{TX_2}) \in \mathbf{BRel}(\mathbb{C})(T^{[\Delta]}(m, v)X, T^{[\Delta]}(n, w)X)$  for all  $X$  whenever  $m \leq n$  and  $v \leq w$ . From the monotonicity of  $\Delta$ , for all  $I \in \mathbb{C}$ ,  $c'_1, c'_2 \in U(TI)$ ,  $n' \in M, w' \in \mathcal{Q}$ , we have

$$\begin{aligned} (c'_1, c'_2) \in \tilde{\Delta}(m \cdot n', v + w')I & \iff \Delta_I^{m \cdot n'}(c'_1, c'_2) \leq v + w' \implies \Delta_I^{n \cdot n'}(c'_1, c'_2) \leq v + w' \implies \Delta_I^{n \cdot n'}(c'_1, c'_2) \leq w + w' \\ & \iff (c'_1, c'_2) \in \tilde{\Delta}(n \cdot n', w + w')I. \end{aligned}$$

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:

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

(Condition 2) We next show  $(\eta_{X_1}, \eta_{X_2}) : X \dot{\rightarrow} T^{[\Delta]}(1, 0)X$ . 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:

$$(x_1, x_2) \in X$$



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

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

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

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

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

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

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

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  $(k_1, k_2) : Y \rightarrow \tilde{\Delta}(n'', w'')I$  are arbitrary, we conclude  $(f_1^\sharp, f_2^\sharp) : T^{[\Delta]}(n, w)X \rightarrow T\Delta(n \cdot m, w + v)$  as follows:

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

This completes the proof. □

*Proof.* (Proof of Proposition 40) Since  $\dot{T}$  lifts the Kleisli extension (Condition 3 of Definition 37), and satisfy the fundamental property, we obtain

$$\begin{aligned} (c_1, c_2) &\in \dot{T}(m, v)X \\ \implies \forall I \in \mathbb{C}, n \in M, q \in \mathcal{Q}, (k_1, k_2): X &\dot{\rightarrow} \dot{T}(n, w)(EI) \cdot (k_1^\# c_1, k_2^\# c_2) \in \dot{T}(mn, v + w)(EI) \\ \iff \forall I \in \mathbb{C}, n \in M, q \in \mathcal{Q}, (k_1, k_2): X &\dot{\rightarrow} \tilde{\Delta}(n, w)I \cdot (k_1^\# c_1, k_2^\# c_2) \in \tilde{\Delta}(mn, v + w)I \\ \iff (c_1, c_2) &\in T^{[\Delta]}(m, v)X. \end{aligned}$$

□

### Appendix E. Proofs for Section 9 (Case Study I: Higher-Order Probabilistic Programs)

**Lemma 51.** *The mapping*

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

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

*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  $\mathbb{R} \times \mathbb{R}_{\neq 0} \rightarrow [0, 1]$  where  $\mathbb{R}_{\neq 0}$  is the subspace of  $\mathbb{R}$  whose underlying set is  $\{r \in \mathbb{R} \mid r \neq 0\}$ . 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  $\mathbb{R} \times \mathbb{R}_{\neq 0} \times \mathbb{R} \rightarrow \mathbb{R}$ ; 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  $f_{A \cap [k, k + 1]}: \mathbb{R} \times \mathbb{R}_{\neq 0} \rightarrow [0, 1]$  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  $f_A(x, \sigma^2) = \text{ev}_A \circ \mathcal{N}(x, \sigma^2)$ , the mapping  $g(x, \sigma) \triangleq \mathcal{N}(x, \sigma^2)$  forms a measurable function of type  $\mathbb{R} \times \mathbb{R}_{\neq 0} \rightarrow G\mathbb{R}$ . Next, it is obvious that the mapping  $g'(x, \sigma) \triangleq \mathbf{d}_x$  forms a measurable function of type  $\mathbb{R} \times \{0\} \rightarrow G\mathbb{R}$ . Finally, the following mapping  $g''$  forms a measurable function of type  $\mathbb{R} \times \mathbb{R} \rightarrow (\mathbb{R} \times \mathbb{R}_{\neq 0} + \mathbb{R} \times \{0\})$ :

$$g''(x, \sigma) \triangleq \begin{cases} \iota_1(x, \sigma) & \sigma \neq 0 \\ \iota_2(x, 0) & \sigma = 0 \end{cases}.$$

Let  $A \in \Sigma_{\mathbb{R} \times \mathbb{R}_{\neq 0} + \mathbb{R} \times \{0\}}$ . By definition of  $g''$ , we have  $(g'')^{-1}(A) = \iota_1^{-1}(A) \cup \iota_2^{-1}(A)$ . Since the coprojections are measurable, we have  $\iota_1^{-1}(A) \in \Sigma_{\mathbb{R} \times \mathbb{R}_{\neq 0}}$  and  $\iota_2^{-1}(A) \in \Sigma_{\mathbb{R} \times \{0\}}$ . Since  $\mathbb{R}_{\neq 0}$  and  $\{0\}$  are measurable subsets of  $\mathbb{R}$ , we have  $\Sigma_{\mathbb{R} \times \mathbb{R}_{\neq 0}}, \Sigma_{\mathbb{R} \times \{0\}} \subseteq \Sigma_{\mathbb{R} \times \mathbb{R}}$ . Thus,  $(g'')^{-1}(A) \in \Sigma_{\mathbb{R} \times \mathbb{R}}$ . Since  $A$  is arbitrary,  $g''$  is measurable. Therefore, we conclude the measurability of the composition  $[g, g'] \circ g'': \mathbb{R} \times \mathbb{R} \rightarrow G\mathbb{R}$ , which is exactly the mapping in the statement. □

**Corollary 52.**  $\llbracket \text{norm} \rrbracket \in \text{QBS}(K\mathbb{R} \times K\mathbb{R}, PK\mathbb{R})$ .

**Lemma 53.** *The mapping*

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

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

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

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

The density function  $h(x, \lambda, r) = \frac{1}{2\lambda} \exp\left(-\frac{|x-r|}{\lambda}\right)$  is continuous function of type  $\mathbb{R} \times \mathbb{R}_{0\leq} \times \mathbb{R} \rightarrow \mathbb{R}$  where  $\mathbb{R}_{0\leq}$  is the subspace of  $\mathbb{R}$  whose underlying set is  $\{r \in \mathbb{R} \mid 0 \leq r\}$ . The measurability of  $\text{Lap}(x, \lambda)$  is proved in the same way as  $\mathcal{N}(x, \sigma^2)$ . The rest of proof is the same routine as Lemma 51. □

**Corollary 54.**  $\llbracket \text{lap} \rrbracket \in \text{QBS}(K\mathbb{R} \times K\mathbb{R}, PK\mathbb{R})$ .

---

**Cite this article:** Sato T and Katsumata S (2023). Divergences on monads for relational program logics. *Mathematical Structures in Computer Science* 33, 427–485. <https://doi.org/10.1017/S0960129523000245>