DISTANCES BETWEEN STATES AND BETWEEN PREDICATES

BART JACOBS AND ABRAHAM WESTERBAAN

Institute for Computing and Information Sciences, Radboud University Nijmegen, The Netherlands.
November 28, 2017
e-mail address: bart@cs.ru.nl, bram@westerbaan.name

Abstract. This paper gives a systematic account of various metrics on probability distributions (states) and on predicates. These metrics are described in a uniform manner using the validity relation between states and predicates. The standard adjunction between convex sets (of states) and effect modules (of predicates) is restricted to convex complete metric spaces and directed complete effect modules. This adjunction is used in two state-and-effect triangles, for classical (discrete) probability and for quantum probability.

1. Introduction

Metric structures have a long history in program semantics, see the overview book [2]. They occur naturally, for instance on sequences, of inputs, outputs, or states. In complete metric spaces solutions of recursive (suitably contractive) equations exist via Banach's fixed point theorem. The Hausdorff distance on subsets is used to model non-deterministic (possibilistic) computation.

This paper looks at metrics on probability distributions, often called states. It covers standard distance functions on *classical* discrete probability distributions and on *quantum* distributions. For discrete probability we use the total variation distance, which is a special case of the Kantorovich distance, see *e.g.* [10, 4, 26, 3, 25]. For quantum probability we use the trace distance for states (quantum distributions) on Hilbert spaces, and the operator norm distance for states on von Neumann algebras. One contribution of this paper is a uniform description of all these distances on states as 'validity' distances.

In each of these cases we shall describe a validity relation $\models$ between states ω and predicates p, so that the validity $\omega \models p$ is a number in the unit interval $[0, 1]$. What we call

1998 ACM Subject Classification: F.1.1 Models of Computation.

Key words and phrases: Total variation distance, Kantorovich distance, trace distance, convex set, effect module, state and effect triangle.

The research leading to these results has received funding from the European Research Council under the European Union's Seventh Framework Programme (FP7/2007-2013) / ERC grant agreement nr. 320571.

LOGICAL METHODS
IN COMPUTER SCIENCE

DOI:10.2168/LMCS-???

the 'validity' distance on states is given by the supremum (join) over predicates p in:

$$d(\omega_1, \omega_2) = \bigvee\nolimits_p \big|\, \omega_1 \models p - \omega_2 \models p \,\big|. \tag{1.1}$$

In general, states are closed under convex combinations. We shall thus study combinations of convex and complete metric spaces, in a category **ConvCMet**.

We also study metrics on predicates. The algebraic structure of predicates will be described in terms of effect modules. Here we show that suitably order complete effect modules are Archimedean, and thus carry an induced metric, such that limits and joins of ascending Cauchy sequences coincide. In our main examples, we use fuzzy predicates on sets and effects of von Neumann algebras as predicates; their distance can also be formulated via validity, but now using a join over states ω in:

$$d(p_1, p_2) = \bigvee\nolimits_\omega \big|\, \omega \models p_1 - \omega \models p_2 \,\big|. \tag{1.2}$$

The 'duality' between the distance formulas (1.1) for states and (1.2) for predicates is a new insight.

In a next step this paper restricts the standard adjunction $\mathbf{EMod}^{\mathrm{op}} \rightleftarrows \mathbf{Conv}$ between effect modules and convex sets to an adjunction $\mathbf{DcEMod}^{\mathrm{op}} \rightleftarrows \mathbf{ConvCMet}$ between *directed complete* effect modules and convex *complete metric* spaces. This restriced adjunction is used in two 'state-and-effect' triangles, of the form:

$$\begin{tikzcd}
\mathbf{DcEMod}^{\mathrm{op}} \arrow[rr, bend left] \arrow[rr, phantom, "\top"] & & \mathbf{ConvCMet} \arrow[ll, bend left] \\
& \mathcal{K}\ell_{\mathrm{fin}}(\mathcal{D}) \arrow[ul, "\mathrm{Pred}"] \arrow[ur, "\mathrm{Stat}"'] &
\end{tikzcd}
\qquad
\begin{tikzcd}
\mathbf{DcEMod}^{\mathrm{op}} \arrow[rr, bend left] \arrow[rr, phantom, "\top"] & & \mathbf{ConvCMet} \arrow[ll, bend left] \\
& \mathbf{vNA}^{\mathrm{op}} \arrow[ul, "\mathrm{Pred}"] \arrow[ur, "\mathrm{Stat}"'] &
\end{tikzcd}$$

Details will be provided in Section 4. Thus, the paper culminates in suitable order/-metrically complete versions of the state-and-effect triangles that emerge in the effectus-theoretic [13, 7] description of state and predicate transformer semantics for probability (see also [15, 17]).

2. Distances between states

This section will describe distance functions (metrics) on various forms of probability distributions, which we collectively call 'states'. In separate subsections it will introduce discrete probability distributions on sets and on metric spaces, and quantum distributions on Hilbert spaces and on von Neumann algebras. A unifying formulation will be identified, namely what we call a *validity* formulation of the metrics involved, where the distance between two states is expressed via a join over all predicates using the validities of these predicates in the two states, as in (1.1).

2.1. **Discrete probability distributions on sets.**

A finite discrete probability distribution on a set X is given by 'probability mass' function $\omega\colon X \to [0,1]$ with finite support and $\sum_x \omega(x) = 1$. This support $\mathrm{supp}(\omega) \subseteq X$ is the set $\{x \in X \mid \omega(x) \neq 0\}$. We sometimes simply say 'distribution' instead of 'finite discrete probability distribution'. Often such a distribution is called a 'state'. The 'ket' notation $|-\rangle$ is useful to describe specific distributions. For instance, on a set $X = \{a, b, c, d\}$ we may write a distribution as $\omega = \frac{1}{2}|a\rangle + \frac{1}{8}|b\rangle + \frac{3}{8}|c\rangle$. This corresponds to the probability mass function $\omega\colon X \to [0,1]$ given by $\omega(a) = \frac{1}{2}$, $\omega(b) = \frac{1}{8}$ and $\omega(c) = \frac{3}{8}$.

We write $\mathcal{D}(X)$ for the set of distributions on a set X. The mapping $X \mapsto \mathcal{D}(X)$ forms (part of) a well-known monad on the category of sets, see *e.g.* [12, 14, 15]. We write $\mathcal{K}\ell(\mathcal{D})$ for the associated Kleisli category, and $\mathcal{EM}(\mathcal{D})$ for the category of Eilenberg-Moore algebras. The latter may be identified with *convex sets*, that is, with sets in which formal convex sums can be interpreted as actual sums. Thus we often write $\mathbf{Conv} = \mathcal{EM}(\mathcal{D})$; morphisms in **Conv** are 'affine' functions, that preserve convex sums.

Definition 1. Let $\omega_1, \omega_2 \in \mathcal{D}(X)$ be two distributions on the same set X. Their *total variation distance* $\mathsf{tvd}(\omega_1, \omega_2)$ is the positive real number defined as:

$$\mathsf{tvd}(\omega_1, \omega_2) = \tfrac{1}{2}\sum_{x\in X} \big|\, \omega_1(x) - \omega_2(x)\, \big|. \tag{2.1}$$

The historical origin of this definition is not precisely clear. It is folklore that the total variation distance is a special case of the 'Kantorovich distance' (also known as 'Wasserstein' or 'earth mover's distance') on distributions on metric spaces, when applied to discrete metric spaces (sets), see Subsection 2.2 below.

We leave it to the reader to verify that tvd is a metric on sets of distributions $\mathcal{D}(X)$, and that its values are in the unit interval $[0,1]$.

Example 2. Consider the sets $X = \{a, b\}$ and $Y = \{0, 1\}$ with 'joint' distribution $\omega \in \mathcal{D}(X \times Y)$ given by $\omega = \frac{1}{2}|a, 0\rangle + \frac{1}{2}|b, 1\rangle$. The first and second marginal of ω, written as $\omega_1 \in \mathcal{D}(X)$ and $\omega_2 \in \mathcal{D}(Y)$, are: $\omega_1 = \frac{1}{2}|a\rangle + \frac{1}{2}|b\rangle$ and $\omega_1 = \frac{1}{2}|0\rangle + \frac{1}{2}|1\rangle$. We immediately see that ω is not the same as the product $\omega_1 \otimes \omega_2 \in \mathcal{D}(X \times Y)$ of its marginals, since $\omega_1 \otimes \omega_2 = \frac{1}{4}|a, 0\rangle + \frac{1}{4}|a, 1\rangle + \frac{1}{4}|b, 0\rangle + \frac{1}{4}|b, 1\rangle$. This means ω is 'entwined', see [21, 15]. One way to associated a number with this entwinedness is to take the distance between ω and the product of its marginals. It can be computed as:

$$\begin{aligned}\mathsf{tvd}\big(\omega,\, \omega_1 \otimes \omega_2\big) &= \tfrac{1}{2}\sum_{x\in X, y\in Y} \big|\, \omega(x,y) - (\omega_1 \otimes \omega_2)(x,y)\, \big| \\ &= \tfrac{1}{2}\Big(|\tfrac{1}{2} - \tfrac{1}{4}| + |0 - \tfrac{1}{4}| + |0 - \tfrac{1}{4}| + |\tfrac{1}{2} - \tfrac{1}{4}|\Big) \;=\; \tfrac{1}{2}.\end{aligned}$$

For a function $f\colon X \to \mathcal{D}(Y)$ there are two associated 'transformation' functions, namely state transformation (aka. Kleisli extension) $f_*\colon \mathcal{D}(X) \to \mathcal{D}(Y)$ and predicate transformation $f^*\colon [0,1]^Y \to [0,1]^X$. They are defined as:

$$f_*(\omega)(y) = \textstyle\sum_x f(x)(y)\cdot\omega(x) \qquad\text{and}\qquad f^*(q)(x) = \textstyle\sum_y f(x)(y)\cdot q(y). \tag{2.2}$$

Maps $p \in [0,1]^X$ are called (fuzzy) predicates on X. In the special case where the outcomes $p(x)$ are in the (discrete) subset $\{0,1\} \subseteq [0,1]$, the predicate p is called *sharp*. These sharp predicates correspond to subsets $U \subseteq X$, via the indicator function $\mathbf{1}_U\colon X \to \{0,1\}$.

For a state $\omega \in \mathcal{D}(X)$ we write $\omega \models p$ for the *validity* of predicate p in state ω, defined as the expected value $\sum_x \omega(x)\cdot p(x)$ in $[0,1]$. Thus, $\omega \models \mathbf{1}_U = \sum_{x\in U}\omega(x)$; the latter sum is commonly written as $\omega(U)$. Further, the fundamental validity transformation equality holds: $f_*(\omega) \models q \;=\; \omega \models f^*(q)$.

We conclude this subsection with a standard redescription of the total variation distance, see *e.g.* [10, 29]. It uses validity $\models$, as described above. Such 'validity' based distances will form an important theme in this paper. The proof of the next result is standard but not trivial and is included in the appendix, for the convenience of the reader.

Proposition 3. Let X be an arbitrary set, with states $\omega_1, \omega_2 \in \mathcal{D}(X)$. Then:

$$\mathsf{tvd}\big(\omega_1, \omega_2\big) = \bigvee_{p \in [0,1]^X} \Big| \omega_1 \models p - \omega_2 \models p \Big| = \max_{U \subseteq X} \omega_1 \models \mathbf{1}_U - \omega_2 \models \mathbf{1}_U \qquad \square$$

We write maximum 'max' instead of join $\bigvee$ to express that the supremum is actually reached by a subset (sharp predicate).

Lemma 4. If X is a finite set, then $\mathcal{D}(X)$ is a complete metric space.

Proof Let $X = \{x_1, \ldots, x_N\}$ and $\omega_i \in \mathcal{D}(X)$ be a Cauchy sequence. For each n we have $\big|\omega_i(x_n) - \omega_j(x_n)\big| \leq 2 \cdot \mathsf{tvd}(\omega_i, \omega_j)$. Hence, the sequence $\omega_i(x_n) \in [0, 1]$ is Cauchy too, say with limit r_n. Take $\omega = \sum_n r_n | x_n \rangle \in \mathcal{D}(X)$. This is the limit of the ω_i. $\square$

2.2. **Discrete probability distributions on metric spaces.**

A metric d on a set X is called 1-bounded if it takes values in the unit interval $[0, 1]$, that is, if it has type $d\colon X \times X \to [0, 1]$. We write **Met** for the category with such 1-bounded metric spaces as objects, and with non-expansive functions f between them, satisfying $d(f(x), f(y)) \leq d(x, y)$. From now on we assume that all metric spaces in this paper are 1-bounded. For example, each set carries a discrete metric, where points x, y have distance 0 if they are equal, and 1 otherwise.

For a metric space X and two functions $f, g\colon A \to X$ from some set A to X there is the *supremum* distance given by:

$$\mathsf{spd}(f, g) = \bigvee_{a \in A} d\big(f(a), g(a)\big). \tag{2.3}$$

A 'metric predicate' on a metric space X is a non-expansive function $p\colon X \to [0, 1]$. These predicates carry the above supremum distance spd. We use them in the following definition of Kantorovich distance, which transfers the validity description of Proposition 3 to the metric setting.

Definition 5. Let ω_1, ω_2 be two discrete distributions on (the underlying set of) a metric space X. The *Kantorovich* distance between them is defined as:

$$\mathsf{kvd}(\omega_1, \omega_2) = \bigvee_{p \in \mathbf{Met}(X, [0,1])} \Big| \omega_1 \models p - \omega_2 \models p \Big|. \tag{2.4}$$

This makes $\mathcal{D}(X)$ a (1-bounded) metric space.

The Kantorovich-Wasserstein duality Theorem gives an equivalent description of this distance in terms of joint states and 'couplings', see [24, 29] for details. Here we concentrate on relating the Kantorovich distance to the monad structure of distributions.

Lemma 6. Let X, Y be metric spaces.

(1) The unit function $\eta\colon X \to \mathcal{D}(X)$ given by $\eta(x) = 1|x\rangle$ is non-expansive.
(2) For each non-expansive function $f\colon X \to \mathcal{D}(Y)$ the corresponding state transformer $f_*\colon \mathcal{D}(X) \to \mathcal{D}(Y)$ from (2.2) is non-expansive.

As special cases, the multiplication map $\mu = (\mathrm{id})_*\colon \mathcal{D}(\mathcal{D}(X)) \to \mathcal{D}(X)$ of the monad $\mathcal{D}$ is non-expansive, and validity $(-) \models p = p_*\colon \mathcal{D}(X) \to \mathcal{D}(2) = [0, 1]$ in its first argument as well.

(3) If $f\colon X \to \mathcal{D}(Y)$ and $q\colon Y \to [0,1]$ are non-expansive, then so is $f^*(q)\colon X \to [0,1]$. Moreover, the function $f^*\colon \mathbf{Met}(Y,[0,1]) \to \mathbf{Met}(X,[0,1])$ is itself non-expansive, wrt. the supremum distance (2.3).

As a result, validity $\omega \models (-) = \omega^*\colon \mathbf{Met}(X,[0,1]) \to \mathbf{Met}(1,[0,1]) = [0,1]$ is non-expansive in its second argument too.

(4) Taking convex combinations of distributions σ_i, τ_i satisfies: for $r+s=1$,

$$\mathsf{kvd}\big(r\cdot\sigma_1 + s\cdot\sigma_2,\ r\cdot\tau_1 + s\cdot\tau_2\big) \leq r\cdot\mathsf{kvd}(\sigma_1,\tau_1) + s\cdot\mathsf{kvd}(\sigma_2,\tau_2).$$

Proof We do points (1) and (4) and leave the others to the reader. The crucial point that we use to show for (1) is that the unit map $\eta\colon X \to \mathcal{D}(X)$ is non-expansive is: $\eta(x) \models p = p(x)$. Hence we are done because the join in (2.4) is over non-expansive functions p in:

$$\begin{aligned}\mathsf{kvd}(\eta(x_1),\eta(x_2)) \;=\; \bigvee\nolimits_p \Big|\eta(x_1)\models p - \eta(x_2)\models p\Big| &= \bigvee\nolimits_p \Big|p(x_1)-p(x_2)\Big| \\ &\leq \bigvee\nolimits_p d(x_1,x_2) \\ &= d(x_1,x_2).\end{aligned}$$

For point (4) we first notice that for $\Omega \in \mathcal{D}^2(X)$ and $p\colon X \to [0,1]$,

$$\begin{aligned}\mu(\Omega)\models p \;=\; \textstyle\sum_x \mu(\Omega)(x)\cdot p(x) &= \textstyle\sum_x \big(\sum_\omega \Omega(\omega)\cdot\omega(x)\big)\cdot p(x) \\ &= \textstyle\sum_\omega \Omega(\omega)\cdot\big(\sum_x \omega(x)\cdot p(x)\big) \\ &= \textstyle\sum_\omega \Omega(\omega)\cdot\big(\omega\models p\big) \\ &= \Omega \models \big((-)\models p\big),\end{aligned}$$

where $(-)\models p\colon \mathcal{D}(X) \to [0,1]$ is used as (non-expansive) predicate on $\mathcal{D}(X)$. Hence for $r,s \in [0,1]$ with $r+s=1$,

$$\begin{aligned}&\mathsf{kvd}\big(r\cdot\sigma_1 + s\cdot\sigma_2,\ r\cdot\tau_1 + s\cdot\tau_2\big) \\ &= \mathsf{kvd}\big(\mu(r|\sigma_1\rangle + s|\sigma_2\rangle),\ \mu(r|\tau_1\rangle + s|\tau_2\rangle)\big) \\ &= \bigvee\nolimits_p \big|\,\mu(r|\sigma_1\rangle + s|\sigma_2\rangle)\models p - \mu(r|\tau_1\rangle + s|\tau_2\rangle)\models p\,\big| \\ &= \bigvee\nolimits_p \big|\,r|\sigma_1\rangle + s|\sigma_2\rangle \models \big((-)\models p\big) - r|\tau_1\rangle + s|\tau_2\rangle \models \big((-)\models p\big)\,\big| \\ &= \bigvee\nolimits_p \big|\,r\cdot(\sigma_1\models p) + s\cdot(\sigma_2\models p) - r\cdot(\tau_1\models p) + s\cdot(\tau_2\models p)\,\big| \\ &\leq \bigvee\nolimits_p r\cdot\big|\,\sigma_1\models p - \tau_1\models p\big| + \bigvee\nolimits_p s\cdot\big|\,\sigma_2\models p - \tau_2\models p\,\big| \\ &= r\cdot\mathsf{kvd}(\sigma_1,\tau_1) + s\cdot\mathsf{kvd}(\sigma_2,\tau_2).\end{aligned}$$

□

Corollary 7. The monad $\mathcal{D}$ on **Sets** lifts to a monad, also written as $\mathcal{D}$, on the category **Met**, and commutes with forgetful functors, as in:

$$\begin{tikzcd}\mathbf{Met} \arrow[r, "\mathcal{D}"] \arrow[d] & \mathbf{Met} \arrow[d] \\ \mathbf{Sets} \arrow[r, "\mathcal{D}"] & \mathbf{Sets}\end{tikzcd} \tag{2.5}$$

We write **ConvMet** for the category $\mathcal{EM}(\mathcal{D})$ of Eilenberg-Moore algebras of this lifted monad, with 'convex metric spaces' as objects, see below.

The lifting (2.5) can be seen as a finite version of a similar lifting result for the 'Kantorovich' functor $\mathcal{K}$ in [4]. This $\mathcal{K}(X)$ captures the tight Borel probability measures on a metric space X. The above lifting (2.5) is a special case of the generic lifting of functors on sets to functors on metric spaces described in [3] (see esp. Example 3.3).

The category **ConvMet** $= \mathcal{EM}(\mathcal{D})$ of the monad $\mathcal{D}\colon \mathbf{Met} \to \mathbf{Met}$ contains *convex metric spaces*, consisting of:

(1) a convex set X, that is, a set X with an Eilenberg-Moore algebra $\alpha\colon \mathcal{D}(X) \to X$ of the distribution monad $\mathcal{D}$ on **Sets**;
(2) a metric $d_X\colon X \times X \to [0,1]$ on X;
(3) a connection between the convex and the metric structure, via the requirement that the algebra map $\alpha\colon \mathcal{D}(X) \to X$ is non-expansive: $d_X(\alpha(\omega_1), \alpha(\omega_2)) \le \mathsf{kvd}(\omega_1, \omega_2)$, for all distributions $\omega_1, \omega_2 \in \mathcal{D}(X)$.

The maps in **ConvMet** are both affine and non-expansive. We shall write **ConvCMet** $\hookrightarrow$ **ConvMet** for the full subcategory of convex *complete* metric spaces.

Example 8. The unit interval $[0,1]$ is a convex metric space, via its standard (Euclidean) metric, and its standard convex structure, given by the algebra map $\alpha\colon \mathcal{D}([0,1]) \to [0,1]$ defined by the 'expected value' operation:

$$\alpha(\omega) = \textstyle\sum_{x\in\mathbb{R}} \omega(x) \cdot x \qquad \text{that is} \qquad \alpha\big(\sum_i r_i|x_i\rangle\big) = \sum_i r_i \cdot x_i.$$

The identity map $\mathrm{id}\colon [0,1] \to [0,1]$ is a predicate on $[0,1]$ that satisfies:

$$\omega \models \mathrm{id} = \textstyle\sum_x \omega(x) \cdot \mathrm{id}(x) = \sum_x \omega(x) \cdot x = \alpha(\omega).$$

This allows us to show that α is non-expansive:

$$\begin{aligned} \big|\,\alpha(\omega_1) - \alpha(\omega_2)\,\big| &= \big|\,\omega_1 \models \mathrm{id} - \omega_2 \models \mathrm{id}\,\big| \\ &\le \textstyle\bigvee_p \big|\,\omega_1 \models p - \omega_2 \models p\,\big| \;=\; \mathsf{kvd}(\omega_1, \omega_2). \end{aligned}$$

In fact, we can see this as a special case of non-expansiveness of multiplication maps μ from Lemma 6 (2): indeed, $\mathcal{D}(2) \cong [0,1]$, for the two-element set $2 = \{0,1\}$, and the algebra $\alpha\colon \mathcal{D}([0,1]) \to [0,1]$ corresponds to the multiplication $\mu\colon \mathcal{D}(\mathcal{D}(2)) \to \mathcal{D}(2)$.

2.3. **Density matrices on Hilbert spaces.**

The analogue of a probability distribution in quantum theory is often simply called a state. We first consider states of Hilbert spaces, and consider the more general (and abstract) situation of states on von Neumann algebras in the next subsection.

A state of a Hilbert space $\mathscr{H}$ is a density operator, that is, it is a positive linear map $\varrho\colon \mathscr{H} \to \mathscr{H}$ whose trace is one: $\mathrm{tr}(\varrho) = 1$. Recall that the trace of a positive operator $T\colon \mathscr{H} \to \mathscr{H}$ is given by $\mathrm{tr}(T) = \sum \langle e_i, T(e_i)\rangle$, where $(e_i)_i$ is any orthonormal basis for $\mathscr{H}$; this value $\mathrm{tr}(T)$ does not depend on the choice of basis $(e_i)_i$, but might equal $+\infty$ [1, Def. 2.51]. The same formula also works for when T is not necessarily positive, but bounded with $\mathrm{tr}(|T|) < \infty$ — where $|T| \coloneqq \sqrt{T^\dagger T}$ and $T^\dagger$ is the adjoint of T. Such T, which are aptly called *trace-class operators*, always have finite trace: $\mathrm{tr}(T) < \infty$, see [1, Def. 2.5{4,6}]. When $\mathscr{H}$ is finite dimensional, any operator $T\colon \mathscr{H} \to \mathscr{H}$ is trace-class, and when represented as a matrix, its trace can be computed as the sum of all elements on the diagonal. If T is a density operator, then the associated matrix is called a density matrix. We refer for more information to for instance [1], and to [27, 28, 30] for the finite-dimensional case.

A predicate on $\mathscr{H}$ is a linear map $p\colon \mathscr{H} \to \mathscr{H}$ with $0 \le p \le \mathrm{id}$. It is called *sharp* (or a *projection*) if $p^2 = p$. Predicates are also called *effects*. We write $\mathcal{E}f(\mathscr{H})$ for the set of effects of $\mathscr{H}$. For a state ϱ of $\mathscr{H}$ the validity $\varrho \models p$ is defined as the trace $\mathrm{tr}(\varrho\, p)$. To make sense of this definition we should mention that the product AB of bounded operators $A, B\colon \mathscr{H} \to \mathscr{H}$ is trace-class when either A or B is trace-class [1, Def. 2.54] — so $\varrho\, p$ is trace-class because ϱ is.

Definition 9. Let ϱ_1, ϱ_2 be two quantum states of the same Hilbert space. The trace distance $\mathsf{trd}(\varrho_1, \varrho_2)$ between them is defined as:

$$\mathsf{trd}(\varrho_1, \varrho_2) = \tfrac{1}{2}\mathrm{tr}\big(\big|\, \varrho_1 - \varrho_2 \,\big|\big) = \tfrac{1}{2}\mathrm{tr}\Big(\sqrt{(\varrho_1 - \varrho_2)^\dagger(\varrho_1 - \varrho_2)}\Big). \tag{2.6}$$

This definition involves the square root of a positive operator B. With the following examples in mind it is worth pointing out that in the finite-dimensional case — when B is essentially a positive matrix — the square root of B can be computed by first diagonalising the matrix $B = VDV^\dagger$, where D is a diagonal matrix; then one forms the diagonal matrix $\sqrt{D}$ by taking the square roots of the elements on the diagonal in D; finally the square root of B is $V\sqrt{D}V^\dagger$.

The trace distance trd is an extension of the total variation distance tvd: given two discrete distributions ω_1, ω_2 on the same set, then the union of their supports $\mathrm{supp}(\omega_1) \cup \mathrm{supp}(\omega_2)$ is a finite set, say with n elements. We can represent ω_1, ω_2 via diagonal $n \times n$ matrices as density operators $\widehat{\omega_1}, \widehat{\omega_2}$. They are states, by construction. Then $\mathsf{trd}(\widehat{\omega_1}, \widehat{\omega_2}) = \mathsf{tvd}(\omega_1, \omega_2)$.

Example 10. We describe the quantum analogue of Example 2, involving the 'Bell' state. As a vector in $\mathbb{C}^2 \otimes \mathbb{C}^2$ the Bell state is usually described as $|\, b\,\rangle = \frac{1}{\sqrt{2}}(|\,00\,\rangle + |\,11\,\rangle)$. The corresponding density matrix $\beta = |\, b\,\rangle\langle\, b\,|$ is the following 4×4 matrix.

$$\beta = \tfrac{1}{2}\begin{pmatrix} 1&0&0&1 \\ 0&0&0&0 \\ 0&0&0&0 \\ 1&0&0&1 \end{pmatrix}$$

Its two marginals β_1, β_2 are equal 2×2 matrices, namely:

$$\beta_1 = \beta_2 = \tfrac{1}{2}\begin{pmatrix} 1&0 \\ 0&1 \end{pmatrix} \qquad \text{so that} \qquad \beta_1 \otimes \beta_2 = \tfrac{1}{4}\begin{pmatrix} 1&0&0&0 \\ 0&1&0&0 \\ 0&0&1&0 \\ 0&0&0&1 \end{pmatrix}$$

The product state $\beta_1 \otimes \beta_2$ is obtained as Kronecker product, see *e.g.* [27].

We can now ask the same question as in Example 2, namely what is the distance between the Bell state β and the product of its marginals. We recall that the Bell state is 'maximally entangled' and that the quantum theory allows, informally stated, higher levels of entanglement than in classical probability theory. Hence we expect an outcome that is higher than the value $\frac{1}{2}$ obtained in Example 2 for the classical maximally entwined state.

The key steps are:

$$\beta - \beta_1 \otimes \beta_2 = \begin{pmatrix} 1/4 & 0 & 0 & 1/2 \\ 0 & -1/4 & 0 & 0 \\ 0 & 0 & -1/4 & 0 \\ 1/2 & 0 & 0 & 1/4 \end{pmatrix} \qquad \text{so that} \qquad \big|\beta - \beta_1 \otimes \beta_2\big| = \begin{pmatrix} 1/2 & 0 & 0 & 1/2 \\ 0 & 1/4 & 0 & 0 \\ 0 & 0 & 1/4 & 0 \\ 1/2 & 0 & 0 & 1/2 \end{pmatrix}$$

Hence:

$$\mathsf{trd}\big(\beta,\, \beta_1 \otimes \beta_2\big) = \tfrac{1}{2}\mathrm{tr}\big(\big|\beta - \beta_1 \otimes \beta_2\big|\big) = \tfrac{1}{2}\Big(1/2 + 1/4 + 1/4 + 1/2\Big) = \tfrac{3}{4}.$$

In the earlier version of this paper [16] these distance computations are generalised to n-ary products, both for classical and for quantum states. Both distances then tend to 1, as n

goes to infinity, but the classical distance is one step behind, via formulas $\frac{2^{n-1}-1}{2^{n-1}}$ versus $\frac{2^n-1}{2^n}$. Here we only consider $n = 2$.

The following result is a quantum analogue of Proposition 3. Our formulation generalises the standard formulation of *e.g.* [27, §9.2] and its proof to arbitrary, not necessarily finite-dimensional Hilbert spaces. We'll see an even more general version involving von Neumann algebras later on.

Proposition 11. For states ϱ_1, ϱ_2 on the same Hilbert space $\mathscr{H}$,

$$\mathsf{trd}(\varrho_1, \varrho_2) = \bigvee_{p \in \mathcal{E}f(\mathscr{H})} \big|\, \varrho_1 \models p - \varrho_2 \models p \,\big| = \max_{s \in \mathcal{E}f(\mathscr{H}) \text{ sharp}} \varrho_1 \models s - \varrho_2 \models s. \qquad \square$$

As before, the maximum means the supremum is actually reached by a sharp effect. The proof of this result is in the appendix.

2.4. **States of von Neumann algebras.**

We will not even try to explain the basics of the theory of von Neumann algebras here; for the details we refer to [23]. We just recall some elementary facts which are relevant here. For us, a C^*-algebra $\mathscr{A}$ always has a unit element $1 \in \mathscr{A}$ and is a complex vector space. We write $[0,1]_{\mathscr{A}} \subseteq \mathscr{A}$ for subset of effects $0 \leq e \leq 1$; they will be used as quantum predicates. Such an effect e is called sharp (or a projection) if $e^2 = e$. A C^*-algebra $\mathscr{A}$ is a *von Neumann* algebra (aka. W^*-algebra) if firstly this unit interval $[0,1]_{\mathscr{A}}$ is a directed complete partial order (dcpo), and secondly the positive linear functionals $\omega\colon \mathscr{A} \to \mathbb{C}$ that preserve these (directed) suprema separate the elements of $[0,1]_{\mathscr{A}}$. In the notation introduced below this means that $e_1 = e_2$ follows if $\omega \models e_1 = \omega \models e_2$ for all states ω. There are several equivalent alternative definitions of the notion of 'von Neumann algebra', but this one, essentially due to Kadison (see [22]), is most convenient here.

We consider as morphisms $f\colon \mathscr{A} \to \mathscr{B}$ between von Neumann algebras: linear maps which are unital (that is, $f(1) = 1$), positive ($a \geq 0$ implies $f(a) \geq 0$) and normal. The latter normality requirement means that the restriction $f\colon [0,1]_{\mathscr{A}} \to [0,1]_{\mathscr{B}}$ preserves directed joins (*i.e.* is Scott continuous). This yields a category $\mathbf{vNA}$ of von Neumann algebras. It occurs naturally in opposite form, as $\mathbf{vNA}^{\mathrm{op}}$.

Each non-zero map f in $\mathbf{vNA}$ has operator norm equal to 1, *i.e.* $\|f\|_{\mathrm{op}} = 1$, where $\|f\|_{\mathrm{op}} = \bigvee\{\|f(x)\| \mid \|x\| = 1\}$. Below we apply the operator norm to a (pointwise) difference $\|f - g\|_{\mathrm{op}}$ of parallel maps f, g in $\mathbf{vNA}$. Using $\|f - g\|_{\mathrm{op}}$ as distance, each homset of $\mathbf{vNA}$ is a complete metric space.

A state of a von Neumann algebra $\mathscr{A}$ is a morphism $\varrho\colon \mathscr{A} \to \mathbb{C}$ in $\mathbf{vNA}$. We write $\mathrm{Stat}(\mathscr{A}) = \mathrm{Hom}(\mathscr{A}, \mathbb{C})$ for the set of states; it is easy to see that it is a convex set. For an effect $e \in [0,1]_{\mathscr{A}}$ we write $\varrho \models e$ for the value $\varrho(e) \in [0,1]$. When $\mathscr{A}$ is the von Neumann algebra $\mathcal{B}(\mathscr{H})$ of bounded operators on a Hilbert space $\mathscr{H}$, then 'effect' has a consistent meaning, since $[0,1]_{\mathscr{A}} = \mathcal{E}f(\mathscr{H})$. Moreover, density operators ϱ on $\mathscr{H}$ are in one–one correspondence with states of $\mathcal{B}(\mathscr{H})$, via $\varrho \mapsto \mathrm{tr}(\varrho(\,\cdot\,))$; in fact, this correspondence extends to a linear bipositive isometry between trace-class operators on $\mathscr{H}$ and *normal* — but not necessarily positive — functionals on $\mathcal{B}(\mathscr{H})$ (see [1, Thm 2.68]).

For states of von Neumann algebras we use *half* of the operator norm as distance, since it coincides with the 'validity' distance whose formulation is by now familiar. The proof is again delegated to the appendix.

Proposition 12. Let $\varrho_1, \varrho_2 \colon \mathscr{A} \to \mathbb{C}$ be two states of a von Neumann algebra $\mathscr{A}$. Their *validity distance* $\mathsf{vld}(\varrho_1, \varrho_2)$, as defined on the left below, satisfies:

$$\mathsf{vld}(\varrho_1, \varrho_2) \coloneqq \bigvee_{e \in [0,1]_{\mathscr{A}}} \Big| \varrho_1 \models e - \varrho_2 \models e \Big| = \max_{s \in [0,1]_{\mathscr{A}} \text{ sharp}} \varrho_1 \models s - \varrho_2 \models s \;=\; \tfrac{1}{2} \|\varrho_1 - \varrho_2\|_{\mathrm{op}}$$

Via the last equation it is easy to see that vld is a complete metric. □

Corollary 13. Let $\mathscr{A}$ be a von Neumann algebra.

(1) For each predicate $e \in [0,1]_{\mathscr{A}}$ the 'evaluate at e' map $\mathrm{ev}_e = (-)(e) = (-) \models e \colon \mathrm{Stat}(\mathscr{A}) \to [0,1]$ is both affine and non-expansive.
(2) The convex map $\alpha \colon \mathcal{D}(\mathrm{Stat}(\mathscr{A})) \to \mathrm{Stat}(\mathscr{A})$ is non-expansive.
(3) The 'states' functor $\mathrm{Stat} = \mathrm{Hom}(-, \mathbb{C}) \colon \mathbf{vNA}^{\mathrm{op}} \to \mathbf{Conv}$ restricts to a functor $\mathrm{Stat} \colon \mathbf{vNA}^{\mathrm{op}} \to \mathbf{ConvCMet}$.

Proof (1) It is standard that the map ev_e is affine, so we concentrate on its non-expansiveness: for states ϱ_1, ϱ_2 we have:

$$\big| \mathrm{ev}_e(\varrho_1) - \mathrm{ev}_e(\varrho_2) \big| = \big| \varrho_1 \models e - \varrho_2 \models e \big| \leq \bigvee\nolimits_{a \in [0,1]_{\mathscr{A}}} \big| \varrho_1 \models a - \varrho_2 \models a \big| = \mathsf{vld}(\varrho_1, \varrho_2).$$

(2) Suppose we have two formal convex combinations $\Omega = \sum_i r_i |\omega_i\rangle$ and $\Psi = \sum_j s_j |\varrho_j\rangle$ in $\mathcal{D}(\mathrm{Stat}(\mathscr{A}))$. The map $\alpha \colon \mathcal{D}(\mathrm{Stat}(\mathscr{A})) \to \mathrm{Stat}(\mathscr{A})$ is non-expansive since:

$$\begin{array}{rcl}
\mathsf{vld}\big(\alpha(\Omega), \alpha(\Psi)\big) & = & \bigvee_e \big| (\sum_i r_i \cdot \omega_i) \models e - (\sum_j s_j \cdot \varrho_j) \models e \big| \\
& = & \bigvee_e \big| \sum_i r_i \cdot \omega_i(e) - \sum_j s_j \cdot \varrho_j(e) \big| \\
& = & \bigvee_e \big| \sum_i r_i \cdot \mathrm{ev}_e(\omega_i) - \sum_j s_j \cdot \mathrm{ev}_e(\varrho_j) \big| \\
& = & \bigvee_e \big| \Omega \models \mathrm{ev}_e - \Psi \models \mathrm{ev}_e \big| \\
& \leq & \bigvee_{p \in \mathbf{Met}(\mathrm{Stat}(\mathscr{A}), [0,1])} \big| \Omega \models p - \Psi \models p \big| \\
& \stackrel{(2.4)}{=} & \mathsf{kvd}\big(\Omega, \Psi\big).
\end{array}$$

(3) We have to prove that for a positive unital map $f \colon \mathscr{A} \to \mathscr{B}$ between von Neumann algebras the associated state transformer $f_* = (-) \circ f \colon \mathrm{Hom}(\mathscr{B}, \mathbb{C}) \to \mathrm{Hom}(\mathscr{A}, \mathbb{C})$ is affine and non-expansive. The former is standard, so we concentrate on non-expansiveness. Let $\varrho_1, \varrho_2 \colon \mathscr{B} \to \mathbb{C}$ be states of $\mathscr{B}$. Then:

$$\begin{array}{rcl}
\mathsf{vld}\big(f_*(\varrho_1), f_*(\varrho_2)\big) & = & \bigvee_{e \in [0,1]_{\mathscr{A}}} \big| f_*(\varrho_1)(e) - f_*(\varrho_2)(e) \big| \\
& = & \bigvee_{e \in [0,1]_{\mathscr{A}}} \big| \varrho_1(f(e)) - \varrho_2(f(e)) \big| \\
& \leq & \bigvee_{d \in [0,1]_{\mathscr{B}}} \big| \varrho_1(d) - \varrho_2(d) \big| \\
& = & \mathsf{vld}\big(\varrho_1, \varrho_2\big).
\end{array}$$

□

3. Distances between effects (predicates)

There are several closely connected views on what are predicates in a probabilistic setting. Informally, one can consider fuzzy predicates $X \to [0,1]$ on a space X, or only the sharp ones $X \to \{0,1\}$. Instead of restricting oneself to truth values in $[0,1]$, one can use $\mathbb{R}$-valued predicates $X \to \mathbb{R}$, which are often called 'observables'. Alternatively, one can restrict to the non-negative ones $X \to [0, \infty)$. There are ways to translate between these views, by restriction, or by completion. The relevant underlying mathematical structures are: effect modules, order unit spaces, and ordered cones. Via suitable restrictions, the categories of these structures are equivalent. Here we choose to use effect modules because they capture $[0,1]$-valued predicates, which we consider to be most natural. Moreover, there is a standard adjunction between effect modules and the convex sets that we have been using in the previous section. This adjunction will be explored in the next section.

In this section we recall some basic facts from the theory of effect modules (see [13, 7, 20]), and add a few new ones, especially related to ω-joins and metric completeness, see Proposition 16. With these results in place, we observe that in our main examples — fuzzy predicates on a set and effects in a von Neumann algebras — the induced 'Archimedean' metric can also be expressed using validity $\models$, but now in dual form wrt. the previous section: for the distance between two predicates we now take a join over all states and use the validities of the two predicates in these states.

We briefly recall what an effect module is, and refer to [13] and its references for more details. This involves three steps.

(1) A *partial commutative monoid* (PCM) is given by a set E with an element $0 \in E$ and a partial binary operation $\ovee \colon E \times E \to E$ which is commutative and associative, in a suitably partial sense, and has 0 has unit element.
(2) An *effect algebra* is a PCM E in which each element $x \in E$ has a unique orthosupplement $x^{\perp} \in E$ with $x \ovee x^{\perp} = 1$, where $1 = 0^{\perp}$. Moreover, if $x \ovee 1$ is defined, then $x = 0$. Each effect algebra carries a partial order given by: $x \leq y$ iff $x \ovee z = y$ for some z. It satisfies $x \leq y$ iff $y^{\perp} \leq x^{\perp}$. For more information on effect algebras we refer to [9].
(3) An *effect module* is an effect algebra E with a (total) scalar multiplication operation $[0,1] \times E \to E$ which is an action a 'bihomomorphism'.

We write **EMod** for the category of effect modules. A map $f \colon E \to D$ in **EMod** preserves 1, sums $\ovee$, when they exist, and scalar multiplication; such an f then also preserves orthosupplements. There are (non-full) subcategories $\mathbf{DcEMod} \hookrightarrow \omega\text{-}\mathbf{EMod} \hookrightarrow \mathbf{EMod}$ of *directed complete* and *ω-complete* effect modules. The sum $\ovee$ and scalar multiplication $\cdot$ operations are required to preserve these joins in each argument separately[1]. The existence of orthosupplements $(-)^{\perp}$ implies that ω/directed meets exist, and that $(-)^{\perp}$ sends joins to meets and vice-versa. Morphisms in **DcEMod** and ω-**EMod** are homormorphisms of effect modules that additionally preserve the relevant joins.

Below it is shown how this effect module structure arises naturally in our main examples. The predicate functors Pred are special cases of constructions for 'effectuses', see [13].

Lemma 14. (1) For the distribution monad $\mathcal{D}$ on **Sets** there is a 'predicate' functor on its Kleisli category:

$$\mathcal{K}\ell(\mathcal{D}) \xrightarrow{\text{Pred}} \mathbf{DcEMod}^{\text{op}} \qquad \text{given by} \qquad \begin{cases} X & \mapsto \; [0,1]^X \\ \big(X \xrightarrow{f} \mathcal{D}(Y)\big) & \mapsto \; \big([0,1]^Y \xrightarrow{f^*} [0,1]^X\big) \end{cases}$$

This functor is faithful, and it is full (& faithful) if we restrict it to the subcategory $\mathcal{K}\ell_{\text{fin}}(\mathcal{D}) \hookrightarrow \mathcal{K}\ell(\mathcal{D})$ with *finite* sets as objects.

[1] In fact, it can be shown that maps $(-) \ovee y$ preserve joins automatically, see Lemma 15 (1i). Preservation by scalar multiplication can also be proved, but is outside the scope of this paper.

(2) There is also a 'predicate' functor:

$$\mathbf{vNA}^{\mathrm{op}} \xrightarrow{\mathrm{Pred}} \mathbf{DcEMod}^{\mathrm{op}} \qquad \text{given by} \qquad \begin{cases} \mathscr{A} \mapsto [0,1]_{\mathscr{A}} \\ \mathscr{B} \xrightarrow{f} \mathscr{A} \mapsto [0,1]_{\mathscr{B}} \xrightarrow{f} [0,1]_{\mathscr{A}} \end{cases}$$

This functor is full and faithful.

Writing $(-)^{\mathrm{op}}$ on both sides in point (2) looks rather formal, but makes sense since the category **vNA** of von Neumann algebra is naturally used in opposite form, see also the next section.

Proof (1) It is easy to see that the set $[0,1]^X$ of fuzzy predicate on a set X is an effect module, in which a sum $p \ovee q$ exists if $p(x) + q(x) \leq 1$ for all $x \in X$, and in that case $(p \ovee q)(x) = p(x) + q(x)$. Clearly, $p^{\perp}(x) = 1 - p(x)$ and $(r \cdot p)(x) = r \cdot p(x)$ for a scalar $r \in [0,1]$. The induced order on $[0,1]^X$ is the pointwise order, which is (directed) complete.

For a Kleisli map $f\colon X \to \mathcal{D}(Y)$ the predicate transformation map $f^*\colon [0,1]^Y \to [0,1]^X$ from (2.2) preserves the effect module structure. Moreover, it is Scott-continuous by the following argument. Let $q_i \in [0,1]^X$ be a directed collection of predicates, and let $x \in X$. Write the support of $f(x) \in \mathcal{D}(Y)$ as $\{y_1, \ldots, y_n\}$. Then:

$$\begin{array}{rcll} f^*\big(\bigvee_i q_i\big)(x) & \overset{(2.2)}{=} & f(x)(y_1) \cdot \big(\bigvee_i q_i\big)(y_1) + \cdots + f(x)(y_n) \cdot \big(\bigvee_i q_i\big)(y_n) & \\ & = & \big(\bigvee_i f(x)(y_1) \cdot q_i(y_1)\big) + \cdots + \big(\bigvee_i f(x)(y_n) \cdot q_i(y_n)\big) & \\ & = & \bigvee_{i_1} \cdots \bigvee_{i_n} f(x)(y_1) \cdot q_{i_1}(y_1) + \cdots + f(x)(y_n) \cdot q_{i_n}(y_n) & \\ & = & \bigvee_i f(x)(y_1) \cdot q_i(y_1) + \cdots + f(x)(y_n) \cdot q_i(y_n) & \text{since } (q_i) \text{ is directed} \\ & = & \bigvee_i f^*(q_i)(x) & \\ & = & \big(\bigvee_i f^*(q_i)\big)(x). & \end{array}$$

Assume $f^* = g^*$ for $f, g\colon X \to \mathcal{D}(Y)$, and let $x \in X$, $y \in Y$. Write $\mathbf{1}_{\{y\}} \in [0,1]^Y$ for the singleton predicate that is 1 on $y \in Y$ and zero everywhere else. Then $f(x)(y) = f^*(\mathbf{1}_{\{y\}})(x) = g^*(\mathbf{1}_{\{y\}})(x) = g(x)(y)$. Hence $f = g$, showing that Pred is faithful.

Now let X, Y be finite sets and $h\colon [0,1]^Y \to [0,1]^X$ a map in **DcEMod**. Define $f(x)(y) = h(\mathbf{1}_{\{y\}})(x) \in [0,1]$. We claim that $f(x)$ is a distribution on $Y = \{y_1, \ldots, y_n\}$, say, and that $f^* = h$. This works as follows.

$$\begin{array}{rl} \sum_{y \in Y} f(x)(y) & = \sum_i h(\mathbf{1}_{\{y_i\}})(x) \\ & = \big(\ovee_i h(\mathbf{1}_{\{y_i\}})\big)(x) \\ & = h\big(\ovee_i \mathbf{1}_{\{y_i\}}\big)(x) \\ & = h(\mathbf{1}_Y)(x) \\ & = \mathbf{1}_X(x) \\ & = 1. \end{array} \qquad \begin{array}{rl} f^*(q)(x) & = \sum_i f(x)(y_i) \cdot q(y_i) \\ & = \ovee_i h(\mathbf{1}_{\{y_i\}})(x) \cdot q(y_i) \\ & = h(\ovee_i q(y_i) \cdot \mathbf{1}_{\{y_i\}})(x) \\ & = h(q)(x). \end{array}$$

(2) It is not hard to see that the unit interval $[0,1]_{\mathscr{A}}$ of a C^*-algebra $\mathscr{A}$ is an effect module, see also [13]. If $\mathscr{A}$ is a von Neumann algebra, then this interval is a dcpo, by definition. Each map f of von Neumann algebra restricts to these intervals, and is in fact entirely determined by its behaviour on unit intervals: an arbitrary element

can be written as a linear combination of (four) positive elements; the latter can be scaled down with a scalar, if needed, so that they fit in the unit interval. □

We recall that the Archimedean property of an order unit space (with unit 1) is typically formulated as follows. Let x be an arbitrary element that satisfies $x \leq \frac{1}{n} \cdot 1$, for all $n \geq 1$, then $x \leq 0$. This Archimedean property is crucial for defining a norm on order unit spaces.

An analogous Archimedean property is given for effect modules in [18, 19]. Its formulation is more subtle, and runs as follows. For arbitrary elements x, y, if $\frac{1}{2} \cdot x \leq \frac{1}{2} \cdot y \ovee \frac{1}{2n} \cdot 1$ for all $n \geq 1$, then $x \leq y$. This formulation uses the fact that sums $r \cdot x \ovee s \cdot y$ with $r + s \leq 1$ always exist in an effect module.

Also for Archimedean effect modules one can define an 'Archimedean' distance function ard as:

$$\mathsf{ard}(x,y) = \max\Big(\bigwedge\{r \in (0,1] \mid \tfrac{1}{2} \cdot x \leq \tfrac{1}{2} \cdot y \ovee \tfrac{r}{2} \cdot 1\},\ \bigwedge\{r \in (0,1] \mid \tfrac{1}{2} \cdot y \leq \tfrac{1}{2} \cdot x \ovee \tfrac{r}{2} \cdot 1\}\Big)$$

In this situation we can write $\|x\| = \mathsf{ard}(0,x) \in [0,1]$, so that $x \leq \|x\| \cdot 1$. But we need be careful that we cannot express the distance ard in terms of $\| - \|$ since there is no general subtraction in effect modules — but there is a partial operation $\ominus$, see below.

In [18, 19] it is shown that:

- the full subcategory **AEMod** of Archimedean effect modules is equivalent to the category of order unit spaces; the 'Archimedean' distances on order unit spaces and effect modules coincide;
- Archimedean effect modules carry this (1-bounded) metric ard, and all maps of effect modules are automatically non-expansive. This gives a functor $\mathbf{AEMod} \to \mathbf{Met}$.

We need to collect a few basic facts about this Archimedean distance function ard, especially about its relation to (partial) subtraction $\ominus$ in the last point below.

Lemma 15. Let E be an Archimedean effect module. For $x, y \in E$ with $x \leq y$ one can define $y \ominus x = (y^\perp \ovee x)^\perp$. Then:

(1) This minus operation $\ominus$ satisfies the following properties:
  (a) $x \ominus 0 = x$ and $1 \ominus y = y^\perp$ and $x \ominus x = 0$;
  (b) if $y \leq z$ then: $x \ovee y = z$ iff $x = z \ominus y$; in particular, $x = (x \ovee y) \ominus y$ and $(z \ominus y) \ovee y = z$;
  (c) $x \ovee y \leq z$ iff $x \leq z \ominus y$
  (d) if $x \leq y$ then $(y \ovee z) \ominus x = (y \ominus x) \ovee z$;
  (e) if $x \leq y \leq z$ then $y \ominus x \leq z \ominus x$;
  (f) if $x \geq y$ then $x \leq y \ovee z$ iff $x \ominus y \leq z$;
  (g) if $x \leq y$ then $r \cdot y \ominus r \cdot x = r \cdot (y \ominus x)$ for $r \in [0,1]$;
  (h) if $r \leq s$ in $[0,1]$, then $s \ominus r = s - r$ and $(s - r) \cdot x = s \cdot x \ominus r \cdot x$;
  (i) the operation $(-) \ovee y$ preserves all joins and meets that exist;
  (j) $(-) \ominus y$ preserves all joins and meets, and $x \ominus (-)$ sends joins to meets and vice-versa;
  (k) if scalar multiplication preserves joins in one of its arguments, then it also preserves meets in that argument.

(2) The sum $\ovee$, orthosupplement $(-)^\perp$, and scalar multiplication $\cdot$ operations are continuous in each argument wrt. the Archimedean metric ard, and so subtraction $\ominus$ too.

(3) If $x \leq y$, then $y \ominus x \leq \mathsf{ard}(x,y) \cdot 1$.

Proof (1) The first point is trivial, and left to the reader. For (1b) we use: $x = z \ominus y = (z^\perp \ovee y)^\perp$ iff $x^\perp = z^\perp \ovee y$ iff $x \ovee z^\perp \ovee y = 1$ iff $z = x \ovee y$. Next, for (1c),

$$\begin{aligned} x \ovee y \leq z \;\Leftrightarrow\; \exists w.\, x \ovee y \ovee w = z \;&\Leftrightarrow\; \exists w.\, x \ovee w = z \ominus y \qquad \text{as just shown} \\ &\Leftrightarrow\; x \leq z \ominus y. \end{aligned}$$

Point (1d) is obtained as follows. We have:

$$(y \ovee z)^\perp \ovee x \ovee (y \ominus x) \ovee z \overset{(1b)}{=} (y \ovee z)^\perp \ovee y \ovee z = 1,$$

so that $(y \ominus x) \ovee z = \big((y \ovee z)^\perp \ovee x\big)^\perp = ((y \ovee z) \ominus x$.

For (1e) let $x \leq y \leq z$, say via $z = y \ovee w$. Then $z \ominus x = (y \ovee w) \ominus x = (y \ominus x) \ovee w$ by the previous point. Hence $y \ominus x \leq z \ominus x$.

Assume now $x \geq y$ for (1f). In one direction, if $x \leq y \ovee z$, then, by the previous point, $x \ominus y \leq (y \ovee z) \ominus y = z$. The other direction follows similary by adding y on both sides.

For (1g) let $x \leq y$ and $r \in [0,1]$; the equation $r \cdot y \ominus r \cdot x = r \cdot (y \ominus x)$ follows from:

$$\big(r \cdot y \ominus r \cdot x\big) \ovee r \cdot x = r \cdot y = r \cdot \big((y \ominus x) \ovee x\big) = r \cdot (y \ominus x) \ovee r \cdot x.$$

Point (1h) is easy and left to the reader; for (1i), let $x = \bigvee x_i$. Then:

$$\begin{aligned} (\textstyle\bigvee x_i) \ovee y \leq z &\iff \textstyle\bigvee x_i \leq z \ominus y \qquad \text{by (1c)} \\ &\iff x_i \leq z \ominus y, \text{ for all } i \\ &\iff x_i \ovee y \leq z, \text{ for all } i \\ &\iff \textstyle\bigvee (x_i \ovee y) \leq z. \end{aligned}$$

Similary, meets are preserved by point (1f).

For (1j) we use that $(\bigvee x_i)^\perp = \bigwedge x_i^\perp$ since $x \leq y$ iff $y^\perp \leq x^\perp$. Then by (1i):

$$\begin{aligned} (\textstyle\bigvee x_i) \ominus y \;=\; \big((\bigvee x_i)^\perp \ovee y\big)^\perp \;=\; \big((\bigwedge x_i^\perp) \ovee y\big)^\perp &= \big(\textstyle\bigwedge (x_i^\perp \ovee y)\big)^\perp \\ &= \textstyle\bigvee \big(x_i^\perp \ovee y\big)^\perp \;=\; \bigvee (x_i \ominus y). \end{aligned}$$

Similarly one proves that $(-) \ominus y$ preserves meets, and that $x \ominus (-)$ sends joins to meets and vice-versa.

Finally, for (1k), let $(-) \cdot x$ preserve joins. Then:

$$\begin{aligned} \big(\textstyle\bigwedge r_i\big) \cdot x \;=\; \big(\bigvee r_i^\perp\big)^\perp \cdot x &= \big(1 - \textstyle\bigvee r_i^\perp\big) \cdot x \\ &= 1 \cdot x \ominus \big(\textstyle\bigvee r_i^\perp\big) \cdot x \qquad \text{by (1h)} \\ &= 1 \cdot x \ominus \big(\textstyle\bigvee r_i^\perp \cdot x\big) \\ &= \textstyle\bigwedge 1 \cdot x \ominus r_i^\perp \cdot x \qquad \text{by (1j)} \\ &= \textstyle\bigwedge r_i \cdot x. \end{aligned}$$

Next let $r \cdot (-)$ preserve joins. We use that $r \cdot (x^\perp) = (r \cdot x \ovee r^\perp \cdot 1)^\perp$ in:

$$\begin{aligned} r \cdot \big(\textstyle\bigwedge x_i\big) \;=\; r \cdot \big(\bigvee x_i^\perp\big)^\perp &= \Big(r \cdot \big(\textstyle\bigvee x_i^\perp\big) \ovee r^\perp \cdot 1\Big)^\perp \\ &= \Big(\textstyle\bigvee r \cdot x_i^\perp \ovee r^\perp \cdot 1\Big)^\perp \\ &= \textstyle\bigwedge r \cdot x_i. \end{aligned}$$

(2) Proving continuity of the effect module operations is tedious, but we can exploit the equivalence of (Archimedean) effect modules and order unit spaces and use continuity of the corresponding operations for order unit spaces.

(3) Let $x \leq y$. Then $\bigwedge\{r \mid \frac{1}{2}\cdot x \leq \frac{1}{2}\cdot y \otimes \frac{r}{2}\cdot 1\} = 0$, so that $\mathsf{ard}(x,y) = \bigwedge\{r \mid \frac{1}{2}\cdot y \leq \frac{1}{2}\cdot x \otimes \frac{r}{2}\cdot 1\}$. We recall that in an ω-complete effect module scalar multiplication preserves ω-joins in each argument, by definition. Hence point (1k) can be used, in:

$$\begin{aligned}
\mathsf{ard}(x,y)\cdot 1 &= \Big(\bigwedge\{r \mid \tfrac{1}{2}\cdot y \leq \tfrac{1}{2}\cdot x \otimes \tfrac{r}{2}\cdot 1\}\Big)\cdot 1 \\
&\overset{(1k)}{=} \bigwedge\{r\cdot 1 \mid \tfrac{1}{2}\cdot y \leq \tfrac{1}{2}\cdot x \otimes \tfrac{r}{2}\cdot 1\} \\
&\overset{(1f)}{=} \bigwedge\{r\cdot 1 \mid \tfrac{1}{2}\cdot y \ominus \tfrac{1}{2}\cdot x \leq \tfrac{r}{2}\cdot 1\} \\
&\geq \big(\tfrac{1}{2}\cdot y \ominus \tfrac{1}{2}\cdot x\big) \otimes \big(\tfrac{1}{2}\cdot y \ominus \tfrac{1}{2}\cdot x\big) \\
&\overset{(1g)}{=} \tfrac{1}{2}\cdot\big(y\ominus x\big) \otimes \tfrac{1}{2}\cdot\big(y\ominus x\big) \\
&= y \ominus x.
\end{aligned}$$

□

Proposition 16. Let E be an ω-complete effect module. Then:

(1) E is Archimedean;

(2) E is metrically complete for the above Archimedean distance function ard;

(3) for each ascending sequence $e_1 \leq e_2 \leq e_3 \leq \cdots$ which is Cauchy, one has $\bigvee e_n = \lim e_n$.

Proof (1) Assume $\frac{1}{2}\cdot x \leq \frac{1}{2}\cdot y \otimes \frac{1}{2n}\cdot 1$ for all $n \geq 1$. We need to prove $x \leq y$. This follows directly from the existence of ω-meets:

$$\begin{aligned}
\tfrac{1}{2}\cdot x \;\leq\; \textstyle\bigwedge_n \tfrac{1}{2}\cdot y \otimes \tfrac{1}{2n}\cdot 1 &= \tfrac{1}{2}\cdot y \otimes \big(\textstyle\bigwedge_n \tfrac{1}{2n}\cdot 1\big) && \text{by Lemma 15 (1i)} \\
&= \tfrac{1}{2}\cdot y \otimes \big(\textstyle\bigwedge_n \tfrac{1}{2n}\big)\cdot 1 && \text{by Lemma 15 (1k)} \\
&= \tfrac{1}{2}\cdot y \otimes 0\cdot 1 \\
&= \tfrac{1}{2}\cdot y.
\end{aligned}$$

(2) We first prove an auxiliary result:

$$\begin{gathered}
\text{assume that for each sequence } a_1, a_2, \ldots \in E \text{ for which } \textstyle\sum_n \|a_n\| \leq 1, \\
\text{the sums } b_N \coloneqq \textstyle\bigotimes_{n\leq N} a_n \text{ converge;} \qquad (*) \\
\text{then } E \text{ is complete.}
\end{gathered}$$

We first remark that the sums b_N exist since $\bigotimes_{n\leq N} a_n \leq \bigotimes_{n\leq N} \|a_n\|\cdot 1 = \big(\bigotimes_{n\leq N}\|a_n\|\big)\cdot 1 \leq 1\cdot 1 = 1$.

We start by proving the statement $(*)$. Let $x_1, x_2, \ldots \in E$ be a Cauchy sequence; we need to prove that it converges, given the assumption in $(*)$. We replace $x_1, x_2, \ldots$ by $\frac{1}{2}\cdot x_1$, $\frac{1}{2}\cdot x_2$, $\ldots$ so that we may assume that $\|x_n\| \leq \frac{1}{2}$ for all n, because if $(\frac{1}{2}\cdot x_n)_n$ converges, then so does $(x_n)_n$. Similarly, by replacing $(x_n)_n$ by an appropriate subsequence we may assume that $\mathsf{ard}(x_m, x_n) \leq (\frac{1}{2})^n$ for all $m \geq n$. In particular, $\mathsf{ard}(x_{n+1}, x_n) \leq (\frac{1}{2})^{n+1}$, that is,

$$x_n \leq x_{n+1} \otimes (\tfrac{1}{2})^{n+1}\cdot 1 \qquad \text{and} \qquad x_{n+1} \leq x_n \otimes (\tfrac{1}{2})^{n+1}\cdot 1.$$

We then have:

$$\begin{aligned} x_1 \;\leq\; x_2 \ovee (\tfrac{1}{2})^2 \cdot 1 &\leq x_3 \ovee (\tfrac{1}{2})^3 \cdot 1 \ovee (\tfrac{1}{2})^2 \cdot 1 \\ &\leq \cdots \\ &\leq x_{n+1} \ovee (\tfrac{1}{2})^{n+1} \cdot 1 \ovee \cdots \ovee (\tfrac{1}{2})^2 \cdot 1 \\ &= x_{n+1} \ovee (\tfrac{1}{2} - (\tfrac{1}{2})^{n+1}) \cdot 1. \end{aligned}$$

The trick is to consider the elements $a_n \coloneqq (x_{n+1} \ovee (\frac{1}{2})^{n+1} \cdot 1) \ominus x_n$. We check that these a_n satisfy the requirement in $(*)$:

$$a_n \;=\; (x_{n+1} \ovee (\tfrac{1}{2})^{n+1} \cdot 1) \ominus x_n \;\leq\; (x_n \ovee (\tfrac{1}{2})^{n+1} \cdot 1 \ovee (\tfrac{1}{2})^{n+1} \cdot 1) \ominus x_n \;=\; (\tfrac{1}{2})^n \cdot 1.$$

Thus we have $\|a_n\| \leq (\frac{1}{2})^n$, and so $\sum_n \|a_n\| \leq 1$. We may now additionally assume that the sums $b_N \coloneqq \bigovee_{n \leq N} a_n$ converge. These sums can be re-organised as:

$$\begin{aligned} b_N \;&=\; \textstyle\bigovee_{n \leq N} a_n \\ &=\; \big((x_{N+1} \ovee (\tfrac{1}{2})^{N+1} \cdot 1) \ominus x_N\big) \ovee \big((x_N \ovee (\tfrac{1}{2})^N) \cdot 1 \ominus x_{N-1}\big) \ovee \cdots \\ &\qquad\quad \ovee \big((x_2 \ovee (\tfrac{1}{2})^2) \cdot 1 \ominus x_1\big) \\ &=\; \big(x_{N+1} \ovee (\tfrac{1}{2})^{N+1} \cdot 1 \ovee (\tfrac{1}{2})^N \cdot 1 \ovee \cdots \ovee (\tfrac{1}{2})^2 \cdot 1\big) \ominus x_1 \\ &=\; \big(x_{N+1} \ovee (\tfrac{1}{2} - (\tfrac{1}{2})^{N+1}) \cdot 1\big) \ominus x_1. \end{aligned}$$

We claim that we can now also show that the sequence of x_N converges, since:

$$x_{N+1} \;=\; (b_N \ovee x_1) \ominus (\tfrac{1}{2} - (\tfrac{1}{2})^{N+1}) \cdot 1,$$

This right-hand-side converges too, as N goes to infinity.

Thus we have proven $(*)$. We now use it to prove our original aim, namely that E is complete. So let $a_1, a_2, \ldots \in E$ for which $s \coloneqq \sum_n \|a_n\| \leq 1$ and sums $b_N \coloneqq \bigovee_{n \leq N} a_n$ exist. These b_N form an ascending chain, so by ω-completeness of E, the suppremum $b \coloneqq \bigvee_N b_N$ exits. We are done if we can show that b is the limit of the b_N.

For $M \leq N$ we have:

$$\begin{aligned} b_N \ominus b_M \;=\; a_N \ovee \cdots \ovee a_{M+1} &\leq \|a_N\| \cdot 1 \ovee \cdots \ovee \|a_{M+1}\| \cdot 1 \\ &= \big(\|a_N\| + \cdots + \|a_{M+1}\|\big) \cdot 1. \end{aligned}$$

This means:

$$\begin{aligned} b \ominus b_M \;&=\; \big(\textstyle\bigvee_N b_N\big) \ominus b_M \\ &=\; \big(\textstyle\bigvee_{N \geq M} b_N\big) \ominus b_M \\ &=\; \textstyle\bigvee_{N \geq M} b_N \ominus b_M \qquad \text{by Lemma 15 (1i)} \\ &\leq\; \textstyle\bigvee_{N \geq M} \big(\|a_N\| + \cdots + \|a_{M+1}\|\big) \cdot 1 \\ &=\; \big(\textstyle\bigvee_{N \geq M} \|a_N\| + \cdots + \|a_{M+1}\|\big) \cdot 1 \\ &=\; \big(s - (\|a_M\| + \cdots + \|a_1\|)\big) \cdot 1, \qquad \text{where, recall, } s \coloneqq \textstyle\sum_n \|a_n\| \in [0,1]. \end{aligned}$$

The latter scalar becomes arbitrarily small as M goes to infinity. This means that $\mathsf{ard}(b, b_M)$ can be made arbitrarily small. Hence $\lim_M b_M = b$, as required.

(3) Let $e_1 \leq e_2 \leq \cdots$ be a Cauchy sequence and let $\epsilon > 0$. We can find an $N \in \mathbb{N}$ such that $\mathsf{ard}(e_n, e_m) < \epsilon$ for all $n, m \geq N$. For $m \geq N$ we have $e_m \leq \bigvee_n e_n$, so that:

$$\begin{aligned}
(\textstyle\bigvee_n e_n) \ominus e_m &= (\textstyle\bigvee_{n \geq m} e_n) \ominus e_m \\
&= \textstyle\bigvee_{n \geq m} (e_n \ominus e_m) && \text{by Lemma 15 (1i)} \\
&\leq \textstyle\bigvee_{n \geq m} \mathsf{ard}(e_n, e_m) \cdot 1 && \text{by Lemma 15 (3)} \\
&= \big(\textstyle\bigvee_{n \geq m} \mathsf{ard}(e_n, e_m)\big) \cdot 1 \\
&\leq \epsilon \cdot 1.
\end{aligned}$$

Lemma 15 (1f) gives $\bigvee_n e_n \leq e_m \owedge \epsilon \cdot 1$, and in particular $\frac{1}{2} \cdot (\bigvee_n e_n) \leq \frac{1}{2} \cdot e_m \owedge \frac{\epsilon}{2} \cdot 1$. Hence $\mathsf{ard}(\bigvee_n e_n, e_m) \leq \epsilon$, so that $\lim_m e_m = \bigvee_n e_n$. □

With this information about distances and joins and their relation in effect modules we return to our main examples from Lemma 14. We describe the Archimedean metrics in these cases in more detail, and discover that we can describe them also as 'validity' metrics, but in dual form: here they involve joins over states, and not over predicates like in Section 2.

Proposition 17. (1) Let X be an arbitrary set. The Archimedean metric ard induced on the effect module $[0, 1]^X$ of fuzzy predicates on X is the supremum metric (2.3), as observed in [18, 19]. But this metric can alternatively be described via 'validities', as (in the last equation) in:

$$\mathsf{ard}(p, q) = \mathsf{spd}(p, q) \overset{(2.3)}{=} \bigvee_{x \in X} \big|p(x) - q(x)\big| = \bigvee_{\omega \in \mathcal{D}(X)} \big|\, \omega \models p - \omega \models q \,\big|.$$

(2) Let $\mathscr{A}$ be a von Neumann algebra. The Archimedean metric ard on the effect module $[0, 1]_{\mathscr{A}}$ of effects of $\mathscr{A}$ is the distance induced by the norm $\| - \|$ of $\mathscr{A}$. Moreover, this distance can be described as on the right below.

$$\mathsf{ard}(e, d) = \big\|\, e - d \,\big\| = \bigvee_{\omega \colon \mathscr{A} \to \mathbb{C}} \big|\, \omega \models e - \omega \models d \,\big|.$$

Proof (1) Let $p, q \in [0, 1]^X$. We abbreviate $s \coloneqq \bigvee_x |p(x) - q(x)|$ and $t \coloneqq \bigvee_\omega |\omega \models p - \omega \models q|$. First note that for $x \in X$ the unit (or 'Dirac') distribution $\eta(x) = 1|x\rangle$ satisfies $\eta(x) \models p = p(x)$. This yields $s \leq t$. The converse inequality $t \leq s$ follows from:

$$\begin{aligned}
t = \textstyle\bigvee_\omega \big|\, \omega \models p - \omega \models q \,\big| &= \textstyle\bigvee_\omega \big|\, \sum_x \omega(x) \cdot p(x) - \sum_x \omega(x) \cdot q(x) \,\big| \\
&\leq \textstyle\bigvee_\omega \sum_x \omega(x) \cdot \big|p(x) - q(x)\big| \\
&\leq \textstyle\bigvee_\omega \sum_x \omega(x) \cdot s \\
&= \textstyle\bigvee_\omega (\sum_x \omega(x)) \cdot s \\
&= s.
\end{aligned}$$

(2) From [23, Cor. 4.3.10] we see that for a self-adjoint element $a \in \mathscr{A}$ we have $\|a\| = \bigvee_\omega |\omega(a)|$, where ω ranges over (normal) states $\mathscr{A} \to \mathbb{C}$. Thus:

$$\mathsf{ard}(e, d) = \big\|\, e - d \,\big\| = \textstyle\bigvee_\omega \big|\, \omega(e - d) \,\big| = \bigvee_\omega \big|\, \omega(e) - \omega(d) \,\big| = \bigvee_\omega \big|\, \omega \models e - \omega \models d \,\big|. \quad \square$$

4. State-and-effect triangles

In this section the results from the two previous sections are combined. This will happen via the adjunction $\mathbf{EMod}^{\mathrm{op}} \rightleftarrows \mathbf{Conv}$ between effect modules and convex sets from [11]. This adjunction is restricted by imposing completeness requirements on both sides. Then it is shown how our standard examples give rise to commuting state-and-effect triangles with full and faithful state and predicate functors.

Recall from Section 2 that we write **ConvMet** for the category of convex metric spaces, and **ConvCMet** for the subcategory of convex *complete* metric spaces.

Lemma 18. The adjunction from [11] on the left below restricts to the adjunction on the right.

$$\begin{tikzcd} \mathbf{EMod}^{\mathrm{op}} \arrow[r, bend left] \arrow[r, phantom, "\top"] & \mathbf{Conv} \arrow[l, bend left] \end{tikzcd} \qquad \begin{tikzcd} \mathbf{DcEMod}^{\mathrm{op}} \arrow[r, bend left] \arrow[r, phantom, "\top"] & \mathbf{ConvCMet} \arrow[l, bend left] \end{tikzcd} \tag{4.1}$$

All functors are given by 'homming into $[0,1]$'.

Proof The proof boils down to two points:

(1) For a directed complete effect module E, the convex set $\mathbf{DcEMod}\big(E,[0,1]\big)$ is a (convex) complete metric space.
(2) For a convex complete metric space X, the effect module $\mathbf{ConvCMet}\big(X,[0,1]\big)$ is directed complete.

As to point (1), let E be a directed complete effect module. The homset $\mathbf{DcEMod}(E,[0,1])$ carries the supremum metric (2.3). This metric is complete with pointwise limits: $(\lim h_n)(e) = \lim h_n(e)$. It is easy to see that such a limit map $\lim h_n$ preserves sums $\ovee$ and scalar multiplication. Hence it is a map of effect modules, and thus automatically a non-expansive (and continuous) function. In order to see that it is also Scott continuous, let (e_i) be directed collection of elements in E. Writing $h = \lim h_n$ we have to prove $h(\bigvee e_i) = \bigvee h(e_i)$. This works as follows. For each n and j we have:

$$\begin{aligned} \big|h(\textstyle\bigvee_i e_i) - \bigvee_i h(e_i)\big| \;&\leq\; \big|h(\textstyle\bigvee_i e_i) - h_n(\bigvee_i e_i)\big| + \big|\textstyle\bigvee_i h_n(e_i) - h_n(e_j)\big| \\ &\qquad + \big|h_n(e_j) - h(e_j)\big| + \big|h(e_j) - \textstyle\bigvee_i h(e_i)\big| \\ &\leq\; \mathsf{spd}(h,h_n) + \big|\textstyle\bigvee_i h_n(e_i) - h_n(e_j)\big| + \mathsf{spd}(h,h_n) + \big|h(e_j) - \textstyle\bigvee_i h(e_i)\big|. \end{aligned}$$

By choosing n suitable large, the two spd distances can be made arbitrarily small. Having fixed n, the term $|\bigvee_i h_n(e_i) - h_n(e_j)|$ can be made arbitrary small too by choosing j suitable large, since the directed net $(\,h_n(e_i)\,)_i$ in $[0,1]$ converges to its supremum $\bigvee_i h_n(e_i)$. Since the final term $|h(e_j) - \bigvee_i h(e_i)|$ vanishes too as j increases we see that $|h(\bigvee_i e_i) - \bigvee_i h(e_i)| = 0$, and so $h(\bigvee_i e_i) = \bigvee_i h(e_i)$.

The homset $\mathbf{DcEMod}\big(E,[0,1]\big)$ also has a convex structure, given by the map:

$$\mathcal{D}\Big(\mathbf{DcEMod}\big(E,[0,1]\big)\Big) \xrightarrow{\;\alpha\;} \mathbf{DcEMod}\big(E,[0,1]\big) \quad \text{with} \quad \alpha(\omega)(e) = \textstyle\sum_h \omega(h)\cdot h(e),$$

where h ranges over $\mathbf{DcEMod}\big(E,[0,1]\big)$. Notice that each element $e \in E$ gives rise to a non-expansive predicate $\mathrm{ev}_e\colon \mathbf{DcEMod}(E,[0,1]) \to [0,1]$ via $\mathrm{ev}_e(h) = h(e)$. It satisfies for $\omega \in \mathcal{D}\big(\mathbf{DcEMod}(E,[0,1])\big)$,

$$\omega \models \mathrm{ev}_e \;=\; \textstyle\sum_h \omega(h)\cdot \mathrm{ev}_e(h) \;=\; \sum_h \omega(h)\cdot h(e) \;=\; \alpha(\omega)(e).$$

Now we can show that the algebra map α on $\mathbf{DcEMod}\big(E,[0,1]\big)$ is non-expansive, using the Kantorovich metric (2.4) on distributions:

$$\begin{aligned}
\mathsf{spd}\big(\alpha(\omega_1),\alpha(\omega_2)\big) \stackrel{(2.3)}{=} \bigvee\nolimits_e \big|\,\alpha(\omega_1)(e)-\alpha(\omega_2)(e)\,\big| \;&=\; \bigvee\nolimits_e \big|\,\omega_1\models \mathrm{ev}_e - \omega_2\models \mathrm{ev}_e\,\big| \\
&\leq\; \bigvee\nolimits_p \big|\,\omega_1\models p - \omega_2\models p\,\big| \\
&\stackrel{(2.4)}{=}\; \mathsf{kvd}(\omega_1,\omega_2).
\end{aligned}$$

Each map $f\colon E \to D$ in **EMod** gives an affine map $(-)\circ f\colon \mathrm{Hom}(D,[0,1]) \to \mathrm{Hom}(E,[0,1])$ in **Conv**; it is easy to show that it is also non-expansive.

For point (2) we have to prove that for each convex complete metric space X the set $\mathbf{ConvCMet}(X,[0,1])$ of affine non-expansive maps is a directed complete effect module. We concentrate on directed completeness, since the effect module structure is standard, see [11]. Hence let (p_i) be a directed collection of non-expansive affine maps $p_i\colon X\to[0,1]$. We take $p=\bigvee_i p_i$ pointwise. This map is affine since affine sums are by definition finite, so that they commute with *directed* joins:

$$\begin{aligned}
p(\textstyle\sum_n r_n|x_n\rangle) \;=\; \big(\bigvee_i p_i\big)(\sum_n r_n|x_n\rangle) &= \textstyle\bigvee_i p_i(\sum_n r_n|x_n\rangle) \\
&= \textstyle\bigvee_i \sum_n r_n\cdot p_i(x_n) \\
&= \textstyle\sum_n r_n\cdot\big(\bigvee_i p_i(x_n)\big) \;=\; \sum_n r_n\cdot p(x_n).
\end{aligned}$$

It is not hard to see that p is non-expansive. $\square$

The next two results summarise our main concrete findings.

Proposition 19. The Kleisli subcategory $\mathcal{K}\ell_{\mathrm{fin}}(\mathcal{D})$, with finite sets only, of the distribution monad $\mathcal{D}$ on **Sets** gives rise to a triangle as below, in which the two up-going functors are full and faithful and make the two corresponding triangles commute up-to isomorphism.

$$\begin{tikzcd}
\mathbf{DcEMod}^{\mathrm{op}} \arrow[rr, bend left=10] \arrow[rr, phantom, "\top"] & & \mathbf{ConvCMet} \arrow[ll, bend left=10] \\
& \mathcal{K}\ell_{\mathrm{fin}}(\mathcal{D}) \arrow[ul, "\mathrm{Hom}(-,2)=\mathrm{Pred}"] \arrow[ur, "\mathrm{Stat}=\mathrm{Hom}(1,-)"'] &
\end{tikzcd}$$

Proof We use the full and faithful predicate functor $\mathrm{Pred} = [0,1]^{(-)}\colon \mathcal{K}\ell_{\mathrm{fin}}(\mathcal{D}) \to \mathbf{DcEMod}^{\mathrm{op}}$ from Lemma 14 (1). The states functor $\mathrm{Stat}\colon \mathcal{K}\ell_{\mathrm{fin}}(\mathcal{D}) \to \mathbf{Conv} = \mathcal{EM}(\mathcal{D})$ is the full and faithful Kleisli extension functor, restricted to finite sets. The functor restricts to *metric* spaces $\mathbf{ConvMet} \hookrightarrow \mathbf{Conv}$ by Lemma 6 and to *complete* spaces $\mathbf{ConvCMet} \hookrightarrow \mathbf{ConvMet}$ by Lemma 4. We need to check that the two triangles commute.

In one direction we have, for a finite set X,

$$\begin{aligned}
\Big(\mathbf{DcEMod}\big(-,[0,1]\big)\circ\mathrm{Pred}\Big)(X) &= \mathbf{DcEMod}\big([0,1]^X,[0,1]\big) \\
&\cong \mathcal{K}\ell_{\mathrm{fin}}(\mathcal{D})\big(1,X\big) \qquad \text{since Pred is full \& faithful} \\
&\cong \mathcal{D}(X) \\
&= \mathrm{Stat}(X)
\end{aligned}$$

In the other direction:

$$\begin{aligned}\Big(\mathbf{ConvCMet}\big(-,[0,1]\big)\circ \mathrm{Stat}\Big)(X) &= \mathbf{ConvMet}\big(\mathcal{D}(X),[0,1]\big) \\ &\cong \mathbf{Met}(X,[0,1]) \quad \text{using } X \text{ with discrete metric} \\ &= \mathbf{Sets}(X,[0,1]) \\ &= \mathrm{Pred}(X). \qquad \square\end{aligned}$$

The description, in the above triangle, of the predicate and state functors via homsets $\mathrm{Hom}(-,2)$ and $\mathrm{Hom}(1,-)$ comes from effectus theory [13, 7]. It also applies to von Neumann algebras, when we use their category in opposite form, as $\mathbf{vNA}^{\mathrm{op}}$. For instance, the initial object in $\mathbf{vNA}$ is the algebra $\mathbb{C}$ of complex numbers; it forms the final object 1 in $\mathbf{vNA}^{\mathrm{op}}$. Thus, a map $1 \to \mathscr{A}$ in $\mathbf{vNA}^{\mathrm{op}}$ is a state $\mathscr{A} \to \mathbb{C}$, as we have described before. In a similar way one can check that maps $\mathscr{A} \to 2 = 1+1$ in $\mathbf{vNA}^{\mathrm{op}}$ correspond to effects in the unit interval $[0,1]_{\mathscr{A}}$, see [13] for details.

Proposition 20. The opposite of the category $\mathbf{vNA}$ of von Neumann algebras fits in a triangle as below, in which the predicate and state functors are full and faithful and make the triangles commute up-to isomorphism.

$$\begin{tikzcd} \mathbf{DcEMod}^{\mathrm{op}} \arrow[rr, bend left, ""] \arrow[rr, phantom, "\top"] & & \mathbf{ConvCMet} \arrow[ll, bend left, ""] \\ & \mathbf{vNA}^{\mathrm{op}} \arrow[ul, "\mathrm{Hom}(-,2)=\mathrm{Pred}"] \arrow[ur, "\mathrm{Stat}=\mathrm{Hom}(1,-)"'] & \end{tikzcd}$$

Proof In Lemma 14 (2) we have seen that the predicate functor $\mathrm{Pred} = [0,1]_{(-)}\colon \mathbf{vNA}^{\mathrm{op}} \to \mathbf{DcEMod}^{\mathrm{op}}$ is full and faithful. For convenience we abbreviate $\mathcal{F} = \mathbf{ConvCMet}(-,[0,1])$ and $\mathcal{G} = \mathbf{DcEMod}(-,[0,1])$ so that $\mathcal{F} \dashv \mathcal{G}$ at the top of the above triangle.

Starting from the predicate functor Pred the above triangle commutes, since Pred is full and faithful:

$$\begin{aligned}\mathcal{G}\mathrm{Pred}(\mathscr{A}) \;=\; \mathbf{DcEMod}\Big(\mathrm{Pred}(\mathscr{A}),[0,1]\Big) &= \mathbf{DcEMod}\Big(\mathrm{Pred}(\mathscr{A}),\mathrm{Pred}(\mathbb{C})\Big) \\ &\cong \mathbf{vNA}\big(\mathscr{A},\mathbb{C}\big) \\ &= \mathrm{Stat}(\mathscr{A}).\end{aligned}$$

Commutation of the second triangle is less obvious. It relies on the fact that the linear combinations of normal states on $\mathscr{A}$ form a closed linear subspace $\mathscr{A}_*$ of the continuous dual $\mathscr{A}^*$ of $\mathscr{A}$. This "pre-dual" $\mathscr{A}_*$ of $\mathscr{A}$ determines the order and norm of $\mathscr{A}$ in the sense that the map $a \mapsto \hat{a}\colon \mathscr{A} \to (\mathscr{A}_*)^*$ which sends $a \in \mathscr{A}$ to the bounded functional $\hat{a}\colon \mathscr{A}_* \to \mathbb{C}$ given by $\hat{a}(\varphi) = \varphi(a)$ is a linear isomorphism $\mathscr{A} \to (\mathscr{A}_*)^*$ that preserves (and reflects) both the norm and the order, see *e.g.* [1, Thm 2.92]. Restricted to effects, we get a natural isomorphism $\mathrm{Pred} = [0,1]_{(-)} \Rightarrow [0,1]_{((-)_*)^*}$. Since a bounded linear functional $f\colon \mathscr{A}_* \to \mathbb{C}$ on the pre-dual $\mathscr{A}_*$ of a von Neumann algebra $\mathscr{A}$ is completely determined by its action on the states of $\mathscr{A}$, and this action is affine, contractive and maps into $[0,1]$ when f is from $[0,1]_{(\mathscr{A}_*)^*}$, restricting such f to states gives a natural isomorphism $[0,1]_{((-)_*)^*} \to \mathbf{ConvCMet}(\mathrm{Stat}(-),[0,1])$. Composing this with the natural isomorphism mentioned before we get a natural isomorphism $\mathrm{Pred} \Rightarrow \mathcal{F}\mathrm{Stat}$.

With these isomorphisms $\mathcal{G}\mathrm{Pred} \cong \mathrm{Stat}$ and $\mathrm{Pred} \cong \mathcal{F}\mathrm{Stat}$ in place we can show that the functor $\mathrm{Stat}\colon \mathbf{vNA}^{\mathrm{op}} \to \mathbf{ConvCMet}$ is full and faithful, since for two von Neumann algebras $\mathscr{A}$ and $\mathscr{B}$ we have:

$$\begin{aligned}\mathbf{ConvCMet}\Big(\mathrm{Stat}(\mathscr{A}), \mathrm{Stat}(\mathscr{B})\Big) &\cong \mathbf{ConvCMet}\Big(\mathrm{Stat}(\mathscr{A}), \mathcal{G}\mathrm{Pred}(\mathscr{B})\Big) \\ &\cong \mathbf{DcEMod}^{\mathrm{op}}\Big(\mathcal{F}\mathrm{Stat}(\mathscr{A}), \mathrm{Pred}(\mathscr{B})\Big) \\ &\cong \mathbf{DcEMod}^{\mathrm{op}}\Big(\mathrm{Pred}(\mathscr{A}), \mathrm{Pred}(\mathscr{B})\Big) \\ &\cong \mathbf{vNA}^{\mathrm{op}}\Big(\mathscr{A}, \mathscr{B}\Big). \qquad \square\end{aligned}$$

5. Concluding remarks

In this paper we have given a systematic unifying description of metrics on states and predicates from the perspective of the duality between state transformers and predicate transformers, notably in state-and-effect triangles. This unifying perspective is most prominent in the use of 'validity' metrices, both on states (via joins over predicates) and on predicates (via joins over states).

We have concentrated on the *discrete* version of classical probability and on quantum probability. What about *continuous* classical probability? Most of it has already been done in [5], see also [8], albeit in slightly different form, using ω-complete ordered cones instead of directed complete effect modules, together with a 'cone duality' result of the form $\mathrm{Hom}(L_p^+(X,\mu), \mathbb{R}_{\geq 0}) \cong L_q^+(X,\mu)$ when $\frac{1}{p}+\frac{1}{q}=1$; here, X is a measurable space with measure μ. In the language of triangles, this duality corresponds to commutation of the triangles, as in the above Propositions 19 and 20. In a next step, as in [8], a category of 'kernels' can be formed, as comma category $(1 \downarrow \mathbf{B})$ of the base category $\mathbf{B}$ that we use in triangles. For instance, the comma category $(1 \downarrow \mathcal{K}\ell(\mathcal{D}))$ contains distributions as objects, and distribution preserving maps between them. They can be used to define Bayesian inversion in the form of a dagger functor on such a comma category, see notable [8] — and [6] for a wider perspective on inversion and disintegration.

References

[1] E.M. Alfsen and F.W. Shultz. *State spaces of operator algebras*. Springer, 2012.

[2] J. de Bakker and E. Vink. *Control Flow Semantics*. MIT Press, Cambridge, MA, 1996.

[3] P. Baldan, F. Bonchi, H. Kerstan, and B. König. Behavioral metrics via functor lifting. In V. Raman and S. Suresh, editors, *FSTTCS 2014: Foundations of Software Technology and Theoretical Computer Science*, volume 29 of *LIPIcs*, pages 403–415. Schloss Dagstuhl, 2014.

[4] F. van Breugel, C. Hermida, M. Makkai, and J. Worrell. Recursively defined metric spaces without contraction. *Theor. Comp. Sci.*, 380:143–163, 2007.

[5] P. Chaput, V. Danos, P. Panangaden, and G. Plotkin. Approximating Markov processes by averaging. *Journ. ACM*, 61(1), 2014.

[6] K. Cho and B. Jacobs. Disintegration and Bayesian inversion, both abstractly and concretely. See `arxiv.org/abs/1709.00322`, 2017.

[7] K. Cho, B. Jacobs, A. Westerbaan, and B. Westerbaan. An introduction to effectus theory. see `arxiv.org/abs/1512.05813`, 2015.

[8] F. Clerc, F. Dahlqvist, V. Danos, and I. Garnier. Pointless learning. In J. Esparza and A. Murawski, editors, *Foundations of Software Science and Computation Structures*, number 10203 in Lect. Notes Comp. Sci., pages 355–369. Springer, Berlin, 2017.

[9] A. Dvurečenskij and S. Pulmannová. *New Trends in Quantum Structures*. Kluwer Acad. Publ., Dordrecht, 2000.

[10] A. Gibbs and F. Su. On choosing and bounding probability metrics. *Int. Statistical Review*, 70(3):419–435, 2002.

[11] B. Jacobs. Convexity, duality, and effects. In C. Calude and V. Sassone, editors, *IFIP Theoretical Computer Science 2010*, number 82(1) in IFIP Adv. in Inf. and Comm. Techn., pages 1–19. Springer, Boston, 2010.
[12] B. Jacobs. Probabilities, distribution monads, and convex categories. *Theor. Comp. Sci.*, 412(28):3323–3336, 2011.
[13] B. Jacobs. New directions in categorical logic, for classical, probabilistic and quantum logic. *Logical Methods in Comp. Sci.*, 11(3), 2015. See `https://lmcs.episciences.org/1600`.
[14] B. Jacobs. *Introduction to Coalgebra. Towards Mathematics of States and Observations.* Number 59 in Tracts in Theor. Comp. Sci. Cambridge Univ. Press, 2016.
[15] B. Jacobs. From probability monads to commutative effectuses. *Journ. of Logical and Algebraic Methods in Programming*, 156, 2017. See `http://dx.doi.org/10.1016/j.jlamp.2016.11.006`.
[16] B. Jacobs. A note on distances between probabilistic and quantum distributions. In A. Silva, editor, *Math. Found. of Programming Semantics*, Elect. Notes in Theor. Comp. Sci. Elsevier, Amsterdam, 2017.
[17] B. Jacobs. A recipe for state and effect triangles. *Logical Methods in Comp. Sci.*, 13(2), 2017. See `https://lmcs.episciences.org/3660`.
[18] B. Jacobs and J. Mandemaker. The expectation monad in quantum foundations. In B. Jacobs, P. Selinger, and B. Spitters, editors, *Quantum Physics and Logic (QPL) 2011*, number 95 in Elect. Proc. in Theor. Comp. Sci., pages 143–182, 2012.
[19] B. Jacobs, J. Mandemaker, and R. Furber. The expectation monad in quantum foundations. *Information & Computation*, 250:87–114, 2016.
[20] B. Jacobs and A. Westerbaan. An effect-theoretic account of Lebesgue integration. In D. Ghica, editor, *Math. Found. of Programming Semantics*, number 319 in Elect. Notes in Theor. Comp. Sci., pages 239–253. Elsevier, Amsterdam, 2015.
[21] B. Jacobs and F. Zanasi. A formal semantics of influence in bayesian reasoning. In K. Larsen, H. Bodlaender, and J.-F. Raskin, editors, *Math. Found. of Computer Science*, volume 83 of *LIPIcs*. Schloss Dagstuhl, 2017.
[22] R. Kadison. Operator algebras with a faithful weakly-closed representation. *Annals of Mathematics*, 64(1):175–181, 1956.
[23] R.V. Kadison and J.R. Ringrose. *Fundamentals of the Theory of Operator Algebras*. American Mathematical Society, 2015.
[24] T. Lindvall. *Lectures on the Coupling Method*. John Wiley & Sons, New York, 1987.
[25] R. Madare, P. Panangaden, and G. Plotkin. Quantitative algebraic reasoning. In *Logic in Computer Science*. IEEE, Computer Science Press, 2016.
[26] A. McIver, C. Morgan, and T. Rabehaja. Abstract hidden Markov models: A monadic account of quantitative information flow. In *Logic in Computer Science*, pages 597–608. IEEE, Computer Science Press, 2015.
[27] M. Nielsen and I. Chuang. *Quantum Computation and Quantum Information*. Cambridge Univ. Press, 2000.
[28] E. Rieffel and W. Polak. *Quantum Computing. A Gentle Introduction*. MIT Press, Cambridge, MA, 2011.
[29] C. Villani. *Optimal Transport — Old and New*. Springer, Berlin Heidelberg, 2009.
[30] N.S. Yanovsky and M.A. Manucci. *Quantum Computing for Computer Scientists*. Cambridge Univ. Press, 2008.

Appendix A. Missing proofs from Section 2

Proof of Proposition 3 Let $\omega_1, \omega_2 \in \mathcal{D}(X)$ be two discrete probability distributions on the same set X. Recall from (2.1) that by definition: $\mathsf{tvd}\big(\omega_1, \omega_2\big) = \frac{1}{2}\sum_{x\in X} \big|\,\omega_1(x) - \omega_2(x)\,\big|$. We will prove the two inequalities labeled (a) and (b) in:

$$\mathsf{tvd}\big(\omega_1, \omega_2\big) \overset{(a)}{\leq} \max_{U\subseteq X} \omega_1 \models \mathbf{1}_U - \omega_2 \models \mathbf{1}_U \leq \bigvee_{p\in[0,1]^X} \Big|\,\omega_1 \models p - \omega_2 \models p\,\Big| \overset{(b)}{\leq} \mathsf{tvd}\big(\omega_1, \omega_2\big).$$

This proves Proposition 3 since the inequality in the middle is trivial.

We start with some preparatory definitions. Let $U \subseteq X$ be an arbitrary subset. Recall that we write $\omega_i(U) = \sum_{x\in U}\omega_i(x) = \omega \models \mathbf{1}_U$. We partition U in three disjoint parts, and take the relevant sums:

$$\begin{cases} U_> = \{x \in U \mid \omega_1(x) > \omega_2(x)\} \\ U_= = \{x \in U \mid \omega_1(x) = \omega_2(x)\} \\ U_< = \{x \in U \mid \omega_1(x) < \omega_2(x)\} \end{cases} \qquad \begin{cases} U\!\uparrow = \omega_1(U_>) - \omega_2(U_>) \geq 0 \\ U\!\downarrow = \omega_2(U_<) - \omega_1(U_<) \geq 0. \end{cases}$$

We use this notation in particular for $U = X$. In that case we can use:

$$\begin{aligned} 1 &= \omega_1(X) = \omega_1(X_>) + \omega_1(X_=) + \omega_1(X_<) \\ 1 &= \omega_2(X) = \omega_2(X_>) + \omega_2(X_=) + \omega_2(X_<) \end{aligned}$$

Hence by subtraction we obtain, since $\omega_1(X_=) = \omega_2(X_=)$,

$$0 = \big(\omega_1(X_>) - \omega_2(X_>)\big) + \big(\omega_1(X_<) - \omega_2(X_<)\big)$$

That is,

$$X\!\uparrow = \omega_1(X_>) - \omega_2(X_>) = \omega_2(X_<) - \omega_1(X_<) = X\!\downarrow .$$

As a result:

$$\begin{aligned} \mathsf{tvd}\big(\omega_1, \omega_2\big) &= \tfrac{1}{2}\sum_{x\in X}\big|\,\omega_1(x) - \omega_2(x)\big| \\ &= \tfrac{1}{2}\Big(\sum_{x\in X_>}\big(\omega_1(x) - \omega_2(x)\big) + \sum_{x\in X_<}\big(\omega_2(x) - \omega_1(x)\big)\Big) \\ &= \tfrac{1}{2}\Big(\big(\omega_1(X_>) - \omega_2(X_>)\big) + \big(\omega_2(X_<) - \omega_1(X_<)\big)\Big) \\ &= \tfrac{1}{2}\big(X\!\uparrow + X\!\downarrow\big) \\ &= X\!\uparrow \end{aligned} \tag{A.1}$$

We have prepared the ground for proving the above inequalities (a) and (b).

(a) We'll see that the above maximum is actually reached for the subset $U = X_>$, first of all because:

$$\begin{aligned} \mathsf{tvd}\big(\omega_1, \omega_2\big) \overset{(A.1)}{=} X\!\uparrow = \omega_1(X_>) - \omega_2(X_>) &= \omega_1 \models \mathbf{1}_{X_>} - \omega_2 \models \mathbf{1}_{X_>} \\ &\leq \max_{U\subseteq X} \omega_1 \models \mathbf{1}_U - \omega_2 \models \mathbf{1}_U. \end{aligned}$$

(b) Let $p \in [0,1]^X$ be an arbitrary predicate. We write $\mathbf{1}_U \,\&\, p$ for the pointwise product predicate, with: $\big(\mathbf{1}_U \,\&\, p\big) = \mathbf{1}_U(x)\cdot p(x)$, which is $p(x)$ if $x \in U$ and 0 otherwise.

Then:

$$\begin{array}{rl}
& \big|\omega_1 \models p - \omega_2 \models p\big| \\
= & \Big|\big(\omega_1 \models \mathbf{1}_{X_>} \,\&\, p \;+\; \omega_1 \models \mathbf{1}_{X_=} \,\&\, p \;+\; \omega_1 \models \mathbf{1}_{X_<} \,\&\, p\big) \\
& \quad -\, \big(\omega_2 \models \mathbf{1}_{X_>} \,\&\, p \;+\; \omega_2 \models \mathbf{1}_{X_=} \,\&\, p \;+\; \omega_2 \models \mathbf{1}_{X_<} \,\&\, p\big)\Big| \\
= & \Big|\big(\omega_1 \models \mathbf{1}_{X_>} \,\&\, p \,-\, \omega_2 \models \mathbf{1}_{X_>} \,\&\, p\big) \;-\; \big(\omega_2 \models \mathbf{1}_{X_<} \,\&\, p \,-\, \omega_1 \models \mathbf{1}_{X_<} \,\&\, p\big)\Big| \\
= & \begin{cases} \big(\omega_1 \models \mathbf{1}_{X_>} \,\&\, p \,-\, \omega_2 \models \mathbf{1}_{X_>} \,\&\, p\big) \;-\; \big(\omega_2 \models \mathbf{1}_{X_<} \,\&\, p \,-\, \omega_1 \models \mathbf{1}_{X_<} \,\&\, p\big) \\ \quad \text{if } \big(\omega_1 \models \mathbf{1}_{X_>} \,\&\, p \,-\, \omega_2 \models \mathbf{1}_{X_>} \,\&\, p\big) \overset{(*)}{\geq} \big(\omega_2 \models \mathbf{1}_{X_<} \,\&\, p \,-\, \omega_1 \models \mathbf{1}_{X_<} \,\&\, p\big) \\ \big(\omega_2 \models \mathbf{1}_{X_<} \,\&\, p \,-\, \omega_1 \models \mathbf{1}_{X_<} \,\&\, p\big) \;-\; \big(\omega_1 \models \mathbf{1}_{X_>} \,\&\, p \,-\, \omega_2 \models \mathbf{1}_{X_>} \,\&\, p\big) \\ \qquad \text{otherwise} \end{cases} \\
\leq & \begin{cases} \omega_1 \models \mathbf{1}_{X_>} \,\&\, p \,-\, \omega_2 \models \mathbf{1}_{X_>} \,\&\, p & \text{if } (*) \\ \omega_2 \models \mathbf{1}_{X_<} \,\&\, p \,-\, \omega_1 \models \mathbf{1}_{X_<} \,\&\, p & \text{otherwise} \end{cases} \\
= & \begin{cases} \sum_{x\in X_>} (\omega_1(x) - \omega_2(x)) \cdot p(x) & \text{if } (*) \\ \sum_{x\in X_<} (\omega_2(x) - \omega_1(x)) \cdot p(x) & \text{otherwise} \end{cases} \\
\leq & \begin{cases} \sum_{x\in X_>} \omega_1(x) - \omega_2(x) & \text{if } (*) \\ \sum_{x\in X_<} \omega_2(x) - \omega_1(x) & \text{otherwise} \end{cases} \\
= & \begin{cases} X{\uparrow} & \text{if } (*) \\ X{\downarrow} = X{\uparrow} & \text{otherwise} \end{cases} \\
= & X{\uparrow} \\
\overset{(A.1)}{=} & \mathsf{tvd}\big(\omega_1, \omega_2\big).
\end{array}$$

□

Proof of Proposition 11 Let ϱ_1, ϱ_2 be two states (density operators) of a Hilbert space $\mathscr{H}$. The trick is to split the trace-class operator $\varrho \coloneqq \varrho_1 - \varrho_2$ into its positive and negative parts: we have $\varrho = \varrho_+ - \varrho_-$, where $\varrho_+, \varrho_- \colon \mathscr{H} \to \mathscr{H}$ are positive operators with $\varrho_+\varrho_- = 0$ and $|\varrho| = \varrho_+ + \varrho_-$, see [1, Cor 2.15]. Note that since $\varrho_+, \varrho_- \leq |\varrho|$ the operators ϱ_+ and ϱ_- are trace-class as well. The key is to note that $\mathrm{tr}(\varrho_+) - \mathrm{tr}(\varrho_-) = \mathrm{tr}(\varrho) = \mathrm{tr}(\varrho_1) - \mathrm{tr}(\varrho_2) = 1 - 1 = 0$, so that $\mathrm{tr}(\varrho_+) = \mathrm{tr}(\varrho_-)$. Hence:

$$\mathsf{trd}(\varrho_1, \varrho_2) \overset{(2.6)}{=} \tfrac{1}{2}\mathrm{tr}(|\varrho|) = \tfrac{1}{2}\big(\mathrm{tr}(\varrho_+) + \mathrm{tr}(\varrho_-)\big) = \mathrm{tr}(\varrho_+) = \mathrm{tr}(\varrho_-).$$

Now, given an effect p on $\mathscr{H}$ we have $\varrho_1 \models p - \varrho_2 \models p = \mathrm{tr}(\varrho_1\, p) - \mathrm{tr}(\varrho_2\, p) = \mathrm{tr}(\varrho\, p) = \mathrm{tr}(\varrho_+\, p) - \mathrm{tr}(\varrho_-\, p) \leq \mathrm{tr}(\varrho_+\, p) \leq \mathrm{tr}(\varrho_+) = \mathsf{trd}(\varrho_1, \varrho_2)$, using $p \leq \mathrm{id}$. Since similarly $\varrho_2 \models p - \varrho_1 \models p \leq \mathsf{trd}(\varrho_1, \varrho_2)$, we get:

$$\bigvee_{p\in\mathcal{E}\!f(\mathscr{H})} \big|\, \varrho_2 \models p - \varrho_1 \models p \,\big| \leq \mathsf{trd}(\varrho_1, \varrho_2).$$

The only thing that remains to be shown is that there is a projection s on $\mathscr{H}$ with $\varrho_1 \models s - \varrho_2 \models s = \mathsf{trd}(\varrho_1, \varrho_2)$. It turns out that we need to pick the least projection s in $\mathscr{B}(\mathscr{H})$ with $s\varrho_+ = \varrho_+$ (which exists, see *e.g.* [1, Defn 2.107]). If t denotes the least projection with $\varrho_- t = \varrho_-$ then one can prove that $st = 0$, so that $\varrho_- s = \varrho_- ts = 0$. Whence $\varrho_1 \models s - \varrho_2 \models s = \mathrm{tr}(\varrho_1 s) - \mathrm{tr}(\varrho_2 s) = \mathrm{tr}(\varrho s) = \mathrm{tr}(\varrho_+ s) - \mathrm{tr}(\varrho_- s) = \mathrm{tr}(\varrho_+) = \mathsf{trd}(\varrho_1, \varrho_2)$. □

Proof of Proposition 12 Let $\varrho_1, \varrho_2 \colon \mathscr{A} \to \mathbb{C}$ be two normal states of a von Neumann algebra $\mathscr{A}$ and let $e \in [0,1]_{\mathscr{A}}$ be an arbitrary effect. If we bluntly apply the definition of the operator norm we only get $|\varrho_1 \models e - \varrho_2 \models e| = |(\varrho_1 - \varrho_2)(e)| \leq \|\varrho_1 - \varrho_2\|_{\mathrm{op}} \cdot \|e\| \leq \|\varrho_1 - \varrho_2\|_{\mathrm{op}}$. The factor "$\frac{1}{2}$" from Proposition 12 is then missing, so a more subtle approach is called for. Writing $\varrho \coloneqq \varrho_1 - \varrho_2$ there is by [23, Thm 7.4.7] a sharp predicate $s \in [0,1]_A$ such that both $\varrho_+ \coloneqq \varrho(s(\,\cdot\,)s)$ and $\varrho_- \coloneqq -\varrho(s^\perp(\,\cdot\,)s^\perp)$ are positive and normal, and, moreover,

$$\varrho = \varrho_+ - \varrho_- \qquad \text{and} \qquad \|\varrho\|_{\mathrm{op}} = \|\varrho_+\|_{\mathrm{op}} + \|\varrho_-\|_{\mathrm{op}}.$$

Further, by [23, Thm 4.3.2] we have $\|\varrho_1\|_{\mathrm{op}} = \varrho_1(1)$ and $\|\varrho_2\|_{\mathrm{op}} = \varrho_2(1)$. Then since ϱ_1 and ϱ_2 are states, we have $\varrho(1) = \varrho_1(1) - \varrho_2(1) = 1 - 1 = 0$, so $\varrho_+(1) - \varrho_-(1) = \varrho(1) = 0$, and thus $\|\varrho_+\|_{\mathrm{op}} = \varrho_+(1) = \varrho_-(1) = \|\varrho_-\|_{\mathrm{op}}$. But then, since $\|\varrho\|_{\mathrm{op}} = \|\varrho_+\|_{\mathrm{op}} + \|\varrho_-\|_{\mathrm{op}}$, we get:

$$\|\varrho_+\|_{\mathrm{op}} = \|\varrho_-\|_{\mathrm{op}} = \tfrac{1}{2}\|\varrho_1 - \varrho_2\|_{\mathrm{op}}.$$

Now, given $e \in [0,1]_A$ we have $\varrho_1 \models e - \varrho_2 \models e = \varrho(e) \leq \varrho_+(e) \leq \varrho_+(1) \leq \|\varrho_+\|_{\mathrm{op}} = \frac{1}{2}\|\varrho_1 - \varrho_2\|_{\mathrm{op}}$, and so $\bigvee_{e \in [0,1]_{\mathscr{A}}} \varrho_1 \models e - \varrho_2 \models e \leq \frac{1}{2}\|\varrho_1 - \varrho_2\|_{\mathrm{op}}$. By a similar reasoning, we get $\bigvee_{e \in [0,1]_A} \varrho_2 \models e - \varrho_1 \models e \leq \frac{1}{2}\|\varrho_1 - \varrho_2\|_{\mathrm{op}}$, and so:

$$\bigvee_{e \in [0,1]_{\mathscr{A}}} \big|\, \varrho_1 \models e - \varrho_1 \models e \,\big| \leq \tfrac{1}{2}\|\varrho_1 - \varrho_2\|_{\mathrm{op}}.$$

The only real thing left to prove is that $\frac{1}{2}\|\varrho_1 - \varrho_2\|_{\mathrm{op}} = \varrho_1(s) - \varrho_2(s)$, for the above sharp predicate s, because all the equalities in Proposition 12 follow trivially from it. Since $\varrho_+ = \varrho(s(\,\cdot\,)s)$ we have $\varrho_+(s) = \varrho(s) = \varrho_+(1) = \|\varrho_+\|_{\mathrm{op}} = \frac{1}{2}\|\varrho_1 - \varrho_2\|_{\mathrm{op}}$; and since $\varrho_- = -\varrho(s^\perp(\,\cdot\,)s^\perp)$ we have $\varrho_-(s) = -\varrho(s^\perp s s^\perp) = -\varrho(0) = 0$. Whence $\varrho_1(s) - \varrho_2(s) = \varrho(s) = \varrho_+(s) - \varrho_-(s) = \varrho_+(s) = \frac{1}{2}\|\varrho_1 - \varrho_2\|_{\mathrm{op}}$. $\square$