# Approximate Span Liftings: Compositional Semantics for Relaxations of Differential Privacy

Tetsuya Sato*, Gilles Barthe†, Marco Gaboardi‡, Justin Hsu§ and Shin-ya Katsumata¶
* Seikei University † MPI for Security and Privacy and IMDEA Software Institute
‡ University at Buffalo § University of Wisconsin–Madison ¶ National Institute of Informatics

***Abstract*—We develop new abstractions for reasoning about three relaxations of differential privacy: *Rényi differential privacy*, *zero-concentrated differential privacy*, and *truncated concentrated differential privacy*, which express bounds on statistical divergences between two output probability distributions. In order to reason about such properties compositionally, we introduce *approximate span-lifting*, a novel construction extending the approximate relational lifting approaches previously developed for standard differential privacy to a more general class of divergences, and also to continuous distributions. As an application, we develop a program logic based on approximate span-liftings capable of proving relaxations of differential privacy and other statistical divergence properties.**

## I. Introduction

Differential privacy [1] is a strong, statistical notion of data privacy that has attracted the attention of theoreticians and practitioners alike. One reason for its success is that differential privacy can usually be proved *compositionally*, enabling easy construction of new private algorithms and making formal verification practical. By now, researchers have developed a wide variety of programming languages and program analysis tools to prove differential privacy [2], [3], [4], [5], [6], [7], [8], [9] ([10] provide a recent survey).

Seeking more refined composition properties, researchers have recently proposed new relaxations of differential privacy: *Rényi differential privacy* (RDP) [11], *zero-concentrated differential privacy* (zCDP) [12], and *truncated concentrated differential privacy* (tCDP) [13]. Roughly speaking, standard differential privacy requires a bound on the magnitude of a random variable measuring the privacy loss, while RDP, zCDP, and tCDP model finer bounds on the *moments* of this random variable. (Recall that the first moment of a random variable is its average value, and the second moment of a random variable is its variance.) These relaxations capture fine-grained aspects of the privacy loss, enabling more precise privacy analyses and allowing algorithms to add less random noise to achieve the same privacy level.

RDP, zCDP, and tCDP are all defined in terms of *Rényi divergences* [14], distances on distributions originating from information theory. Inspiring our work, Barthe and Olmedo previously developed abstractions for reasoning about a family of divergences called $f$*-divergences* as part of their work on the program logic $f$pRHL [15], [16]. In particular, the semantic foundation of $f$pRHL is a *2-witness relational lifting* for $f$-divergences, which tracks the $f$-divergence between related pairs of distributions. However, their framework is not sufficient to establish our target properties for two reasons. First, Rényi divergences are not $f$-divergences (for one differenc, $f$-divergences are jointly convex while Rényi divergences are only quasi-convex [17]), moreover, zCDP and tCDP are *supremums* of *Rényi divergences*. As a result, these properties cannot be described in terms of $f$-divergences, nor captured in $f$pRHL. We develop new relational liftings supporting significantly more general divergences, allowing direct reasoning about RDP, zCDP, and tCDP.

Second, the 2-witness relational lifting approach has only been proposed for discrete distributions, while many algorithms satisfying relaxations of differential privacy—indeed, the motivating examples—sample from continuous distributions, such as the Gaussian distribution. Handling these distributions requires a careful treatment of measure theory. Previous work [18] has considered a different semantic model for standard differential privacy over continuous distributions using *witness-free* relational lifting, but it is not clear how to extend this model beyond differential privacy.

To overcome these challenges, we generalize 2-witness liftings in two directions. First, we replace the notion of $f$-divergence with a more general class of divergences, identifying the basic properties needed for compositional reasoning. Second, we generalize to continuous probability measures. The main challenge is establishing a sequential composition principle—the continuous case introduces measurability requirements for composition. Accordingly, we extend the structure of 2-witness liftings to a new notion called *approximate span-liftings*, which have the necessary data to ensure closure under sequential composition. Finally, we instantiate our general model with divergences for RDP, zCDP, and tCDP, establishing categorical properties needed to build approximate span-liftings. As an extended application, we develop a relational program logic that can verify differential privacy, RDP, zCDP, and tCDP within a single logic for programs using discrete or continuous sampling, and interpret the logic via approximate span-liftings.

After motivating the various relaxations of differential pri-

Work done while Tetsuya Sato was at the University at Buffalo, SUNY. This work was partially supported by the NSF under grant #1565365 and #1718220, and a Facebook TAV award. Katsumata was supported by JSPS KAKENHI (Grant-in-Aid for Scientific Research (C)) Grant Number JP15K00014 and JST ERATO HASUO Metamathematics for Systems Design Project (No. JPMJER1603).

978-1-7281-3608-0/19/$31.00 

vacy, summarizing the key technical challenges (Section II), and introducing mathematical preliminaries (Section III), we present our main contributions.

- We identify a general class of divergences supporting basic properties composition properties, and we show that our class can model RDP, zCDP and tCDP (Section IV).
- We extend 2-witness relational liftings to the continuous case by introducing a novel notion of approximate span-lifting. We show how to translate composition properties of specific divergences to their corresponding approximate span-liftings (Section V).
- We develop a program logic supporting four flavors of differential privacy—standard DP, RDP, zCDP, and tCDP—where programs may use both discrete and continuous random sampling, and show soundness (Section VI). We demonstrate our logic on three examples (Section VII).

We survey related work (Section VIII) and then conclude with promising future directions (Section IX).

## II. Motivation and Technical Challenges

For simplicity, in this section we consider probability distributions that have associated density functions.

### A. Differential Privacy and its Relaxations

A *randomized algorithm* is a measurable function $\mathcal{A}\colon X \to \mathrm{Prob}(Y)$ from a set $X$ of inputs to the set $\mathrm{Prob}(Y)$ of *probability distributions* on a set $Y$ of outputs.

**Definition 1** (Differential Privacy (DP) [1])**.** *A randomized algorithm $\mathcal{A}\colon X \to \mathrm{Prob}(Y)$ is $(\varepsilon,\delta)$-*differentially private *w.r.t an adjacency relation $\Phi \subseteq X \times X$ if for any pairs of inputs $(x,x') \in \Phi$, and any measurable subset $S \subseteq Y$, we have $\Pr[\mathcal{A}(x) \in S] \leq e^{\varepsilon}\Pr[\mathcal{A}(x') \in S] + \delta$.*

**Definition 2** (Rényi divergence [14])**.** *Let $\alpha > 1$. The* Rényi divergence *of order $\alpha$ between two probability distributions $\mu_1$ and $\mu_2$ on a measurable space $X$ is defined by:*

$$D_X^{\alpha}(\mu_1||\mu_2) \stackrel{\mathsf{def}}{=} \frac{1}{\alpha-1}\log\int_X \mu_2(x)\left(\frac{\mu_1(x)}{\mu_2(x)}\right)^{\alpha} dx. \quad (1)$$

**Definition 3** (Rényi Differential Privacy (RDP) [11])**.** *A randomized algorithm $\mathcal{A} : X \to \mathrm{Prob}(Y)$ is $(\alpha,\rho)$-*Rényi differentially private *w.r.t an adjacency relation $\Phi \subseteq X \times X$ if for all $(x,x') \in \Phi$, we have $D_X^{\alpha}(\mathcal{A}(x)||\mathcal{A}(x')) \leq \rho$.*

**Definition 4** (zero-Concentrated Differential Privacy (zCDP) [12])**.** *A randomized algorithm $\mathcal{A} : X \to \mathrm{Prob}(Y)$ is $(\xi,\rho)$-*zero concentrated differentially private *w.r.t an adjacency relation $\Phi \subseteq X \times X$ if for all $(x,x') \in \Phi$, we have*

$$\forall \alpha > 1.\ D_Y^{\alpha}(\mathcal{A}(x)||\mathcal{A}(x')) \leq \xi + \alpha\rho. \quad (2)$$

**Definition 5** (Truncated Concentrated Differential Privacy (tCDP) [13])**.** *A randomized algorithm $\mathcal{A} : X \to \mathrm{Prob}(Y)$ is $(\rho,\omega)$-*truncated concentrated differentially private *w.r.t an adjacency relation $\Phi \subseteq X \times X$ if for all $(x,x') \in \Phi$, we have*

$$\forall 1 < \alpha < \omega.\ D_Y^{\alpha}(\mathcal{A}(x)||\mathcal{A}(x')) \leq \alpha\rho. \quad (3)$$

While these notions may appear cryptic at first sight, they can all be understood as bounds on the *privacy loss* of a randomized algorithm:

$$\mathcal{L}^{x\to x'}(y) = \frac{\Pr[\mathcal{A}(x) = y]}{\Pr[\mathcal{A}(x') = y]}$$

where $x, x'$ are two inputs. Intuitively, the privacy loss measures how much information is revealed when the output of a private algorithm is seen to be $y$. While output values with a large privacy loss are highly revealing—they are far more likely to result from input $x$ rather than a different input $x'$—if these outputs are only seen with very small probability, then their overall influence can be discounted. The different privacy definitions bound different aspects of the privacy loss random variable, when $y$ is drawn from the output of the algorithm. The following table summarizes these bounds.

| Privacy | Bound on privacy loss $\mathcal{L}$ |
|---|---|
| $(\varepsilon,\delta)$-DP | $\Pr_{y\sim\mathcal{A}(x)}[\mathcal{L}^{x\to x'}(y) \leq e^{\varepsilon}] \geq 1-\delta$ |
| $(\alpha,\rho)$-RDP | $\mathbb{E}_{y\sim\mathcal{A}(x)}[\mathcal{L}^{x\to x'}(y)^{\alpha}] \leq e^{(\alpha-1)\rho}$ |
| $(\xi,\rho)$-zCDP | $\forall\alpha \in (1,\infty).\ \mathbb{E}_{y\sim\mathcal{A}(x)}[\mathcal{L}^{x\to x'}(y)^{\alpha}] \leq e^{(\alpha-1)(\xi+\alpha\rho)}$ |
| $(\omega,\rho)$-tCDP | $\forall\alpha \in (1,\omega).\ \mathbb{E}_{y\sim\mathcal{A}(x)}[\mathcal{L}^{x\to x'}(y)^{\alpha}] \leq e^{(\alpha-1)\alpha\rho}$ |

In particular, DP bounds the maximum value of the privacy loss,[1] $(\alpha,\cdot)$-RDP bounds the $\alpha$-moment, zCDP bounds all moments, and $(\cdot,\omega)$-tCDP bounds the moments up to some cutoff $\omega$. Many conversions are known between these definitions; for instance, the relaxations RDP, zCDP, and tCDP are known to sit between $(\varepsilon,0)$ and $(\varepsilon,\delta)$-differential privacy, up to some modification in the parameters. While this means that RDP, zCDP, and tCDP can sometimes be analyzed by reduction to standard differential privacy, converting between the different notions requires weakening the parameters and often the privacy analysis is simpler or more precise when working with RDP, zCDP, or tCDP directly. The interested reader can refer to the original papers [12], [11].

Two motivating examples fitting for these definitions are the *Gaussian mechanism* and *Sinh Normal mechanism*, which add noise according to the Gaussian distribution and the sinh-normal distribution respectively.

### B. 2-witness liftings for $f$-divergences in the discrete case

Barthe and Olmedo [15] observed that standard differential privacy can be phrased in terms of a more general class of divergences, called *$f$-divergences*.

**Definition 6.** *A* weight function *is a convex function $f\colon \mathbb{R}_{\geq 0} \to \mathbb{R}$ continuous at 0.*[2]

[1]Technically, this interpretation of DP only holds for mechanisms with a well-behaved privacy loss. We mention it here because this is often used informally to compare DP and its relaxations but there are mechanisms for which this interpretation is problematic [19].

[2]As is conventional [20], we exclude the condition $f(1) = 0$ from the definition of weight function to support the exponential of Rényi divergence of order $\alpha$. We also assume $0f(a/0) = \lim_{t\to 0+} tf(a/t)$ for $a > 0$ and $0f(0/0) = 0$.

**Definition 7** ($f$-divergence)**.** *For a weight function $f$, the $f$-divergence $\Delta^f$ between two distributions $\mu_1, \mu_2$ over a measurable space $X$ is defined as*

$$\Delta_X^f(\mu_1, \mu_2) = \int_X \mu_2(x) f\left(\frac{\mu_1(x)}{\mu_2(x)}\right)\, dx. \tag{4}$$

In particular, the $f$-divergence $\Delta^{\mathtt{DP}(\varepsilon)}$ with weight function $\mathtt{DP}(\varepsilon)(t) = \max(0, 1 - e^{\varepsilon} t)$ models differential privacy [15], [16]. For any randomized algorithm $\mathcal{A} : X \to \mathrm{Prob}(Y)$ and adjacency relation $\Phi \subseteq X \times X$, we have the equivalence

$$\mathcal{A} \text{ is } (\varepsilon, \delta)\text{-DP} \iff \forall (x, x') \in \Phi. \Delta_Y^{\mathtt{DP}(\varepsilon)}(\mathcal{A}(x), \mathcal{A}(x')) \leq \delta.$$

To support their logic $f$pRHL, Barthe and Olmedo introduced a *2-witness relational lifting* for $f$-divergences as a key abstraction to reason about $f$-divergence properties of probabilistic programs. This construction lifts a relation $R \subseteq X \times Y$ over discrete sets $X, Y$ to a relation $R^{\sharp(f,\delta)} \subseteq \mathrm{Dist}(X) \times \mathrm{Dist}(Y)$ over subprobability distributions:

$$R^{\sharp(f,\delta)} = \left\{ (\mu_1, \mu_2) \;\middle|\; \begin{array}{l} \exists \mu_l, \mu_r \in \mathrm{Dist}(R).\ \pi_1(\mu_l) = \mu_1, \\ \pi_2(\mu_r) = \mu_2,\ \Delta_R^f(\mu_L, \mu_R) \leq \delta \end{array} \right\}. \tag{5}$$

Above, $\pi_i(\mu)$ is the $i$-th marginal of $\mu$, that is, $(\pi_1(\mu))(x) = \sum_{y \in Y} \mu(x, y)$ and $(\pi_2(\mu))(y) = \sum_{x \in X} \mu(x, y)$. The distributions $\mu_L$ and $\mu_R$ are called *witness distributions*, since to show that two distributions are related by a lifting, one must show the existence of two appropriate witnesses. These liftings have several attractive features. The equality lifting takes the form

$$\mathrm{Eq}_X^{\sharp(f,\delta)} = \left\{ (\mu_1, \mu_2) \;\middle|\; \Delta_X^f(\mu_1, \mu_2) \leq \delta \right\},$$

thus characterizing differential privacy: a program $\mathcal{A} \colon X \to \mathrm{Dist}(Y)$ is $(\varepsilon, \delta)$-differentially private w.r.t. an adjacency relation $\Phi$, if $(\mathcal{A}(x), \mathcal{A}(x')) \in \mathrm{Eq}_Y^{\sharp(\mathtt{DP}(\varepsilon),\delta)}$ for every $(x, x') \in \Phi$. Second, 2-witness liftings satisfy various composition properties, enabling clean verification of probabilistic programs. However, this construction works only in the discrete case and the logic $f$pRHL cannot reason about programs that sample from continuous distributions, like the Gaussian distribution.

### C. Challenge 1: Handling Richer Divergences

Much like how standard differential privacy can be viewed in terms of $f$-divergences, we would like to view RDP, zCDP, and tCDP as bounds on more general divergences. A natural candidate for Rényi differential privacy is Rényi divergence $D^{\alpha}$, as in its original definition. Indeed, we have:

$$\mathcal{A} \text{ is } (\alpha, \rho)\text{-RDP iff } \forall (x, x') \in \Phi,\ D_Y^{\alpha}(\mathcal{A}(x) || \mathcal{A}(x')) \leq \rho.$$

However, the Rényi divergence $D^{\alpha}(\mu_1 || \mu_2)$ of order $\alpha$ is not an $f$-divergence, and so it does not fit in the 2-witness lifting framework. Likewise, zCDP [12] and tCDP [13] can be defined via uniform bounds on families of Rényi divergence:

$$\Delta_X^{\mathtt{zCDP}(\xi)}(\mu_1, \mu_2) = \sup_{1<\alpha} \frac{1}{\alpha} \left(D_X^{\alpha}(\mu_1 || \mu_2) - \xi\right)\ (0 \leq \xi), \tag{6}$$

$$\Delta_X^{\mathtt{tCDP}(\omega)}(\mu_1, \mu_2) = \sup_{1<\alpha<\omega} \frac{1}{\alpha} \left(D_X^{\alpha}(\mu_1 || \mu_2)\right)\ (1 < \omega), \tag{7}$$

letting us reformulate zCDP and tCDP as

$$\mathcal{A} \text{ is } (\xi, \rho)\text{-zCDP iff } \forall (x, x') \in \Phi. \Delta_Y^{\mathtt{zCDP}(\xi)}(\mathcal{A}(x), \mathcal{A}(x')) \leq \rho.$$

$$\mathcal{A} \text{ is } (\rho, \omega)\text{-tCDP iff } \forall (x, x') \in \Phi. \Delta_Y^{\mathtt{tCDP}(\omega)}(\mathcal{A}(x), \mathcal{A}(x')) \leq \rho.$$

These divergences are also not $f$-divergences. Furthermore, the RDP, zCDP and tCDP divergences may take negative values when applied to sub-probability distributions, which can arise from probabilistic computations that may not terminate with probability 1. Accordingly, we generalize the notion of divergence to go beyond $f$-divergences and also to handle sub-probability distributions. Starting from families of real valued functions from pairs of distributions, we introduce basic properties needed to give good composition properties for their corresponding liftings.

### D. Challenge 2: 2-witness Liftings for the Continuous Case

In order to support natural examples for RDP, zCDP, and tCDP, we need a framework supporting continuous distributions, such as Gaussian, Laplace, and sinh-normal distributions. Unfortunately, extending 2-witness relational liftings to the continuous case presents further technical challenges related to composition. The relational lifting $(-)^{\sharp(\mathtt{DP}(\varepsilon),\delta)}$ for standard differential privacy satisfies a sequential composition principle:

$$\frac{(f, g) \colon R \dot{\to} S^{\sharp(\mathtt{DP}(\varepsilon_1),\delta_1)}}{(f^{\sharp}, g^{\sharp}) \colon R^{\sharp(\mathtt{DP}(\varepsilon_2),\delta_2)} \dot{\to} S^{\sharp(\mathtt{DP}(\varepsilon_1+\varepsilon_2),\delta_1+\delta_2)}}$$

Here, we denote by $R_1 \dot{\to} R_2$ a relation-preserving map from $R_1$ to $R_2$; $f^{\sharp}$ and $g^{\sharp}$ are the Kleisli liftings of $f$ and $g$ with respect to the monad Dist of (discrete) subprobability distributions; this composition property gives 2-witness relational liftings a *graded monad* structure [21], [22]. Since the 2-witness liftings are defined through the existence of witness distributions, for any $(d_1, d_2) \in R^{\sharp(\mathtt{DP}(\varepsilon_2),\delta_2)}$, we need witness distributions showing $(f^{\sharp}(d_1), g^{\sharp}(d_2)) \in S^{\sharp(\mathtt{DP}(\varepsilon_1+\varepsilon_2),\delta_1+\delta_2)}$. In the discrete case, these witnesses can be constructed in two steps:

1) For any $(x, y) \in R$, there exist witnesses $d_L', d_R' \in \mathrm{Dist}(S)$ proving $(f(x), g(y)) \in S^{\sharp(\mathtt{DP}(\varepsilon_1),\delta_1)}$. By applying the axiom of choice, we obtain a selection function
$$\langle l_1, l_2 \rangle \colon R \to \left\{ (d_L', d_R') \mid \Delta_S^{\mathtt{DP}(\varepsilon_1)}(d_L', d_R') \leq \delta_1 \right\}.$$
2) For any witnesses $d_L, d_R \in \mathrm{Dist}(R)$ proving $(d_1, d_2) \in R^{\sharp(\mathtt{DP}(\varepsilon_2),\delta_2)}$, $(l_1^{\sharp}(d_L), l_2^{\sharp}(d_R))$ is a pair of witness distributions proving $(f^{\sharp}(d_1), g^{\sharp}(d_2)) \in S^{\sharp(\mathtt{DP}(\varepsilon_1+\varepsilon_2),\delta_1+\delta_2)}$ by composability of $\Delta^{\mathtt{DP}(\varepsilon)}$.

The first step is problematic to extend to the continuous case because the witness-selecting functions $l_1$ and $l_2$ obtained by the axiom of choice may not be measurable—the Kleisli extensions $l_1^{\sharp}$ and $l_2^{\sharp}$ in the second step may not be well-defined in the continuous case. To resolve this difficulty, we introduce a novel notion of *approximate span-liftings*. The key idea is that morphisms between span-liftings carry a built-in measurable witness selection function, making it unnecessary to use the axiom of choice when proving sequential composition.

## III. Mathematical Preliminaries

*a) Measure Theory:* We briefly review some definitions from measure theory; readers should consult a textbook for more details [23]. Given a set $X$, a *$\sigma$-algebra* on $X$ is a collection $\Sigma$ of subsets of $X$ including the empty set, closed under complements, countable unions, and countable intersections; a *measurable space* $X$ is a set $|X|$ with a $\sigma$-algebra $\Sigma_X$, called the measurable sets. A countable set $X$ yields the *discrete* measurable space where all subsets are measurable: $\Sigma_X = 2^X$. A map $f\colon X \to Y$ between measurable spaces is *measurable* if $f^{-1}(A) \in \Sigma_X$ for all $A \in \Sigma_Y$. Any subset $S$ of measurable space $X$ forms a *subspace* where the $\sigma$-algebra is given by $\Sigma_S = \{\, A \cap S \mid A \in \Sigma_X \,\}$. $\Sigma_S$ is given as the coarsest one making the inclusion map $S \hookrightarrow X$ measurable. A *measure* on a measurable space is a map $\mu\colon \Sigma_X \to \mathbb{R}_{\geq 0} \cup \{\infty\}$ such that $\mu(\varnothing) = 0$ and $\mu(\cup_i X_i) = \sum_i \mu(X_i)$ for any countable family of disjoint measurable sets $X_i$. Measures with $\mu(X) = 1$ are called *probability measures*, and measures with $\mu(X) \leq 1$ are called *subprobability measures*. For any pair of subprobability measures $\mu_1$ on $X$ and $\mu_2$ on $Y$, the *product measure* $\mu_1 \otimes \mu_2$ of $\mu_1$ and $\mu_2$ is the unique measure on $X \times Y$ satisfying $(\mu_1 \otimes \mu_2)(A \times B) = \mu_1(A) \cdot \mu_2(B)$. For any measurable space $X$ and element $x \in X$, we write $\mathbf{d}_x$ for the Dirac measure on $X$ centered at $x$, defined as $\mathbf{d}_x(A) = 1$ if $x \in A$, and $\mathbf{d}_x(A) = 0$ otherwise. Measurable spaces and measurable functions form a category $\mathbf{Meas}$; this category has all limits and colimits, and finite products distribute over finite coproducts. We denote by $\mathbf{Fin}$ the full subcategory of $\mathbf{Meas}$ consisting of all finite discrete spaces.

*b) The Sub-Giry Monad:* The *sub-Giry monad* $\mathcal{G}$ is the subprobabilistic variant of the Giry monad [24]. In brief, $\mathcal{G}X$ is the set of subprobability measures on $X$ with suitable $\sigma$-algebra for any $X \in \mathbf{Meas}$; functor action $(\mathcal{G}f)(\mu) = \mu(f^{-1}(-))$ for $f\colon X \to Y$; unit $\eta_X(x) = \mathbf{d}_x$ for $X \in \mathbf{Meas}$ and $x \in X$; and Kleisli lifting $f^\sharp(\mu)(A) = \int_X f(x)(A)\, d\mu(x)$ for $f\colon X \to \mathcal{G}Y$ and $A \in \Sigma_Y$. The monad $\mathcal{G}$ is commutative strong with respect to binary products in $\mathbf{Meas}$; the *double strength* $\mathrm{dst}_{X,Y}\colon \mathcal{G}(X) \times \mathcal{G}(Y) \Rightarrow \mathcal{G}(X \times Y)$ is given by the product measure $\mathrm{dst}_{X,Y}(\nu_1, \nu_2) = \nu_1 \otimes \nu_2$.

*c) Graded Monads:* A *graded monad* [21], [22] is a monad refined by indices from a monoid. Let $E = (E, \cdot, 1_E, \preceq)$ be a preordered monoid. An $E$-graded monad on a category $\mathbb{C}$ consists of

- a family $\{T_e\}_{e \in E}$ of endofunctors $T_e$ on $\mathbb{C}$,
- a morphism $\eta_X\colon X \to T_{1_E} X$ for $X \in \mathbb{C}$ (unit),
- a morphism $(-)^{e_1 \sharp e_2}\colon \mathbb{C}(X, T_{e_2} Y) \to \mathbb{C}(T_{e_1} X, T_{e_1 e_2} Y)$ for $X, Y \in \mathbb{C}$ and $e_1, e_2 \in A$ (Kleisli lifting),
- a family $\{\sqsubseteq^{e_1,e_2}\}_{e_1 \preceq e_2}$ of natural transformations $\sqsubseteq^{e_1,e_2}\colon T_{e_1} \Rightarrow T_{e_2}$ (inclusion)

satisfying the following compatibility condition: for any $f\colon X \to T_{e_1} Y$ and $g\colon Y \to T_{e_2} Z$,

$$\sqsubseteq_Z^{(e_2 e_1),(e_2 e_3)} \circ f^{e_2 \sharp e_1} = (\sqsubseteq_Y^{e_1,e_2} \circ f)^{e_2 \sharp e_3},$$
$$f^{1 \sharp e_1} \circ \eta_X = f, f^{e_3 \sharp e_1} \circ \sqsubseteq_X^{e_2,e_3} = \sqsubseteq_Y^{(e_2 e_1),(e_3 e_1)} \circ f^{e_2 \sharp e_1},$$
$$\eta_X^{1_E \sharp e} = \mathrm{id}_{T_e X}, (g^{e_1 \sharp e_2} \circ f)^{e_0 \sharp e_1 e_2} = g^{e_0 e_1 \sharp e_2} \circ f^{e_0 \sharp e_1}.$$

A typical way of constructing a graded monad is by refining a plain monad with indices. An *$E$-graded lifting* of a monad $(T, \eta^T, (-)^\sharp)$ on $\mathbb{D}$ along a functor $U\colon \mathbb{C} \to \mathbb{D}$ is an $E$-graded monad $\{T_e\}_{e \in A}$ on $\mathbb{C}$ satisfying $U \circ T_e = T \circ U$, $U(f^{e_2 \sharp e_1}) = (Uf)^\sharp$, $U(\eta_D) = \eta^T_{UD}$, and $U(\sqsubseteq_D^{e_1,e_2}) = \mathrm{id}_{TUD}$. The functor $U$ erases the grading of $T_e$, yielding the original monad $T$.

*d) The Category of Spans on Measurable Spaces:* To extend the relational lifting approach to the continuous setting, we work with the category of *spans*, whose objects generalize relations by taking arbitrary functions in place of projections.

**Definition 8.** *The category* $\mathbf{Span}(\mathbf{Meas})$ *of spans in* $\mathbf{Meas}$ *consists of:*

- *Objects* $(X, Y, \Phi, \rho_1, \rho_2)$ *given by span* $X \xleftarrow{\rho_1} \Phi \xrightarrow{\rho_1} Y$ *in* $\mathbf{Meas}$.
- *Morphisms* $(X, Y, \Phi, \rho_1, \rho_2) \to (Z, W, \Psi, \rho_1', \rho_2')$ *given by triples* $(h, k, l)$ *of morphisms* $h\colon X \to Z$, $k\colon Y \to W$, *and* $l\colon \Phi \to \Psi$ *in* $\mathbf{Meas}$ *satisfying* $h \circ \rho_1 = \rho_1' \circ l$ *and* $k \circ \rho_2 = \rho_2' \circ l$.

For simplicity, we often denote a $\mathbf{Span}(\mathbf{Meas})$-object $(X, Y, \Phi, \rho_1, \rho_2)$ by $\Phi$. The category $\mathbf{Span}(\mathbf{Meas})$ has all limits, this gives us useful properties. First, the category has binary products:

$$(X, Y, \Phi, \rho_1, \rho_2) \mathbin{\dot\times} (Z, W, \Psi, \rho_1', \rho_2') = (X \times Z, Y \times W, \Phi \times \Psi, \rho_1 \times \rho_1', \rho_2 \times \rho_2').$$

We will frequently use two notions of pairing on functions. Let $f_1\colon X \to Y$, $f_2\colon X \to W$, we have $\langle f_1, f_2 \rangle\colon X \to Y \times W$ and $f_1 \times f_2\colon X \times X \to Y \times W$. As functions, $\langle f_1, f_2 \rangle$ takes a single input $x$ and returns a pair $(f_1(x), f_2(x))$ while $f_1 \times f_2$ take a pair of inputs $(x, y)$ and returns $(f_1(x), f_2(y))$. The category $\mathbf{Span}(\mathbf{Meas})$ also has coproducts:

$$(X, Y, \Phi, \rho_1, \rho_2) \mathbin{\dot+} (X', Y', \Phi', \rho_1', \rho_2') = (X + X', Y + Y', \Phi + \Phi', \rho_1 + \rho_1', \rho_2 + \rho_2').$$

Standard binary relations can be interpreted as spans. For $X, Y \in \mathbf{Meas}$, any binary relation $\Phi \subseteq |X| \times |Y|$ determines a span $X \xleftarrow{\pi_1} \Phi \xrightarrow{\pi_2} Y$ in $\mathbf{Meas}$, where $\pi_1$ and $\pi_2$ are projections and $\Phi$ is regarded as a subspace of $X \times Y$.

Finally, relation-preserving maps can be interpreted as morphisms of spans. Consider two binary relations $\Phi \subseteq |X| \times |Y|$ and $\Psi \subseteq |Z| \times |W|$, and suppose that they are interpreted as spans $(X, Y, \Phi, \pi_1, \pi_2)$ and $(Z, W, \Psi, \pi_1, \pi_2)$ as above. If $f\colon X \to Z$ and $g\colon Y \to W$ in $\mathbf{Meas}$ satisfy $(f(x), g(y)) \in \Psi$ for any $(x, y) \in \Phi$, then we have the following morphism

$$(f, g, f \times g|_\Phi)\colon (X, Y, \Phi, \pi_1, \pi_2) \to (Z, W, \Psi, \pi_1, \pi_2)$$

in $\mathbf{Span}(\mathbf{Meas})$, where $f \times g|_\Phi$ is the restriction of $f \times g$ on $\Phi$ (we often write just $f \times g$). These features are crucial to interpret probabilistic program logics (see Section VI).

## IV. General Statistical Divergences

Now that we have covered the preliminaries, our goal is to build a suitable graded monad on $\mathbf{Span}(\mathbf{Meas})$—this will be our abstraction for relational reasoning about divergences.

We proceed in two stages. In this section, we introduce a general class of *divergences*, real-valued functions on two measures over the same space. Then, we identify important composition properties inspired from analogous properties of $f$-divergences [15], [20]. We will leverage these properties to give a graded monad structure on $\mathbf{Span}(\mathbf{Meas})$ capturing these divergences in the next section.

Throughout, we write $\overline{\mathbb{R}}$ for the set $\mathbb{R} \cup \{-\infty, +\infty\}$ of extended reals. We regard both $\overline{\mathbb{R}}$ and $\mathbb{R}_{\geq 0}$ as partially ordered additive monoids. For the former one, the addition is extended by $\infty + (-\infty) = -\infty$.

**Definition 9.** *A* divergence *is a family* $\Delta = \{\Delta_X\}_{X \in \mathbf{Meas}}$ *of functions*

$$\Delta_X \colon |\mathcal{G}X| \times |\mathcal{G}X| \to \overline{\mathbb{R}}.$$

To describe composition of divergences, it is useful to work with indexed families of divergences; often, two divergences can be combined to give a new divergence with different indices. For instance, the notion of zCDP can be characterized by the family $\{\Delta^{\mathrm{zCDP}(\xi)}\}_{0 \leq \xi}$ introduced in Section II (Equation 6).

**Definition 10.** *Let* $(A, \cdot, 1_A, \preceq)$ *be a preordered monoid. An* $A$-graded family of divergences *is a family* $\boldsymbol{\Delta} = \{\Delta^\alpha\}_{\alpha \in A}$ *such that whenever* $\alpha \preceq \beta$*, for any* $X \in \mathbf{Meas}$*,*

$$\forall \mu_1, \mu_2 \in \mathcal{G}X.\ \Delta^\beta_X(\mu_1, \mu_2) \leq \Delta^\alpha_X(\mu_1, \mu_2).$$

Note that the preorder on the grading is contravariant. We will regard a divergence $\Delta$ as a singleton-graded family $\{\Delta\}$.

### A. Basic Properties of Divergences

We use several properties of graded families of divergences.

**Definition 11.** *An* $A$*-graded family* $\boldsymbol{\Delta} = \{\Delta^\alpha\}_{\alpha \in A}$ *of divergences is:*

**reflexive:** *if* $\Delta^\alpha_X(\mu, \mu) \leq 0$.

**substitutive:** *if for any* $f \colon X \to \mathcal{G}Y$*,*
$\Delta^\alpha_Y(f^\sharp \mu_1, f^\sharp \mu_2) \leq \Delta^\alpha_X(\mu_1, \mu_2)$.

**additive:** *if*
$\Delta^{\alpha \cdot \beta}_{X \times Y}(\mu_1 \otimes \mu_3, \mu_2 \otimes \mu_4) \leq \Delta^\alpha_X(\mu_1, \mu_2) + \Delta^\beta_Y(\mu_3, \mu_4)$.

**continuous:** *if*
$\Delta^\alpha_X(\mu_1, \mu_2) = \sup_{I \in \mathbf{Fin},\ k \colon X \to I} \Delta^\alpha_I(\mathcal{G}k(\mu_1), \mathcal{G}k(\mu_2))$.

**composable:** *if for any* $f, g \colon X \to \mathcal{G}Y$*,*
$\Delta^{\alpha \cdot \beta}_Y(f^\sharp \mu_1, g^\sharp \mu_2) \leq \Delta^\alpha_X(\mu_1, \mu_2) + \sup_{x \in X} \Delta^\beta_Y(f(x), g(x))$.

*All functions are assumed to be measurable.*

These properties are inspired by properties from the literature on $f$-divergences and differential privacy. For instance, substitutivity generalizes the *data-processing inequality* for $f$-divergences [25, Chapter 2], while functoriality is the special case where the data-processing function is deterministic. These two properties are also known in the privacy literature as *resilience to post-processing* [26, Proposition 2.1]. Composability corresponds to composition in differential privacy, which states that we can adaptively compose two differentially private mechanisms. Additivity corresponds to a composition where the second mechanism does not depend on the result of the first. Continuity generalizes continuity of $f$-divergences [25, Theorem 16]; divergences of continuous distributions are approximated by divergences of discrete distributions.

Reflexivity and composability are key properties to give a structure of graded monad. Intuitively, reflexivity and composability of a graded family of divergences give unit and a (graded) Kleisli lifting respectively. We also need additivity to give a *strength* of the graded monad, allowing a lifting on real-valued distributions—often available from known results in probability theory—to be converted into a lifting on distributions over larger spaces (e.g., program memories). In some ways, composability is the most important property: reflexivity is usually immediate, and additivity is a consequence.

**Theorem 1.** *An* $A$*-graded family* $\boldsymbol{\Delta}$ *is additive if it is continuous and composable.*

Although these properties have been studied before in the discrete case, there are subtleties when passing to our continuous ones. For example, in the case of discrete distributions, additivity is an instance of composability [15, Proposition 4]. In the case of continuous distributions, this may no longer hold. However, one can recover additivity from composability by using continuity.

To prove composability, it is easier to establish two other properties of families of divergences first: approximability and finite-composability.

**Definition 12.** *An* $A$*-graded family* $\boldsymbol{\Delta} = \{\Delta^\alpha\}_{\alpha \in A}$ *of divergences is:*

**approximable:** *if for any* $X \in \mathbf{Meas}$ *and* $I \in \mathbf{Fin}$*,* $f, g \colon X \to \mathcal{G}I$*, and* $\mu_1, \mu_2 \in \mathcal{G}X$*, there are* $J_n \in \mathbf{Fin}$ *and* $m^*_n \colon X \to J_n$ *and* $m_n \colon J_n \to X$ *in* $\mathbf{Meas}$ *such that*

$$\Delta^\alpha_I(f^\sharp(\mu_1), g^\sharp(\mu_2)) = \lim_{n \to \infty} \Delta^\alpha_I((f \circ m_n \circ m^*_n)^\sharp(\mu_1), (g \circ m_n \circ m^*_n)^\sharp(\mu_2)).$$

**finite-composable:** *if for any* $I, J \in \mathbf{Fin}$*,* $f, g \colon I \to \mathcal{G}J$*, and* $d_1, d_2 \in \mathcal{G}I$*,*

$$\Delta^{\alpha \cdot \beta}_J(f^\sharp d_1, g^\sharp d_2) \leq \Delta^\alpha_I(d_1, d_2) + \sup_{i \in I} \Delta^\beta_J(f(i), g(i)).$$

*The function* $m^*_n$ *in the definition of the approximability of* $\boldsymbol{\Delta}$ *discretizes points in* $X$ *to* $J_n$*, and* $m_n$ *reconstructs points in* $X$ *from* $J_n$*. Finite-composability of* $\boldsymbol{\Delta}$ *means the composability of* $\boldsymbol{\Delta}$ *in the discrete case.*

These properties allow us to extend composability of divergences in the discrete case, witnessed by finite-composability, to the continuous case. Finite-composability is often known for standard divergences, or can be established by direct calculations. If $\boldsymbol{\Delta}$ is approximable and continuous, finite-composability implies composability.

**Theorem 2.** *A continuous approximable* $A$*-graded family* $\boldsymbol{\Delta}$ *is composable if finite-composable.*

### B. Basic Properties of $f$-divergences

To discuss basic properties of divergences for DP, RDP, zCDP, and tCDP, we begin with basic properties of $f$-divergences since DP can be formulated by a graded family $\mathbf{\Delta}^{\mathtt{DP}} = \{\Delta^{\mathtt{DP}(\varepsilon)}\}_{0\le\varepsilon}$ of $f$-divergences, and Rényi divergences are logarithms of $f$-divergences. An $f$-divergence $\Delta^f$ of subprobability measures is defined in the same way as $f$-divergence of probability measures (4). The $f$-divergences are not necessarily positive for subprobability measures, though they are positive for proper probability measures. We can extend the continuity of $f$-divergences [20, Theorem 16] to support subprobability measures.

**Theorem 3** (Cf. [20, Theorem 16])**.** *For any weight function $f$, the $f$-divergence $\Delta^f$ is continuous:*[3] *for any* subprobability measures $\mu_1, \mu_2 \in \mathcal{G}X$ *on $X$, we have*

$$\Delta^f_X(\mu_1,\mu_2) = \sup_{\substack{\{A_i\}_{i=0}^n \\ \text{partition of } X}} \sum_{i=0}^n \mu_2(A_i) f\left(\frac{\mu_1(A_i)}{\mu_2(A_i)}\right).$$

As we have seen, DP can be formulated by the $\mathbb{R}_{\ge 0}$-graded family $\mathbf{\Delta}^{\mathtt{DP}} = \{\Delta^{\mathtt{DP}(\varepsilon)}\}_{0\le\varepsilon}$ of $f$-divergences, while the Rényi divergences supporting RDP, zCDP, and tCDP are logarithms of $f$-divergences. Before proving basic properties of divergences for DP, RDP, zCDP, and tCDP, we first need two important basic properties of $f$-divergences, continuity and approximability, and we show that finite-composability of $f$-divergences are extended to (proper) composability.

**Theorem 4.** *The $f$-divergence $\Delta^f$ is approximable for any weight function $f$.*

Thus, finite-composable $f$-divergences are composable.

**Theorem 5.** *An $A$-graded family $\mathbf{\Delta} = \{\Delta^{f_\alpha}\}_{\alpha\in A}$ of $f_\alpha$-divergences is composable if it is finite-composable.*

We remark here that any composable family of $f$-divergences is also additive by applying Theorem 1, since $f$-divergences are always continuous (Theorem 3).

### C. Properties of Divergences for DP, RDP, zCDP, and tCDP

As we have seen, DP can be formulated by the $\mathbb{R}_{\ge 0}$-graded family $\mathbf{\Delta}^{\mathtt{DP}}$ of $f$-divergences. By Theorem 1 and 5 and [15, Theorem 1], we obtain the basic properties of the divergences $\mathbf{\Delta}^{\mathtt{DP}}$ for DP as follows:

**Theorem 6** (Cf. [15, Theorem 1])**.** *The $\mathbb{R}_{\ge 0}$-graded family $\mathbf{\Delta}^{\mathtt{DP}} = \{\Delta^{\mathtt{DP}(\varepsilon)}\}_{0\le\varepsilon}$ is reflexive, continuous, approximable, composable, and additive.*

Similarly, we can obtain basic properties for RDP, zCDP, and tCDP. By Theorem 3 and Theorem 4, the exponential $\exp(D^\alpha)$ of the $\alpha$-Rényi divergence is continuous and approximable because it is an $f$-divergence with weight function $t \mapsto \exp(\alpha/(1-\alpha))t^\alpha$.

[3]Note that a measurable finite partition $\{A_i\}_{i=0}^n$ on $X$ is equivalent to a measurable function $k\colon X \to I$ where $I = \{0, 1, \dots, n\}$.

Since the logarithm function is monotone and continuous except at 0, Rényi divergence is continuous and approximable too. Reflexivity and finite-composability of Rényi divergences follow by direct calculations. Theorem 5 yields:

**Theorem 7.** *For any $\alpha > 1$, the Rényi divergence $D^\alpha$ of order $\alpha$ is reflexive, continuous, approximable, composable, and additive (as a singleton-graded family).*

We extend the following properties of Rényi divergences which give the transitive laws of RDP and zCDP to support subprobability measures; an analogous law for tCDP is not currently known.

**Proposition IV.1** (Cf. [17, Theorem 3])**.** *We have*

$$1 < \alpha \le \beta \implies D^\alpha_X(\mu_1||\mu_2) \le D^\beta_X(\mu_1||\mu_2).$$

**Proposition IV.2** (Cf. [27, Lemma 4.1])**.** *For any $\alpha > 1$, $\mu_1, \mu_2, \mu_3 \in \mathcal{G}X$, and $p, q > 1$ satisfying $\frac{1}{p} + \frac{1}{q} = 1$, we have*

$$D^\alpha_X(\mu_1||\mu_3) \le \frac{p\alpha - 1}{p(\alpha-1)} D^{p\alpha}_X(\mu_1||\mu_2) + D^{\frac{q}{p}(p\alpha-1)}_X(\mu_1||\mu_2).$$

As we have seen in Section II-D, we can define divergences for zCDP and tCDP by Equation (6) and Equation (7). Explicitly, we introduce the divergences for zCDP and tCDP by $\Delta^{\mathtt{zCDP}(\xi,\rho)} = \sup_{1<\alpha} \frac{1}{\alpha}(D^\alpha - \xi)$ and $\Delta^{\mathtt{tCDP}(\omega)(\rho)} = \sup_{1<\alpha<\omega} \frac{1}{\alpha} D^\alpha$ respectively. Since two supremums are commutative ($\sup_x \sup_y A(x,y) = \sup_y \sup_x A(x,y)$) in general, the following basic properties of the graded family of zCDP and the divergence of tCDP are obtained from Theorem 7.

**Theorem 8.** *The $\mathbb{R}_{\ge 0}$-graded family $\mathbf{\Delta}^{\mathtt{zCDP}} = \{\Delta^{\mathtt{zCDP}(\xi)}\}_{0\le\xi}$ for zCDP is reflexive, continuous, composable, and additive.*

**Theorem 9.** *For each $1 < \omega$, the divergence $\Delta^{\mathtt{tCDP}(\omega)}$ for $\omega$-tCDP is reflexive, continuous, composable, and additive.*

Note that we may not have approximability, but the family is still composable. These results also hold for subprobability measures where Rényi divergence and divergences for zCDP and tCDP are defined in a way similar to Equation (1) and Equation (2) respectively.

## V. Approximate Span-Lifting

We are now ready to combine graded divergences with spans, leading to our new relational liftings. Given an $A$-graded family $\mathbf{\Delta} = \{\Delta^\alpha\}_{\alpha\in A}$ of divergences, we introduce a graded monad on $\mathbf{Span}(\mathbf{Meas})$ called the *approximate span-lifting*[4] $(-)^{\sharp(\mathbf{\Delta},\alpha,\delta)}$ for the family $\mathbf{\Delta}$, where $\alpha \in A$ and $\delta \in \overline{\mathbb{R}}$. We first define its action on objects.

**Definition 13.** *We define the span-constructor $(-)^{\sharp(\mathbf{\Delta},\alpha,\delta)}$ as follows: for any $(X, Y, \Phi, \rho_1, \rho_2)$ in $\mathbf{Span}(\mathbf{Meas})$, we define the $\mathbf{Span}(\mathbf{Meas})$-object*

$$\begin{aligned}&(X, Y, \Phi, \rho_1, \rho_2)^{\sharp(\mathbf{\Delta},\alpha,\delta)} \\ &= (\mathcal{G}X, \mathcal{G}Y, W(\Phi, \mathbf{\Delta}, \alpha, \delta),\ \mathcal{G}\rho_1 \circ \pi_1,\ \mathcal{G}\rho_1 \circ \pi_2)\end{aligned}$$

[4]The name is inspired from "approximate lifting" [15].

*where* $W(\Phi, \boldsymbol{\Delta}, \alpha, \delta) = \{ (\nu_1, \nu_2) \mid \Delta^{\alpha}_{\Phi}(\nu_1, \nu_2) \leq \delta \}$. *We deal it as a subspace of the measurable space* $\mathcal{G}\Phi \times \mathcal{G}\Phi$.

Intuitively, $(X, Y, \Phi, \rho_1, \rho_2)^{\sharp(\boldsymbol{\Delta},\alpha,\delta)}$ relates subprobability measures with $\Delta^{\alpha}$-distance at most $\delta$. The set $W(\Phi, \boldsymbol{\Delta}, \alpha, \delta)$ contains all possible witness distributions, and $\pi_1$ and $\pi_2$ are canonical projections from $W(\Phi, \boldsymbol{\Delta}, \alpha, \delta)$ to $\mathcal{G}\Phi$. As a special case, the approximate span-lifting $(-)^{\sharp(\boldsymbol{\Delta},\alpha,\delta)}$ recovers the divergence $\Delta^{\alpha}$ by applying the equality relation $(X, X, \mathtt{Eq}_X, \pi_1, \pi_1)^{\sharp(\boldsymbol{\Delta},\alpha,\delta)}$.

**Theorem 10.** *For any* $A$*-graded family* $\boldsymbol{\Delta}$*,* $\alpha \in A$*, and* $\delta \in \overline{\mathbb{R}}$*,*

$$(X, X, X, \mathtt{id}_X, \mathtt{id}_X)^{\sharp(\boldsymbol{\Delta},\alpha,\delta)} = (\mathcal{G}X, \mathcal{G}X, \{ (\mu_1, \mu_2) \mid \Delta^{\alpha}_{X}(\mu_1, \mu_2) \leq \delta \}, \pi_1, \pi_2).$$

Here, the span $(X, X, X, \mathrm{id}_X, \mathrm{id}_X)$ is isomorphic to the equality relation $(X, X, \mathrm{Eq}_X, \pi_1|_{\mathrm{Eq}_X}, \pi_1|_{\mathrm{Eq}_X})$.

Next, we give approximate span-liftings the structure of a graded monad with double strength. We consider the case where $\boldsymbol{\Delta}$ is a reflexive, composable, and additive $A$-graded family of divergences; we can sometimes recover more limited versions of approximate span-liftings by dropping or weakening these properties.

**Theorem 11.** *If an* $A$*-graded family* $\boldsymbol{\Delta}$ *is reflexive, composable, and additive, then the approximate span-lifting* $(-)^{\sharp(\boldsymbol{\Delta},\alpha,\delta)}$ *forms an* $A \times \overline{\mathbb{R}}$*-graded monad with double strength. Namely, there are maps*

**Functor:** *For any morphism* $(h, k, l) \colon (X, Y, \Phi, \rho_1, \rho_2) \to (Z, W, \Psi, \rho'_1, \rho'_2)$ *in* $\mathbf{Span}(\mathbf{Meas})$ *and* $(\alpha, \delta) \in A \times \overline{\mathbb{R}}$*,*

$$(\mathcal{G}h, \mathcal{G}k, \mathcal{G}l \times \mathcal{G}l) \colon (X, Y, \Phi, \rho_1, \rho_2)^{\sharp(\boldsymbol{\Delta},\alpha,\delta)} \to (Z, W, \Psi, \rho'_1, \rho'_2)^{\sharp(\boldsymbol{\Delta},\alpha,\delta)}.$$

**Unit:** *For any morphism* $(X, Y, \Phi, \rho_1, \rho_2)$ *in* $\mathbf{Span}(\mathbf{Meas})$*,*

$$(\eta_X, \eta_Y, \langle \eta_\Phi, \eta_\Phi \rangle) \colon (X, Y, \Phi, \rho_1, \rho_2) \to (X, Y, \Phi, \rho_1, \rho_2)^{\sharp(\boldsymbol{\Delta},1_A,0)}.$$

**Kleisli lifting:** *For any* $(\beta, \gamma) \in A \times \overline{\mathbb{R}}$ *and any morphism* $(h, k, l) \colon (X, Y, \Phi, \rho_1, \rho_2) \to (Z, W, \Psi, \rho'_1, \rho'_2)^{\sharp(\boldsymbol{\Delta},\alpha,\delta)}$ *in* $\mathbf{Span}(\mathbf{Meas})$*,*

$$(h^\sharp, k^\sharp, (\pi_1 \circ l)^\sharp \times (\pi_2 \circ l)^\sharp) \colon (X, Y, \Phi, \rho_1, \rho_2)^{\sharp(\boldsymbol{\Delta},\beta,\gamma)} \to (Z, W, \Psi, \rho'_1, \rho'_2)^{\sharp(\boldsymbol{\Delta},\alpha\beta,\delta+\gamma)}$$

**Inclusions:** *For any* $(X, Y, \Phi, \rho_1, \rho_2)$ *in* $\mathbf{Span}(\mathbf{Meas})$*, and any* $\alpha \preceq \beta$ *and* $\delta \leq \gamma$*,*

$$(\mathtt{id}_{\mathcal{G}X}, \mathtt{id}_{\mathcal{G}Y}, \mathtt{id}_{\mathcal{G}\Phi} \times \mathtt{id}_{\mathcal{G}\Phi}) \colon (X, Y, \Phi, \rho_1, \rho_2)^{\sharp(\boldsymbol{\Delta},\alpha,\delta)} \to (X, Y, \Phi, \rho_1, \rho_2)^{\sharp(\boldsymbol{\Delta},\beta,\gamma)}.$$

**Double strength:** *For all* $(\alpha, \delta)$ *and* $(\beta, \gamma)$ *in* $A \times \overline{\mathbb{R}}$ *and all* $(X, Y, \Phi, \rho_1, \rho_2)$ *and* $(Z, W, \Psi, \rho'_1, \rho'_2)$ *in* $\mathbf{Span}(\mathbf{Meas})$*, by letting* $\theta_i = \mathtt{dst}_{\Phi,\Psi} \circ (\pi_i \times \pi_i)$ *where* $i = 1, 2$*,*

$$(\mathtt{dst}_{X,Z}, \mathtt{dst}_{Y,W}, \langle \theta_1, \theta_2 \rangle) \colon (X, Y, \Phi, \rho_1, \rho_2)^{\sharp(\boldsymbol{\Delta},\alpha,\delta)} \mathbin{\dot{\times}} (Z, W, \Psi, \rho'_1, \rho'_2)^{\sharp(\boldsymbol{\Delta},\beta,\gamma)} \to ((X, Y, \Phi, \rho_1, \rho_2) \mathbin{\dot{\times}} (Z, W, \Psi, \rho'_1, \rho'_2))^{\sharp(\boldsymbol{\Delta},\alpha\beta,\delta+\gamma)}$$

*a) Approximate Span-liftings for Privacy:* Finally, we get approximate span-liftings for DP, RDP, zCDP, and tCDP by combining Theorems 6, 7, 8, and 9 with Theorem 11.

**Theorem 12** (Approximate span-lifting for DP, RDP, zCDP, tCDP)**.** *The span-liftings listed in Figure 1 are actually graded liftings with a double strength of the monad* $\mathcal{G} \times \mathcal{G}$ *along* $U \colon \mathbf{Span}(\mathbf{Meas}) \to \mathbf{Meas} \times \mathbf{Meas}$*.*

## VI. Case Study: the Program Logic Span-apRHL

The previous section showed that the relaxations of differential privacy RDP, zCDP, and tCDP can be captured by relational liftings with categorical properties similar to the ones of the relational liftings used for standard differential privacy. As a result, we can use these liftings to give the semantic foundation for formal verification of these relaxations. To demonstrate a concrete application, we design a relational program logic span-apRHL that can prove DP, RDP, zCDP, and tCDP for randomized algorithms using continuous random sampling.

*a) The Language pWHILE:* We take a standard, first-order language pWHILE, augmenting the usual imperative commands with a random sampling statement (we omit the grammar of expressions, which is largely standard).

$$\tau ::= \mathtt{bool} \mid \mathtt{int} \mid \mathtt{real} \mid \tau^d \ (d \in \mathbb{N}) \mid \dots$$

$$\nu ::= \mathtt{Dirac}(e) \mid \mathtt{Bern}(e) \mid \mathtt{Lap}(e_1, e_2) \mid \mathtt{Gauss}(e_1, e_2) \mid \dots$$

$$c ::= \mathtt{skip} \mid x \xleftarrow{\$} \nu \mid c_1; c_2 \mid \mathtt{if}\ e\ \mathtt{then}\ c_1\ \mathtt{else}\ c_2 \mid \mathtt{while}\ e\ \mathtt{do}\ c$$

The type system is standard, and the value types are interpreted as measurable spaces. To give a semantics to expressions, distribution expressions, and commands, we interpret their associated typing/well-formedness judgments in a context $\Gamma$, interpreted as a product space. We interpret an expression judgment $\Gamma \vdash^t e \colon \tau$ as a map $[\![\Gamma \vdash^t e \colon \tau]\!] \colon [\![\Gamma]\!] \to [\![\tau]\!]$ in $\mathbf{Meas}$; we interpret a probabilistic expression judgment $\Gamma \vdash^p \nu \colon \tau$ as a map $[\![\Gamma \vdash^p \nu \colon \tau]\!] \colon [\![\Gamma]\!] \to \mathcal{G}[\![\tau]\!]$ in $\mathbf{Meas}$; and we interpret a command judgment $\Gamma \vdash c$ as a map $[\![\Gamma \vdash c]\!] \colon [\![\Gamma]\!] \to \mathcal{G}[\![\Gamma]\!]$ in $\mathbf{Meas}$.

*b) Relational Assertions:* Our assertion logic uses formulas of the form

$$\Phi, \Psi ::= \mathcal{E} \mid \Phi \wedge \Psi \mid \Phi \vee \Psi \mid \neg\Phi$$

where $\mathcal{E}$ represents basic relational expressions which are first-order formulas over expressions where program variables are tagged with the symbols $\langle 1 \rangle$ and $\langle 2 \rangle$, e.g. $x\langle 1 \rangle \leq x\langle 2 \rangle$. Relational expressions are interpreted as formulae over pairs of memories, and the symbols $\langle 1 \rangle$ and $\langle 2 \rangle$ indicate whether a variable should be interpreted in the first or second memory.

Since we use span-liftings instead of relational liftings, we interpret relational assertions as spans, that is, as $\mathbf{Span}(\mathbf{Meas})$-objects. This can be done by first interpreting assertions $\Gamma \vdash^R \Phi$ as binary relations $[\![\Phi]\!] \subseteq [\![\Gamma]\!] \times [\![\Gamma]\!]$, and then converting them to spans $([\![\Gamma]\!], [\![\Gamma]\!], [\![\Phi]\!], \pi_1, \pi_2)$.

| Privacy | (Graded family of )Divergence | Approximate span-lifting | Grading Monoid |
|---|---|---|---|
| DP | $\mathbf{\Delta}^{\mathtt{DP}} = \{\Delta^{\mathtt{DP}(\varepsilon)}\}_{0\le\varepsilon}$ | $\{(-)^{\sharp(\mathbf{\Delta}^{\mathtt{DP}},\varepsilon,\delta)}\}_{0\le\varepsilon,0\le\delta}$ | $\mathbb{R}_{\ge 0}\times\mathbb{R}_{\ge 0}$ |
| RDP | $D^\alpha$ (Rényi divergence; see (1)) | $\{(-)^{\sharp(D^\alpha,*,\rho)}\}_{*\in\{*\},\rho\in\overline{\mathbb{R}}}$ | $\overline{\mathbb{R}}$ |
| zCDP | $\mathbf{\Delta}^{\mathtt{zCDP}} = \{\Delta^{\mathtt{zCDP}(\xi)}\}_{0\le\xi}$ (see (6)) | $\{(-)^{\sharp(\mathbf{\Delta}^{\mathtt{zCDP}},\xi,\rho)}\}_{0\le\xi,\rho\in\overline{\mathbb{R}}}$ | $\mathbb{R}_{\ge 0}\times\overline{\mathbb{R}}$ |
| tCDP | $\mathbf{\Delta}^{\mathtt{tCDP}(\omega)} = \{\Delta^{\mathtt{tCDP}(\omega)}\}$ (see (7)) | $\{(-)^{\sharp(\Delta^{\mathtt{tCDP}(\omega)},*,\rho)}\}_{*\in\{*\},\rho\in\overline{\mathbb{R}}}$ | $\overline{\mathbb{R}}$ |

Fig. 1. span-liftings for DP,zCDP, RDP, tCDP

*c) Relational Program Logic Judgments, Axioms and Rules:* In span-apRHL we can prove judgments corresponding to differential privacy, RDP, zCDP, and tCDP. For well-typed commands $\Gamma \vdash c_1$ and $\Gamma \vdash c_2$ and assertions $\Gamma \vdash^R \Phi$ and $\Gamma \vdash^R \Psi$, we define judgments:

$$\Gamma \vdash c_1 \sim^{\mathtt{DP}}_{\varepsilon,\delta} c_2 \colon \Phi \implies \Psi \quad \text{for } (\varepsilon,\delta)\text{-DP}$$
$$\Gamma \vdash c_1 \sim^{\alpha-\mathtt{RDP}}_{\rho} c_2 \colon \Phi \implies \Psi \quad \text{for } (\alpha,\rho)\text{-RDP}$$
$$\Gamma \vdash c_1 \sim^{\mathtt{zCDP}}_{\xi,\rho} c_2 \colon \Phi \implies \Psi \quad \text{for } (\xi,\rho)\text{-zCDP}$$
$$\Gamma \vdash c_1 \sim^{\mathtt{tCDP}(\omega)}_{\rho} c_2 \colon \Phi \implies \Psi \quad \text{for } (\omega,\rho)\text{-tCDP}$$

We divide the proof rules of span-apRHL in four classes: basic rules (Figure 2), rules for basic mechanisms (Figure 3), rules for reasoning about transitivity (Figure 4), and rules for conversions (Figure 5). The basic rules can be used to reason about either differential privacy, RDP, zCDP, or tCDP. We describe the basic rules in a parametric way by considering $\{\sim^{\mathbf{\Delta}}_{\alpha,\delta}\}_{\alpha\in A,0\le\delta}$ to stand for one of the families $\{\sim^{\mathrm{DP}}_{\varepsilon,\delta}\}_{0\le\varepsilon,0\le\delta}$, $\{\sim^{\alpha-\mathrm{RDP}}_{\rho}\}_{*\in\{*\},0\le\rho}$, $\{\sim^{\mathrm{zCDP}}_{\xi,\rho}\}_{0\le\xi,0\le\rho}$, and $\{\sim^{\omega-\mathrm{tCDP}}_{\rho}\}_{0\le\rho}$. We give a selection of the more interesting proof rules in Figure 2, and defer the rest of the rules to the appendix. Here, we comment briefly on the rules. The [assn] rule for assignment is mostly standard, the only detail is that the index $1_A$ now depends on which notion of privacy we want to use. The rule [seq] is the sequential composition of commands and takes the same form no matter which family of divergence we consider. The rule [weak] is our version of the usual consequence rule, where additionally we can weaken also the privacy parameters for the various privacy definitions.

In Figure 3, we show some rules for two basic mechanisms that we support: Gauss and Sinh-normal. Rules for the other mechanisms are in the appendix. The rules [RDP-G], [zCDP-G], [tCDP-G] and [DP-G] are all rules for the Gaussian mechanism. They differ in terms of the privacy definition they provide, and for the values of the privacy parameters they achieve. These values correspond to the ones that can be obtained by analyzing the Gaussian mechanism in the different relaxations of differential privacy. The rule [tCDP-SinhG] is similar to the other rules but it supports the Sinh-normal mechanism as analyzed in [13].

In Figure 4, we show rules for transitivity in span-apRHL. Transitivity is important because it allows one to reason about group privacy [26]. The different flavors of the logic have different numeric parameters for these rules, reflecting the slight differences in group privacy [26], [12], [11]. Finally, Figure 5 gives rules for converting between judgments for different flavors of differential privacy. In some of them we have a loss in the parameters, in others there is no loss. These rules correspond to the different conversion theorems for the different logics [12], [11]. Notice that most of these rules require lossless programs because they have been formulated in terms of distributions, rather than subdistributions.

### A. Denotational Semantics of pWHILE

To prove the soundness of span-apRHL we interpret pWHILE in $\mathbf{Meas}$ using the sub-Giry monad $\mathcal{G}$. We interpret an expression judgment $\Gamma \vdash^t e\colon \tau$ as a measurable function $[\![\Gamma \vdash^t e\colon \tau]\!]\colon [\![\Gamma]\!] \to [\![\tau]\!]$; for instance, the variable case $\Gamma \vdash^t x\colon \tau$ is interpreted as the projection $\pi_x\colon [\![\Gamma]\!] \to [\![\tau]\!]$. Note that all operators $\oplus$ and comparisons $\bowtie$ are interpreted to measurable functions $\oplus\colon [\![\tau]\!]\times[\![\tau]\!] \to [\![\tau]\!]$ and $\bowtie\colon [\![\tau]\!]\times[\![\tau]\!] \to [\![\mathtt{bool}]\!]$ respectively. Likewise, we interpret a distribution expression judgment $\Gamma \vdash^p \nu\colon \tau$ as a measurable function $[\![\Gamma \vdash^p \nu\colon \tau]\!]\colon [\![\Gamma]\!] \to \mathcal{G}[\![\tau]\!]$; for instance, the Gaussian expression $\Gamma \vdash^p \mathtt{Gauss}(e_1,e_2)\colon \mathtt{real}$ is interpreted as a Gaussian distribution: $\mathcal{N}([\![\Gamma \vdash^t e_1\colon \mathtt{real}]\!],[\![\Gamma \vdash^t e_2\colon \mathtt{real}]\!])$. Finally, we interpret a judgment $\Gamma \vdash c$ as a measurable function $[\![\Gamma \vdash c]\!]\colon [\![\Gamma]\!] \to \mathcal{G}[\![\Gamma]\!]$ defined inductively as

$$[\![\Gamma \vdash x \xleftarrow{\$} \nu]\!] = \mathcal{G}(\mathrm{rw}\langle\Gamma \mid x\colon\tau\rangle) \circ \mathrm{st}_{[\![\Gamma]\!],[\![\tau]\!]} \circ \langle \mathrm{id}_{[\![\Gamma]\!]}, [\![\nu]\!]\rangle,$$

$$[\![\Gamma \vdash c_1; c_2]\!] = [\![\Gamma \vdash c_2]\!]^\sharp \circ [\![\Gamma \vdash c_1]\!], \qquad [\![\Gamma \vdash \mathtt{skip}]\!] = \eta_{[\![\Gamma]\!]}$$

$$[\![\Gamma \vdash \mathtt{if}\ b\ \mathtt{then}\ c_1\ \mathtt{else}\ c_2]\!] =$$

$$[[\![\Gamma \vdash c_1]\!],[\![\Gamma \vdash c_2]\!]] \circ \mathrm{br}\langle\Gamma\rangle \circ \langle [\![\Gamma \vdash b]\!], \mathrm{id}_{[\![\Gamma]\!]}\rangle$$

Here, $\mathrm{rw}\langle\Gamma \mid x\colon\tau\rangle\colon [\![\Gamma]\!]\times[\![x\colon\tau]\!] \to [\![\Gamma]\!]$ $(x\colon\tau\in\Gamma)$ is an overwriting operation of memory $((a_1,\dots,a_k,\dots,a_n),b_k) \mapsto (a_1,\dots,b_k,\dots,a_n)$, which is encoded using projection mappings of Cartesian products in $\mathbf{Meas}$. The function $\mathrm{br}\langle\Gamma\rangle\colon 2\times[\![\Gamma]\!] \to [\![\Gamma]\!]+[\![\Gamma]\!]$ comes from the canonical isomorphism $2\times[\![\Gamma]\!] \cong [\![\Gamma]\!]+[\![\Gamma]\!]$ given from the distributivity of $\mathbf{Meas}$.

To interpret loops, we introduce the dummy "abort" command $\Gamma \vdash \mathtt{null}$ that is interpreted by the null/zero measure over $[\![\Gamma]\!]$, and the following commands corresponding to the finite unrollings of the loop:

$$[\mathtt{while}\ b\ \mathtt{do}\ c]_n = \begin{cases} \mathtt{if}\ b\ \mathtt{then}\ \mathtt{null}\ \mathtt{else}\ \mathtt{skip}, & \text{if } n=0 \\ \mathtt{if}\ b\ \mathtt{then}\ c; [\mathtt{while}\ b\ \mathtt{do}\ c]_k, & \text{if } n=k+1 \end{cases}$$

$$\Gamma \vdash x_1 \leftarrow e_1 \sim^{\mathbf{\Delta}}_{1_A,0} x_2 \leftarrow e_2 \colon \Phi\{e_1\langle 1\rangle, e_2\langle 2\rangle / x_1\langle 1\rangle, x_2\langle 2\rangle\} \implies \Phi \quad \text{[assn]}$$

$$\frac{\Gamma \vdash c_1 \sim^{\mathbf{\Delta}}_{\alpha,\delta} c_1' \colon \Phi \implies \Phi' \quad \Gamma \vdash c_2 \sim^{\mathbf{\Delta}}_{\beta,\gamma} c_2' \colon \Phi' \implies \Psi}{\Gamma \vdash c_1; c_2 \sim^{\mathbf{\Delta}}_{\alpha\beta,\delta+\gamma} c_1'; c_2' \colon \Phi \implies \Psi} \text{[seq]}$$

$$\frac{\Gamma \vdash^I \Phi' \implies \Phi \quad \Gamma \vdash^I \Psi \implies \Psi' \quad \Gamma \vdash c_1 \sim^{\mathbf{\Delta}}_{\alpha,\delta} c_2 \colon \Phi \implies \Psi \quad \alpha \le \beta \quad \delta \le \gamma}{\Gamma \vdash c_1 \sim^{\mathbf{\Delta}}_{\beta,\gamma} c_2 \colon \Phi' \implies \Psi'} \text{[weak]}$$

Fig. 2. Selection of span-apRHL basic rules (implications $\vdash^I$ are definecd on the next page).

$$\Gamma \vdash x_1 \xleftarrow{\$} \mathtt{Gauss}(e_1, \sigma^2) \sim^{\alpha-\mathtt{RDP}}_{\alpha r^2/2\sigma^2} x_2 \xleftarrow{\$} \mathtt{Gauss}(e_2, \sigma^2) \colon (|e_1\langle 1\rangle - e_2\langle 2\rangle| \le r) \implies (x_1\langle 1\rangle = x_2\langle 2\rangle) \quad \text{[RDP-G]}$$

$$\Gamma \vdash x_1 \xleftarrow{\$} \mathtt{Gauss}(e_1, \sigma^2) \sim^{\mathtt{zCDP}}_{0,r^2/2\sigma^2} x_2 \xleftarrow{\$} \mathtt{Gauss}(e_2, \sigma^2) \colon (|e_1\langle 1\rangle - e_2\langle 2\rangle| \le r) \implies (x_1\langle 1\rangle = x_2\langle 2\rangle) \quad \text{[zCDP-G]}$$

$$\Gamma \vdash x_1 \xleftarrow{\$} \mathtt{Gauss}(e_1, \sigma^2) \sim^{\mathtt{tCDP}}_{0,r^2/2\sigma^2} x_2 \xleftarrow{\$} \mathtt{Gauss}(e_2, \sigma^2) \colon (|e_1\langle 1\rangle - e_2\langle 2\rangle| \le r) \implies (x_1\langle 1\rangle = x_2\langle 2\rangle) \quad \text{[tCDP-G]}$$

$$\frac{\exists c > \frac{1+\sqrt{3}}{2}.\ (2\log(0.66/\delta) \le c^2) \wedge (\frac{cr}{\varepsilon} \le \sigma)}{\Gamma \vdash x_1 \xleftarrow{\$} \mathtt{Gauss}(e_1, \sigma^2) \sim^{\mathtt{DP}}_{\varepsilon,\ \delta} x_2 \xleftarrow{\$} \mathtt{Gauss}(e_2, \sigma^2) \colon (|e_1\langle 1\rangle - e_2\langle 2\rangle| \le r) \implies (x_1\langle 1\rangle = x_2\langle 2\rangle)} \quad \text{[DP-G]}$$

$$\frac{1 < 1/\sqrt{\rho} \le A/\delta}{\begin{array}{l}\Gamma \vdash x_1 \xleftarrow{\$} e_1 + A \cdot \mathtt{arsinh}\left(\frac{1}{A}\mathtt{Gauss}(0, \delta^2/2\rho)\right) \\ \quad \sim^{A/8\delta-\mathtt{tCDP}}_{16\rho} x_2 \xleftarrow{\$} e_2 + A \cdot \mathtt{arsinh}\left(\frac{1}{A}\mathtt{Gauss}(0, \delta^2/2\rho)\right) : \\ \quad\quad (|e_1\langle 1\rangle - e_2\langle 2\rangle| \le \delta) \implies (x_1\langle 1\rangle = x_2\langle 2\rangle)\end{array}} \quad \text{[tCDP-SinhG]}$$

Fig. 3. Rules for basic mechanisms for DP, RDP, zCDP, and tCDP in span-apRHL.

$$\frac{\Gamma \vdash c_1 \sim^{\mathtt{DP}}_{\varepsilon_1,\delta_1} c_2 \colon \Phi \implies x_1\langle 1\rangle = x_2\langle 2\rangle \quad \Gamma \vdash c_2 \sim^{\mathtt{DP}}_{\varepsilon_2,\delta_2} c_3 \colon \Psi \implies x_2\langle 1\rangle = x_3\langle 2\rangle}{\Gamma \vdash c_1 \sim^{\mathtt{DP}}_{\varepsilon_1+\varepsilon_2,\ \max(e^{\varepsilon_2}\delta_1+\delta_2, e^{\varepsilon_1}\delta_2+\delta_1)} c_3 \colon \Phi \circ \Psi \implies x_1\langle 1\rangle = x_3\langle 2\rangle} \quad \text{[DP-Trans]}$$

$$\frac{\begin{array}{l}\Gamma \vdash c_1 \sim^{p\alpha-\mathtt{RDP}}_{\rho_1} c_2 \colon \Phi \implies x_1\langle 1\rangle = x_2\langle 2\rangle \\ \Gamma \vdash c_2 \sim^{q(p\alpha-1)/p-\mathtt{RDP}}_{\rho_2} c_3 \colon \Psi \implies x_2\langle 1\rangle = x_3\langle 2\rangle \quad \frac{1}{p} + \frac{1}{q} = 1 \quad 1 < p \quad 1 < q\end{array}}{\Gamma \vdash c_1 \sim^{\alpha-\mathtt{RDP}}_{((p\alpha-1)\rho_1/p(\alpha-1))+\rho_2} c_3 \colon \Phi \circ \Psi \implies x_1\langle 1\rangle = x_3\langle 2\rangle} \quad \text{[RDP-Trans]}$$

$$\frac{\begin{array}{l}\Gamma \vdash c_1 \sim^{\mathtt{zCDP}}_{\xi(k-1)\sum_{i=1}^{k-1},(k^2-1)\rho} c_2 \colon \Phi \implies x_1\langle 1\rangle = x_2\langle 2\rangle \\ \Gamma \vdash c_2 \sim^{\mathtt{zCDP}}_{\xi,\rho} c_3 \colon \Psi \implies x_2\langle 1\rangle = x_3\langle 2\rangle \quad k \in \mathbb{N} \quad 1 < k\end{array}}{\Gamma \vdash c_1 \sim^{\mathtt{zCDP}}_{\xi k\sum_{i=1}^{k},k^2\rho} c_3 \colon \Phi \circ \Psi \implies x_1\langle 1\rangle = x_3\langle 2\rangle} \quad \text{[zCDP-Trans]}$$

Fig. 4. Span-apRHL transitivity rules for group privacy

We then interpret loops as:

$$[\![\Gamma \vdash \mathtt{while}\ b\ \mathtt{do}\ c]\!] = \sup_{n\in\mathbb{N}} [\![\Gamma \vdash [\mathtt{while}\ e\ \mathtt{do}\ c]_n]\!].$$

This is well-defined, since the family $\{[\![\Gamma \vdash [\mathtt{while}\ e\ \mathtt{do}\ c]_n]\!]\}_{n\in\mathbb{N}}$ is an $\omega$-chain with respect to the $\omega\mathbf{CPO}_\perp$-enrichment $\sqsubseteq$ of $\mathbf{Meas}_\mathcal{G}$.

### *B. Semantics of Relations*

Since we use span-liftings instead of relational liftings, we need to interpret relation expressions to spans, that is, $\mathbf{Span}(\mathbf{Meas})$-objects. We proceed in two steps: first interpreting expressions as binary relations, and then converting relations to spans. In the first step, we interpret a relation expression $\Gamma \vdash^R \Phi$ as a binary relation over $[\![\Gamma]\!]$:

$$\begin{array}{l}(\!|\Gamma \vdash^R e_1\langle 1\rangle \bowtie e_2\langle 2\rangle|\!) \\ = \left\{ (m_1, m_2) \;\middle|\; [\![\Gamma \vdash^t e_1 \colon \tau]\!](m_1) \bowtie [\![\Gamma \vdash^t e_2 \colon \tau]\!](m_2) \right\} \\ (\!|\Gamma \vdash^R (e_1\langle 1\rangle \otimes_1 e_2\langle 2\rangle) \bowtie (e_3\langle 1\rangle \otimes_2 e_4\langle 2\rangle)|\!) \\ = \left\{ (m_1, m_2) \;\middle|\; \begin{array}{l}[\![\Gamma \vdash^t e_1 \colon \tau]\!](m_1) \otimes_1 [\![\Gamma \vdash^t e_2 \colon \tau]\!](m_2) \\ \bowtie [\![\Gamma \vdash^t e_3 \colon \tau]\!](m_1) \otimes_2 [\![\Gamma \vdash^t e_4 \colon \tau]\!](m_2)\end{array} \right\}\end{array}$$

$$\frac{\Gamma \vdash c_1 \sim^{\mathtt{DP}}_{\varepsilon,\ 0} c_2 \colon \Phi \implies \Psi \quad c_1, c_2 \colon \text{lossless}}{\Gamma \vdash c_1 \sim^{\mathtt{zCDP}}_{\varepsilon^2/2,\ 0} c_2 \colon \Phi \implies \Psi \quad c_1, c_2 \colon \text{lossless}}\ [\text{D/z}] \qquad \frac{\Gamma \vdash c_1 \sim^{\mathtt{zCDP}}_{0,\ \rho} c_2 \colon \Phi \implies \Psi}{\forall \alpha > 1.\ \Gamma \vdash c_1 \sim^{\alpha-\mathtt{RDP}}_{\rho} c_2 \colon \Phi \implies \Psi}\ [\text{z/R}]$$

$$\frac{\Gamma \vdash c_1 \sim^{\mathtt{zCDP}}_{\xi,\ \rho} c_2 \colon \Phi \implies \Psi \quad c_1, c_2 \colon \text{lossless} \quad 0 < \delta < 1}{\Gamma \vdash c_1 \sim^{\mathtt{DP}}_{\xi+\rho+2\sqrt{\rho \log(1/\delta)},\ \delta} c_2 \colon \Phi \implies \Psi}\ [\text{z/D}]$$

$$\frac{\Gamma \vdash c_1 \sim^{\mathtt{tCDP}(\omega)}_{\rho} c_2 \colon \Phi \implies \Psi,\ c_1, c_2 \colon \text{lossless},\ \beta = \min(\omega, 1 + \sqrt{\log(1/\delta)/\rho}),\ 0 < \delta < 1}{\Gamma \vdash c_1 \sim^{\mathtt{DP}}_{\rho\beta+\log(1/\delta)/(\beta-1),\ \delta} c_2 \colon \Phi \implies \Psi}\ [\text{t/D}]$$

$$\frac{\Gamma \vdash c_1 \sim^{\alpha-\mathtt{RDP}}_{\rho} c_2 \colon \Phi \implies \Psi \quad c_1, c_2 \colon \text{lossless} \quad 0 < \delta < 1}{\Gamma \vdash c_1 \sim^{\mathtt{DP}}_{\rho-\log \delta/(\alpha-1),\ \delta} c_2 \colon \Phi \implies \Psi}\ [\text{R/D}]$$

Fig. 5. Rules for conversions between DP, RDP and zCDP in span-apRHL.

We interpret the connectives in the expected way:

$$(\!|\Gamma \vdash^R \Phi \wedge \Psi|\!) = (\!|\Gamma \vdash^R \Phi|\!) \cap (\!|\Gamma \vdash^R \Psi|\!)$$
$$(\!|\Gamma \vdash^R \Phi \vee \Psi|\!) = (\!|\Gamma \vdash^R \Phi|\!) \cup (\!|\Gamma \vdash^R \Psi|\!)$$
$$(\!|\Gamma \vdash^R \neg\Phi|\!) = ([\![\Gamma]\!] \times [\![\Gamma]\!]) \setminus (\!|\Gamma \vdash^R \Phi|\!)$$

The binary relation $(\!|\Gamma \vdash^R \Phi|\!)$ can be converted to the span

$$[\![\Gamma \vdash^R \Phi]\!] = ([\![\Gamma]\!], [\![\Gamma]\!], (\!|\Gamma \vdash^R \Phi|\!), \pi_1|_{(\!|\Gamma \vdash^R \Phi|\!)}, \pi_2|_{(\!|\Gamma \vdash^R \Phi|\!)}).$$

We interpret the implication $\Gamma \vdash^I \Phi \implies \Psi$ by the following morphism in $\mathbf{Span}(\mathbf{Meas})$:

$$[\![\Gamma \vdash^I \Phi \implies \Psi]\!] = (\mathrm{id}_{[\![\Gamma]\!]}, \mathrm{id}_{[\![\Gamma]\!]}, (\mathrm{id}_{[\![\Gamma]\!]} \times \mathrm{id}_{[\![\Gamma]\!]})|_{(\!|\Gamma \vdash^R \Phi|\!)})$$

A judgment $\Gamma \vdash c_1 \sim^{\mathbf{\Delta}}_{\alpha,\delta} c_2 \colon \Phi \implies \Psi$ is valid if there exists $l \colon (\!|\Gamma \vdash^R \Phi|\!) \to W([\![\Gamma \vdash^R \Psi]\!], \mathbf{\Delta}, \alpha, \delta)$ measurable such that the following map is a morphism in $\mathbf{Span}(\mathbf{Meas})$:

$$([\![\Gamma \vdash c_1]\!], [\![\Gamma \vdash c_2]\!], l) \colon [\![\Gamma \vdash^R \Phi]\!] \to [\![\Gamma \vdash^R \Psi]\!]^{\sharp(\mathbf{\Delta},\alpha,\delta)}$$

Finally, we define the validity in span-apRHL as follows:

$$\begin{aligned}
&\Gamma \models c_1 \sim^{\mathtt{DP}}_{\varepsilon,\delta} c_2 \colon \Phi \implies \Psi \\
&\quad \text{iff } \exists l.\ ([\![\Gamma \vdash c_1]\!], [\![\Gamma \vdash c_2]\!], l) \colon [\![\Phi]\!] \to [\![\Psi]\!]^{\sharp(\mathbf{\Delta}^{\mathtt{DP}},\varepsilon,\delta)}, \\
&\Gamma \models c_1 \sim^{\alpha-\mathtt{RDP}}_{\rho} c_2 \colon \Phi \implies \Psi \\
&\quad \text{iff } \exists l.\ ([\![\Gamma \vdash c_1]\!], [\![\Gamma \vdash c_2]\!], l) \colon [\![\Phi]\!] \to [\![\Psi]\!]^{\sharp(D^\alpha,*,\rho)}, \\
&\Gamma \models c_1 \sim^{\mathtt{zCDP}}_{\xi,\rho} c_2 \colon \Phi \implies \Psi \\
&\quad \text{iff } \exists l.\ ([\![\Gamma \vdash c_1]\!], [\![\Gamma \vdash c_2]\!], l) \colon [\![\Phi]\!] \to [\![\Psi]\!]^{\sharp(\mathbf{\Delta}^{\mathtt{zCDP}},\xi,\rho)} \\
&\Gamma \models c_1 \sim^{\omega-\mathtt{tCDP}}_{\rho} c_2 \colon \Phi \implies \Psi \\
&\quad \text{iff } \exists l.\ ([\![\Gamma \vdash c_1]\!], [\![\Gamma \vdash c_2]\!], l) \colon [\![\Phi]\!] \to [\![\Psi]\!]^{\sharp(\mathbf{\Delta}^{\omega-\mathtt{tCDP}},*,\rho)}.
\end{aligned}$$

**Theorem 13.** *If $\Gamma \vdash c_1 \sim^{\mathbf{\Delta}}_{\alpha,\delta} c_2 \colon \Phi \implies \Psi$ is derivable in span-apRHL, then it is valid.*

## VII. Verification Examples

We show how to use span-pRHL to verify concrete programs. Since the guarantees provided by RDP, zCDP, and tCDP can all be converted into guarantees about $(\epsilon, \delta)$-DP, one could analyze all the examples we will show under $(\epsilon, \delta)$-DP. The interest however in performing as much reasoning as possible using these relaxations is that one can achieve better values of the parameters. This will become particularly evident in the last example.

### A. One-way Marginals

As a warm up, we begin with the following classic example of a one-way marginal algorithm with additive noise.

**Algorithm 1** A mechanism estimates the attribute means

```
1: procedure AttMean(n: int, ρ: real (const.), x: bool^n (dataset), i: int, y, z, w: real)
2:     i ← 0; y ← 0;
3:     while i < n do
4:         y ← y + x[i]; i ← i + 1;
5:     z ← y/n;
6:     w ←$ Gauss(z, 1/2n²ρ);
```

We first show the Rényi-differential privacy of `AttMean`. We set a typing context $\Gamma$ of `AttMean` by $x \colon \mathtt{bool}^n$ (dataset), $i \colon \mathtt{int}$, and $y, z, w \colon \mathtt{real}$. We show the following judgment:

$$\Gamma \vdash \mathtt{AttMean} \sim^{\mathtt{RDP}}_{\alpha\rho} \mathtt{AttMean} \colon \mathtt{adj}(x\langle 1\rangle, x\langle 2\rangle) \implies w\langle 1\rangle = w\langle 2\rangle.$$

Here, the adjacent relation $\mathtt{adj}(x\langle 1\rangle, x\langle 2\rangle)$ means that two datasets $x\langle 1\rangle$ and $x\langle 2\rangle$ differs at most in one record. Explicitly, we define it by the following relation expression:

$$\bigwedge_{1 \le i \le n} \big((x[i]\langle 1\rangle \neq x[i]\langle 2\rangle) \implies \bigwedge_{1 \le j < i, i < j \le n} (x[j]\langle 1\rangle = x[j]\langle 2\rangle)\big).$$

The proof of this judgment follows by splitting `AttMean` into two commands `LoopAM; NoiseG` where $\mathtt{NoiseG} = w \xleftarrow{\$} \mathtt{Gauss}(z, 1/2n^2\rho)$, and `LoopAM` is the rest of the program. Since the loop part `LoopAM` is deterministic, by standard reasoning, we obtain:

$$\Gamma \vdash \mathtt{LoopAM} \sim^{\alpha-\mathtt{RDP}}_{0} \mathtt{LoopAM} \colon \mathtt{adj}(x\langle 1\rangle, x\langle 2\rangle) \implies (|z\langle 1\rangle - z\langle 2\rangle| \le 1/n).$$

By [RDP-G], for the noise-adding step NoiseG we have:

$$\Gamma \vdash \texttt{NoiseG} \sim_{\alpha\rho}^{\alpha-\texttt{RDP}} \texttt{NoiseG}\colon \\ (|z\langle 1\rangle - z\langle 2\rangle| \le 1/n) \implies (w\langle 1\rangle = w\langle 2\rangle).$$

Thus, by applying [seq] we complete the proof. A similar proof could have been carried out with both the rules for differential privacy, zCDP, and tCDP. Due to the simplicity of the example (that is, LoopAM is deterministic), the resulting guarantee would have been the same.

**Algorithm 2** A mechanism estimates the attribute means with SinhNormal noise

1: **procedure** AMSinh($n\colon$ int, $\rho\colon$ real (const.), $x\colon \texttt{bool}^n$ (dataset), $i\colon$ int, $y, z, w\colon$ real)
2: $i \leftarrow 0; y \leftarrow 0;$
3: **while** $i < n$ **do**
4: $y \leftarrow y + x[i];\ i \leftarrow i + 1;$
5: $z \leftarrow y/n;$
6: $w \xleftarrow{\$} w + A \cdot \texttt{arsinh}\left(\frac{1}{A}\texttt{Gauss}(0, /2n^2\rho)\right);$

We change the noise in the algorithm AttMean from Gaussian noise to SinhNormal noise. Explicitly, we define a new algorithm $\texttt{AMSinh} = \texttt{LoopAM}; \texttt{NoiseSinh}$ where the noise-adding part is changed to $\texttt{NoiseSinh} = w \xleftarrow{\$} w + A \cdot \texttt{arsinh}\left(\frac{1}{A}\texttt{Gauss}(0, /2n^2\rho)\right)$, where $A$ is a constant satisfying $1 < 1/\sqrt{\rho} \le A/n$. In the similar way as the previous example AttMean, for the loop part LoopAM, we obtain:

$$\Gamma \vdash \texttt{LoopAM} \sim_{0}^{n\cdot A/8-\texttt{tCDP}} \texttt{LoopAM}\colon \\ \texttt{adj}(x\langle 1\rangle, x\langle 2\rangle) \implies (|z\langle 1\rangle - z\langle 2\rangle| \le 1/n).$$

By applying [tCDP-SinhG], the noise-adding part NoiseSinh satisfies

$$\Gamma \vdash \texttt{NoiseSinh} \sim_{16\rho}^{n\cdot A/8-\texttt{tCDP}} \texttt{NoiseSinh}\colon \\ (|z\langle 1\rangle - z\langle 2\rangle| \le 1/n) \implies (w\langle 1\rangle = w\langle 2\rangle).$$

Thus, by applying [seq], we conclude that the algorithm AMSinh is $(16\rho, n \cdot A/8)$-tCDP.

### *B. A k-fold Gaussian mechanism*

Consider a type DATA of dataset and an predicate $\texttt{ADJ}(-,=)$ of adjacency for the type DATA, and consider $K$ queries $q(i,-)\colon \texttt{DATA} \to \texttt{real}$ $(0 \le i < K)$ with sensitivity 1, that is,

$$\texttt{ADJ}(D, D') \implies |q(i, D) - q(i, D')| \le 1.$$

We want now to prove private the following $K$-fold Gaussian mechanism. Even though standard DP can already be handled by other verification techniques, our proof applies the conversion rules between DP and zCDP along with composition in zCDP, yielding a more precise analysis for standard DP.

**Algorithm 3** Sum of $K$ Gaussian mechanisms

1: **procedure** $\texttt{FoldG}_K(K\colon$ int, $\sigma\colon$ real (const.), $D\colon$ DATA, $x, y, z\colon$ real, $i\colon$ int )
2: $i \leftarrow 0; z \leftarrow 0;$
3: **while** $i < K$ **do**
4: $x \leftarrow q(i, D); y \xleftarrow{\$} \texttt{Gauss}(0, \sigma);$
5: $z \leftarrow x + y + z; i \leftarrow i + 1;$

We set a typing context of $\texttt{FoldG}_K$ by $D\colon$ DATA, $x, y, z\colon$ real, and $i\colon$ int. Following sensitivity of queries $q$, for any $0 \le i < K$ we may assume

$$\Gamma \vdash x \leftarrow q(i, D) \sim_{0,0}^{\texttt{zCDP}} x \leftarrow q(i, D)\colon \\ \texttt{ADJ}(D\langle 1\rangle, D\langle 2\rangle) \implies |x\langle 1\rangle - x\langle 2\rangle| \le 1.$$

Thus, for the loop body $c$ (line 5), by applying [zCDP-G], [seq] and [assn], we have

$$\Gamma \vdash c \sim_{0,1/2\sigma^2}^{\texttt{zCDP}} c\colon \\ \texttt{ADJ}(D\langle 1\rangle, D\langle 2\rangle) \wedge (z\langle 1\rangle = z\langle 2\rangle) \implies z\langle 1\rangle = z\langle 2\rangle.$$

Then, by applying [assn], [seq], and [while] (the proof rule for while-loop) rules, we conclude

$$\Gamma \vdash \texttt{FoldG}_K \sim_{0,K/2\sigma^2}^{\texttt{zCDP}} \texttt{FoldG}_K\colon \\ \texttt{ADJ}(D\langle 1\rangle, D\langle 2\rangle) \implies z\langle 1\rangle = z\langle 2\rangle.$$

Hence, the algorithm $\texttt{FoldG}_K$ is $(0, K/2\sigma^2)$-zCDP. Furthermore, by applying [z/D], we conclude that the algorithm $\texttt{FoldG}_K$ is $\left(\frac{K}{2\sigma^2} + \frac{\sqrt{2K\log(1/\delta)}}{\sigma}, \delta\right)$-DP for any $0 < \delta < 1/2$.

This analysis gives a more precise bound compared to reasoning in terms of standard differential privacy. For any $0 < \delta_1 < 1/2$, the loop body $c$ satisfies

$$\Gamma \vdash c \sim_{\max((1+\sqrt{3})/2\sigma, \sqrt{2\log(0.66/\delta_1)}/\sigma), \delta_1}^{\texttt{DP}} c\colon \\ \texttt{adj}(D\langle 1\rangle, D\langle 2\rangle) \wedge (z\langle 1\rangle = z\langle 2\rangle) \implies z\langle 1\rangle = z\langle 2\rangle.$$

Let $\varepsilon = \max((1+\sqrt{3})/2\sigma, \sqrt{2\log(0.66/\delta_1)}/\sigma)$. The algorithm $\texttt{FoldG}_K$ can be seen as $K$-fold adaptive composition of the loop body $c; \cdots; c$. By applying the advanced composition theorem [26, Theorem 3.20], the algorithm $\texttt{FoldG}_K$ is $\left(\varepsilon \cdot \sqrt{2K\log(1/\delta_2)} + K\varepsilon^2, K\delta_1 + \delta_2\right)$-DP for any $0 < \delta_1, \delta_2 < 1/2$. We compare the DP-bounds which we obtained. When $\delta_2 < 0.4$, we have $2\log(0.66/\delta_2) > 1$. We also have $\varepsilon > 1.36/\sigma$ by the definition. Then, we can compute:

$$\frac{K}{2\sigma^2} + \frac{\sqrt{2K\log(1/\delta_2)}}{\sigma} \le \varepsilon \cdot \sqrt{2K\log(1/\delta_2)} + K\varepsilon^2.$$

Hence, $\varepsilon \cdot \sqrt{2K\log(1/\delta_2)} + K\varepsilon^2 > \frac{K}{2\sigma^2} + \frac{\sqrt{2K\log(1/\delta)}}{\sigma}$ whenever $\delta = K\delta_1 + \delta_2$ and $\delta_2 < 0.4$.

We can conclude that verification via zCDP is actually better than advanced composition for the algorithm FoldG. First, in the verification via zCDP, the approximation error $\delta$ is given regardless of the number of queries $K$. Second, if the approximation error satisfies $\delta < 0.4$ then the verification is significantly better than advanced composition. The restriction

$\delta < 0.4$ is quite weak since the approximation error $\delta$ in the $(\varepsilon, \delta)$-DP is thought as the probability of failure of $\varepsilon$-DP. Moreover in practical use of $(\varepsilon, \delta)$-DP, the parameter $\delta$ is usually taken to be quite small (e.g., $\delta \approx 10^{-5}$).

## VIII. Related Works

*a) Relational liftings for $f$-divergences:* Our work is inspired by work on verifying probabilistic relational properties involving $f$-divergences by [15]; we generalize their results to a broader class of divergences and also to handle continuous distributions. Barthe and Olmedo also consider $f$-divergences that satisfy a more limited version of composability, called *weak composability*. Roughly, these composition results only apply when corresponding pairs of distributions have equal weight; the KL-divergence, Hellinger distance, and $\chi^2$ divergences only satisfy this weaker version of composability. While we do not detail this extension, our framework can naturally handle weakly composable divergences in the continuous case.

A similar approach has also been used by [28] in the context of an higher order functional language for reasoning about Bayesian inference. Their type system uses a graded monad to reason about $f$-divergences. The graded monad supports only discrete distributions and is interpreted via a set-theoretic semantics, again using the lifting by [15].

*b) Relational liftings for differential privacy:* Approximate relational liftings were originally proposed for program logics targeting differential privacy. The first such system used a one-witness definition of lifting [4], which was subsequently refined to several notions of two-witness lifting [15], [29]. [18] developed approximate liftings and a program logic for continuous distribution using witness-free lifting based on a categorical monad lifting [30], [31]. A *witness-free* relational lifting for differential privacy was introduced by [18]. This can be seen as an application of the general construction of *graded relational lifting* [21, Section 5] to the Giry monad, using the technique of *codensity lifting* [31, Section 3.3] instead of $\top\top$-lifting. The witness-free relational lifting by [18] sends a binary relation $R$ between measurable spaces $X, Y$ to the following one between $\mathcal{G}X, \mathcal{G}Y$:

$$R^{\top\top(\varepsilon,\delta)} = \bigcap_{(k,l)\colon R \dot{\to} S^{(\varepsilon',\delta')}} (k^\sharp \times l^\sharp)^{-1} S^{(\varepsilon+\varepsilon',\delta+\delta')}$$

$$\text{where} \quad S^{(\varepsilon',\delta')} = \left\{ (x,y) \in \mathcal{G}1 \times \mathcal{G}1 \mid x \le e^{\varepsilon'} y + \delta' \right\}.$$

where $\mathcal{G}$ is the sub-Giry monad, $k^\sharp$ and $l^\sharp$ denote the Kleisli extensions of $k$ and $l$ respectively, $\dot{\to}$ denotes a relation-preserving map, and $\top\top$ is used to denote the codensity lifting and to distinguish it from our 2-witness lifting. Here, the intersection is taken over all measurable functions $k : X \to \mathcal{G}1, l : Y \to \mathcal{G}1$ mapping pairs related by $R$ to those related by $S^{(\varepsilon',\delta')}$. We note that the binary relation $S^{(\varepsilon',\delta')}$ is a parameter of this witness-free lifting, and by changing it, we can derive other graded relational liftings of $\mathcal{G}$.

The main difficulty with the witness-free liftings is checking whether two given distributions are related by $R^{\top\top(\varepsilon,\delta)}$: we have to test the pair $(x, y)$ against every pair $(k, l)$ of measurable functions such that $(k, l)\colon R \dot{\to} S^{(\varepsilon,\delta)}$. Fortunately, since the divergence $\Delta^{\mathtt{DP}(\varepsilon)}$ is defined by a linear inequality of measures, the witness-free lifting $R^{\top\top(\epsilon,\delta)} \subseteq \mathcal{G}X \times \mathcal{G}Y$ can be equivalently defined in a simpler form:

$$R^{\top\top(\varepsilon,\delta)} = \{\, (d_1, d_2) \mid \forall A \subseteq \Sigma_X.\ d_1(A) \le e^{\varepsilon} d_2(R(A)) + \delta \,\}.$$

While we would like to generalize this lifting construction to handle more general divergences for RDP, zCDP, and tCDP, there are at least two obstacles. First, it is not clear how to find a parameter $S$ to derive the suitable graded relational lifting for a given general divergence; this issue is currently under consideration. Second, even if we can find a suitable parameter $S$, it is awkward to work with the lifting unless we can simplify the large intersection into a more convenient form. In contrast, 2-witness liftings seem more concrete and easier to work with: it suffices to give witness distributions to check the membership of lifted relations.

In the discrete case, witness-free liftings are equivalent to the witness-/span-based liftings by [32]. Recent work also considers liftings with more fine-grained parameters that can vary over different pairs of samples [2].

*c) Other techniques for verifying privacy:* Rényi and zero-concentrated differential privacy were recently proposed in the differential privacy literature; to the best of our knowledge, we are the first to verify these properties. In contrast, there are now numerous systems targeting differential privacy using a wide range of techniques beyond program logics, including dynamic analyses [6], linear [7], [5], [33] and dependent [3] type systems, product programs [34], partial evaluation [8], and constraint-solving [9], [2]; see the recent survey [10] for more details.

## IX. Conclusion and Future Work

We have developed a framework for reasoning about three relaxations of differential privacy: Rényi differential privacy, zero concentrated differential privacy, and truncated concentrated differential privacy. We extended the notion of divergences to a more general class, and to support subprobability measures. Additionally, we have introduced a novel notion of approximate span-lifting supporting these divergences and continuous distributions.

One promising direction for future work is to study the moment-accountant composition method [35]. This composition method tracks the moments of the privacy loss random variable, although it does not directly correspond to composition for RDP or zCDP. Another interesting direction would be to analyze recently-proposed RDP mechanisms for posterior sampling [36], and the GAP-Max tCDP algorithm [13].

## References


[1] C. Dwork, F. McSherry, K. Nissim, and A. D. Smith, "Calibrating noise to sensitivity in private data analysis," in *IACR Theory of Cryptography Conference (TCC), New York, New York*, ser. Lecture Notes in Computer Science. Springer-Verlag, 2006, vol. 3876, pp. 265–284.

[2] A. Albarghouthi and J. Hsu, "Synthesizing coupling proofs of differential privacy," *Proceedings of the ACM on Programming Languages*, vol. 2, no. POPL, Jan. 2018, appeared at ACM SIGPLAN–SIGACT Symposium on Principles of Programming Languages (POPL), Los Angeles, California. [Online]. Available: https://arxiv.org/abs/1709.05361

[3] G. Barthe, M. Gaboardi, E. J. Gallego Arias, J. Hsu, A. Roth, and P.-Y. Strub, "Higher-order approximate relational refinement types for mechanism design and differential privacy," in *ACM SIGPLAN–SIGACT Symposium on Principles of Programming Languages (POPL), Mumbai, India*, 2015, pp. 55–68. [Online]. Available: https://arxiv.org/abs/1407.6845

[4] G. Barthe, B. Köpf, F. Olmedo, and S. Zanella-Béguelin, "Probabilistic relational reasoning for differential privacy," *ACM Transactions on Programming Languages and Systems*, vol. 35, no. 3, pp. 9:1–9:49, Nov. 2013. [Online]. Available: https://software.imdea.org/~bkoepf/papers/toplas13.pdf

[5] M. Gaboardi, A. Haeberlen, J. Hsu, A. Narayan, and B. C. Pierce, "Linear dependent types for differential privacy," in *ACM SIGPLAN–SIGACT Symposium on Principles of Programming Languages (POPL), Rome, Italy*, 2013, pp. 357–370. [Online]. Available: https://dl.acm.org/citation.cfm?id=2429113

[6] F. McSherry, "Privacy integrated queries," in *ACM SIGMOD International Conference on Management of Data (SIGMOD), Providence, Rhode Island*, 2009, pp. 19–30. [Online]. Available: https://research.microsoft.com/pubs/80218/sigmod115-mcsherry.pdf

[7] J. Reed and B. C. Pierce, "Distance makes the types grow stronger: A calculus for differential privacy," in *ACM SIGPLAN International Conference on Functional Programming (ICFP), Baltimore, Maryland*, 2010, pp. 157–168. [Online]. Available: http://dl.acm.org/citation.cfm?id=1863568

[8] D. Winograd-Cort, A. Haeberlen, A. Roth, and B. C. Pierce, "A framework for adaptive differential privacy," *Proceedings of the ACM on Programming Languages*, vol. 1, no. ICFP, pp. 10:1–10:29, 2017. [Online]. Available: http://doi.acm.org/10.1145/3110254

[9] D. Zhang and D. Kifer, "LightDP: Towards automating differential privacy proofs," in *ACM SIGPLAN–SIGACT Symposium on Principles of Programming Languages (POPL), Paris, France*, 2017, pp. 888–901. [Online]. Available: https://arxiv.org/abs/1607.08228

[10] G. Barthe, M. Gaboardi, J. Hsu, and B. C. Pierce, "Programming language techniques for differential privacy," *ACM SIGLOG News*, vol. 3, no. 1, pp. 34–53, Jan. 2016. [Online]. Available: http://siglog.hosting.acm.org/wp-content/uploads/2016/01/siglog_news_7.pdf

[11] I. Mironov, "Rényi differential privacy," in *IEEE Computer Security Foundations Symposium (CSF), Santa Barbara, California*, 2017, pp. 263–275. [Online]. Available: https://arxiv.org/abs/1702.07476

[12] M. Bun and T. Steinke, "Concentrated differential privacy: Simplifications, extensions, and lower bounds," in *IACR Theory of Cryptography Conference (TCC), Beijing, China*, ser. Lecture Notes in Computer Science, vol. 9985. Springer-Verlag, 2016, pp. 635–658.

[13] M. Bun, C. Dwork, G. N. Rothblum, and T. Steinke, "Composable and versatile privacy via truncated CDP," in *ACM SIGACT Symposium on Theory of Computing (STOC), Los Angeles, California*, 2018.

[14] A. Renyi, "On measures of entropy and information," in *Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Contributions to the Theory of Statistics*. Berkeley, Calif.: University of California Press, 1961, pp. 547–561. [Online]. Available: http://projecteuclid.org:443/euclid.bsmsp/1200512181

[15] G. Barthe and F. Olmedo, "Beyond differential privacy: Composition theorems and relational logic for $f$-divergences between probabilistic programs," in *International Colloquium on Automata, Languages and Programming (ICALP), Riga, Latvia*, ser. Lecture Notes in Computer Science, vol. 7966. Springer-Verlag, 2013, pp. 49–60. [Online]. Available: https://certicrypt.gforge.inria.fr/2013.ICALP.pdf

[16] F. Olmedo, "Approximate relational reasoning for probabilistic programs," Ph.D. dissertation, Technical University of Madrid, 2014.

[17] T. Van Erven and P. Harremoës, "Rényi divergence and Kullback-Leibler divergence," *IEEE Transactions on Information Theory*, vol. 60, no. 7, pp. 3797–3820, July 2014.

[18] T. Sato, "Approximate relational Hoare logic for continuous random samplings," *Electronic Notes in Theoretical Computer Science*, vol. 325, pp. 277–298, 2016, conference on the Mathematical Foundations of Programming Semantics (MFPS), Pittsburgh, Pennsylvania. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1571066116300949

[19] S. Meiser, "Approximate and probabilistic differential privacy definitions," *IACR Cryptology ePrint Archive*, vol. 2018, p. 277, 2018. [Online]. Available: https://eprint.iacr.org/2018/277

[20] F. Liese and I. Vajda, "On divergences and informations in statistics and information theory," *IEEE Transactions on Information Theory*, vol. 52, no. 10, pp. 4394–4412, Oct 2006.

[21] S.-y. Katsumata, "Parametric effect monads and semantics of effect systems," in *ACM SIGPLAN–SIGACT Symposium on Principles of Programming Languages (POPL), San Diego, California*, 2014, pp. 633–645. [Online]. Available: http://doi.acm.org/10.1145/2535838.2535846

[22] S. Fujii, S. Katsumata, and P. Melliès, "Towards a formal theory of graded monads," in *Foundations of Software Science and Computation Structures - 19th International Conference, FOSSACS 2016, Held as Part of the European Joint Conferences on Theory and Practice of Software, ETAPS 2016, Eindhoven, The Netherlands, April 2-8, 2016, Proceedings*, 2016, pp. 513–530. [Online]. Available: https://doi.org/10.1007/978-3-662-49630-5_30

[23] W. Rudin, *Real and complex analysis*, 3rd ed. New York: McGraw-Hill Book Co., 1987.

[24] M. Giry, "A categorical approach to probability theory," in *Categorical Aspects of Topology and Analysis*, ser. Lecture Notes in Mathematics, B. Banaschewski, Ed. Springer-Verlag, 1982, vol. 915, pp. 68–85. [Online]. Available: http://dx.doi.org/10.1007/BFb0092872

[25] M. C. Pardo and I. Vajda, "About distances of discrete distributions satisfying the data processing theorem of information theory," *IEEE Transactions on Information Theory*, vol. 43, no. 4, pp. 1288–1293, Jul 1997.

[26] C. Dwork and A. Roth, "The algorithmic foundations of differential privacy," *Foundations and Trends® in Theoretical Computer Science*, vol. 9, no. 3–4, 2013. [Online]. Available: http://dx.doi.org/10.1561/0400000042

[27] A. Langlois, D. Stehlé, and R. Steinfeld, "Gghlite: More efficient multilinear maps from ideal lattices," in *Advances in Cryptology – EUROCRYPT 2014*, P. Q. Nguyen and E. Oswald, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2014, pp. 239–256.

[28] G. Barthe, G. P. Farina, M. Gaboardi, E. J. G. Arias, A. Gordon, J. Hsu, and P. Strub, "Differentially private bayesian programming," in *ACM SIGSAC Conference on Computer and Communications Security (CCS), Vienna, Austria*, 2016, pp. 68–79. [Online]. Available: http://doi.acm.org/10.1145/2976749.2978371

[29] G. Barthe, N. Fong, M. Gaboardi, B. Grégoire, J. Hsu, and P.-Y. Strub, "Advanced probabilistic couplings for differential privacy," in *ACM SIGSAC Conference on Computer and Communications Security (CCS), Vienna, Austria*, 2016, pp. 55–67. [Online]. Available: https://arxiv.org/abs/1606.07143

[30] S.-y. Katsumata, "A semantic formulation of TT-lifting and logical predicates for computational metalanguage," in *International Workshop on Computer Science Logic (CSL), Oxford, England*, ser. Lecture Notes in Computer Science, L. Ong, Ed. Springer-Verlag, 2005, vol. 3634, pp. 87–102. [Online]. Available: http://dx.doi.org/10.1007/11538363_8

[31] S.-y. Katsumata and T. Sato, "Codensity liftings of monads," in *6th Conference on Algebra and Coalgebra in Computer Science (CALCO 2015)*, ser. Leibniz International Proceedings in Informatics, vol. 35. Schloss Dagstuhl–Leibniz Center for Informatics, 2015, pp. 156–170. [Online]. Available: https://drops.dagstuhl.de/opus/volltexte/2015/5532

[32] G. Barthe, T. Espitau, J. Hsu, T. Sato, and P.-Y. Strub, "$\star$-liftings for differential privacy," in *International Colloquium on Automata, Languages and Programming (ICALP), Warsaw, Poland*, ser. Leibniz International Proceedings in Informatics, vol. 80. Schloss Dagstuhl–Leibniz Center for Informatics, 2017, pp. 102:1–102:12. [Online]. Available: http://drops.dagstuhl.de/opus/volltexte/2017/7435

[33] A. Azevedo de Amorim, M. Gaboardi, E. J. Gallego Arias, and J. Hsu, "Really natural linear indexed type-checking," in *Implementation of Functional Languages (IFL), Boston, Massachusetts*. ACM Press, 2014, pp. 5:1–5:12. [Online]. Available: http://arxiv.org/abs/1503.04522

[34] G. Barthe, M. Gaboardi, E. J. Gallego Arias, J. Hsu, C. Kunz, and P.-Y. Strub, "Proving differential privacy in Hoare logic," in *IEEE Computer Security Foundations Symposium (CSF), Vienna, Austria*, 2014, pp. 411–424. [Online]. Available: https://arxiv.org/abs/1407.2988

[35] M. Abadi, A. Chu, I. J. Goodfellow, H. B. McMahan, I. Mironov, K. Talwar, and L. Zhang, "Deep learning with differential privacy," in *ACM SIGSAC Conference on Computer and Communications Security*

*(CCS), Vienna, Austria*, 2016, pp. 308–318. [Online]. Available: http://doi.acm.org/10.1145/2976749.2978318

[36] J. Geumlek, S. Song, and K. Chaudhuri, "Renyi differential privacy mechanisms for posterior sampling," in *Conference on Neural Information Processing Systems (NIPS), Long Beach, California*, 2017, pp. 5295–5304. [Online]. Available: http://arxiv.org/abs/1710.00892