A View on Optimal Transport from Noncommutative Geometry*

Francesco D'ANDREA † *and Pierre MARTINETTI* ‡

† *Ecole de Mathématique, Univ. Catholique de Louvain, Chemin du Cyclotron 2, 1348 Louvain-La-Neuve, Belgium*
E-mail: *francesco.dandrea@uclouvain.be*

‡ *Institut für Theoretische Physik, Universität Göttingen, Friedrich-Hund-Platz 1, 37077 Göttingen, Germany*
E-mail: *martinetti@theorie.physik.uni-goettingen.de*

Received April 14, 2010, in final form July 08, 2010; Published online July 20, 2010
doi:10.3842/SIGMA.2010.057

Abstract. We discuss the relation between the Wasserstein distance of order 1 between probability distributions on a metric space, arising in the study of Monge–Kantorovich transport problem, and the spectral distance of noncommutative geometry. Starting from a remark of Rieffel on compact manifolds, we first show that on any – i.e. non-necessary compact – complete Riemannian spin manifolds, the two distances coincide. Then, on convex manifolds in the sense of Nash embedding, we provide some natural upper and lower bounds to the distance between any two probability distributions. Specializing to the Euclidean space $\mathbb{R}^n$, we explicitly compute the distance for a particular class of distributions generalizing Gaussian wave packet. Finally we explore the analogy between the spectral and the Wasserstein distances in the noncommutative case, focusing on the standard model and the Moyal plane. In particular we point out that in the two-sheet space of the standard model, an optimal-transport interpretation of the metric requires a cost function that does not vanish on the diagonal. The latest is similar to the cost function occurring in the relativistic heat equation.

Key words: noncommutative geometry; spectral triples; transport theory

2010 Mathematics Subject Classification: 58B34; 82C70

1 Introduction

The idea at the core of Noncommutative Geometry [11] is the observation that, in many interesting cases, the description of a space as a set of points is inadequate. Think for example of quantum mechanics, where position and momentum are replaced by non-commuting operators: as a consequence, Heisenberg uncertainty relations impose limitations in the precision of their simultaneous measurement so that the notion of "point in phase space" loses any operational meaning. Taking into account General Relativity, one can show by simple arguments that not only phase-space coordinates but also space-time coordinates should be non-commutative (cf. [17, 18] and references therein) making the concept of "points in space-time" also problematic. Noncommutative Geometry provides efficient tools to study these "spaces" that are no longer described by a commutative algebra of coordinate functions, but by some noncommutative operator algebra $\mathcal{A}$.

Losing the notion of points, one also loses the notion of distance between points. However one can still define a distance between *states* of the algebra $\mathcal{A}$, for example with the help of

*This paper is a contribution to the Special Issue "Noncommutative Spaces and Fields". The full collection is available at http://www.emis.de/journals/SIGMA/noncommutative.html a generalized Dirac operator D. The latter is the starting point of Connes theory of *spectral triples* [12], which is the datum $(\mathcal{A}, \mathcal{H}, D)$ of an involutive (non necessarily commutative) algebra $\mathcal{A}$, a representation π of $\mathcal{A}$ as bounded operators on a Hilbert space $\mathcal{H}$ and a self-adjoint operator D, such that $[D, a]$ is bounded and $a(D-\lambda)^{-1}$ is compact for any $a \in \mathcal{A}$ and $\lambda \notin \mathrm{Sp}(D)$ (where the symbol π is omitted). The *spectral distance* between two states φ_1, φ_2 of $\mathcal{A}$ is defined as [10, 13]

$$d_D(\varphi_1, \varphi_2) \doteq \sup_{a\in\mathcal{A}} \big\{ |\varphi_1(a) - \varphi_2(a)|;\ ||[D, a]||_{\mathrm{op}} \leq 1 \big\}, \tag{1.1}$$

where the norm is the operator norm coming from the representation of $\mathcal{A}$ on $\mathcal{H}$. It is easy to check that (1.1) defines a distance in a strict mathematical sense except that it may be infinite.

Recall that a state is by definition a positive linear application $\bar{\mathcal{A}} \to \mathbb{C}$ with norm 1, where $\bar{\mathcal{A}}$ is the C^*-algebra completion of $\mathcal{A}$. When $\mathcal{A}$ is the algebra of observables of a physical system, this notion of state coincides with physicist's intuition, namely density matrices or Gibbs states in statistical physics, state vectors or wave functions in quantum mechanics. For the commutative algebra $C_0^\infty(\mathcal{M})$ – i.e. complex smooth functions vanishing at infinity on some spin manifold $\mathcal{M}$ of dimension[1] m – a canonical spectral triple is

$$\big(C_0^\infty(\mathcal{M}), L^2(\mathcal{M}, \mathcal{S}), D\big), \tag{1.2}$$

where $L^2(\mathcal{M}, \mathcal{S})$ is the Hilbert space of square integrable spinors on which $C_0^\infty(\mathcal{M})$ acts by pointwise multiplication and D is the Dirac operator. The latest is given in local coordinates by

$$D = -i \sum_{\mu=1}^{m} \gamma^\mu \nabla_\mu,$$

where $\nabla_\mu = \partial_\mu + \omega_\mu$ is the covariant derivative associated to the spin connection 1-form ω^μ and $\{\gamma^\mu\}_{\mu=1,2,\dots,m}$ are the self-adjoint Dirac gamma matrices satisfying

$$\gamma^\mu\gamma^\nu + \gamma^\nu\gamma^\mu = 2g^{\mu\nu}\mathbb{I} \tag{1.3}$$

with g the Riemannian metric of $\mathcal{M}$ and $\mathbb{I}$ the identity matrix of dimension 2^n if $m = 2n$ or $2n+1$. In this case $\bar{\mathcal{A}} = C_0(\mathcal{M})$ and the spectral distance between pure states (i.e. states that cannot be written as a convex sum of other states), which by Gelfand transform are in one-to-one correspondence with the points of the manifold

$$x \mapsto \delta_x :\ \delta_x(f) \doteq f(x) \qquad \forall\, f \in C_0(\mathcal{M}), \tag{1.4}$$

coincides [15] with the geodesic distance d of $\mathcal{M}$,

$$d_D(\delta_x, \delta_y) = d(x, y) \qquad \forall\, x, y \in \mathcal{M}. \tag{1.5}$$

In a noncommutative framework it is tempting – inspired by (1.4) – to take the set $\mathcal{P}(\mathcal{A})$ of pure states of $\mathcal{A}$ as the noncommutative analogue of points and d_D as natural generalization of the geodesic distance. This idea has been tested in several examples inspired by physics: finite-dimensional algebras [5, 26], functions on $\mathcal{M}$ with value in a matrix algebra [35, 32, 33] encoding the inner structure of the space-time of the standard model of particles physics [8], non-commutative deformations of $C_0^\infty(\mathcal{M})$ [7]. Most often the computation of the supremum in (1.1) is quite involved, however several explicit results have been obtained. They all indicate that as soon as $\mathcal{A}$ is noncommutative one looses an important feature of the geodesic distance,

[1] Unless otherwise specified, in all the paper we assume that "spin manifolds" are Riemannian, finite-dimensional, connected, complete, without boundary.

namely $\mathcal{P}(\mathcal{A})$ is no longer a path metric space [25]. Explicitly there is no curve $[0,1] \ni t \mapsto \varphi_t \in \mathcal{P}(\mathcal{A})$ such that

$$d_D(\varphi_s, \varphi_t) = |t-s| d_D(\varphi_0, \varphi_1), \tag{1.6}$$

not even a sequence of such curves φ^n such that

$$d_D(\varphi_0, \varphi_1) = \operatorname*{Inf}_n \{\text{length of } \varphi^n \text{ between } \varphi_0 \text{ and } \varphi_1\}.$$

This can be seen on a simple noncommutative examples studied in [26], based on $\mathcal{A} = M_2(\mathbb{C})$ acting on $\mathbb{C}^2$ with $D = D^* \in M_2(\mathbb{C})$. $\mathcal{P}(M_2(\mathbb{C}))$ is weak* homeomorphic to the Euclidean 2-sphere. The image under this homeomorphism of the pure states $\omega_i = (\psi_i, .\psi_i)$, $i = 1, 2$, defined by the eigenvectors ψ_i of D are antipodal, and determine a distinguished vertical axis on S^2. The spectral distance is invariant by rotation around this axis and the connected components (see (2.16) below) are circles parallel to the horizontal plane. Explicitly, on the circle with radius $r \in [0,1]$ one finds

$$d_D(\theta_1, \theta_2) = \frac{2r}{|D_1 - D_2|} \left| \sin \frac{\theta_1 - \theta_2}{2} \right|, \tag{1.7}$$

where $\theta \in [0, 2\pi[$ is the azimuth and D_i are the eigenvalues of D. One then checks[2] that (1.6) has no solution in $\mathcal{P}(M_2(\mathbb{C}))$ [34]. The lack of geodesic curves (1.6) within $\mathcal{P}(\mathcal{A})$ is cured by considering non-pure states. Indeed (1.1) not only generalizes the geodesic distance to noncommutative algebras, it also extends the distance to objects that are not equivalent to points, namely non-pure states. Noticing that (1.7) is the geodesic distance within the Euclidean disk of radius $\frac{2r}{|D_1-D_2|}$ and that the set $\mathcal{S}(M_2(\mathbb{C}))$ of states of $M_2(\mathbb{C})$ is homeomorphic to the 2-dimensional Euclidean ball (see Section 4.1), one easily obtains a curve in $\mathcal{S}(M_2(\mathbb{C}))$ satisfying (1.6), namely

$$\varphi_t = (1-t)\varphi_0 + t\varphi_1.$$

This remains true in full generality since, whatever algebra $\mathcal{A}$ and operator D,

$$d_D(\varphi_s, \varphi_t) = \sup_{a \in \mathcal{A}} \left\{ |(s-t)(\varphi_0 - \varphi_1)(a)|;\ ||[D,a]||_{\text{op}} \leq 1 \right\} = |s-t| d_D(\varphi_0, \varphi_1). \tag{1.9}$$

In other terms, in view of (1.9) and (1.5), the spectral distance is a natural generalization of the geodesic distance to the noncommutative setting as soon as one takes into account the whole space of states, and not only its extremal points. This motivates the study of the spectral distance between non-pure states that we undertake in this paper. We begin by giving a detailed proof of Rieffel's remark [39] (also mentioned in [4]), according to which in the commutative case $\mathcal{A} = C_0^\infty(\mathcal{M})$ the spectral distance coincides with a distance well known in optimal transport theory, namely the Wasserstein distance W of order 1 (see the bibliographic notes of [44, Chapter 6]). We stress in particular that $\mathcal{M}$ has to be complete for that result to hold besides the compact case. Then we present few explicit calculations of d_D between non-pure states, and question on simple examples – including the standard model – the pertinence of the optimal transport interpretation of the spectral distance in a noncommutative framework. These are preliminaries

[2] A curve $t \mapsto \theta(t)$ satisfies (1.6) if and only if

$$\left|\sin \tfrac{\theta(t)-\theta(s)}{2}\right| = K|t-s| \tag{1.8}$$

for any $t, s \in [0,1]$ and K a constant. The right hand side of (1.8) being a function of $t-s$, there exists a function f such that the left hand side is $f(t-s)$. Putting $s = 0$, one obtains $f(t) = |\sin \frac{\theta(t)-\theta(0)}{2}|$. Reinserted in (1.8), this yields Cauchy's functional equation $\theta(t) - \theta(s) = \theta(t-s) - \theta(0)$, whose continuous solutions are linear. Since (1.8) has no linear solutions, this proves the claim.

results, intending to bring the attention of the transport theory community on the metric aspect of noncommutative geometry, and vice-versa.

Notice that some properties of the spectral distance between non-pure states have been investigated in [39]: considering instead of $||[D,a]||_{\rm op}$ an arbitrary semi-norm on $\mathcal{A}$, it is shown that the knowledge of the distance between pure states of a noncommutative $\mathcal{A}$ may not be enough to recover the semi-norm on $\mathcal{A}$. One also needs the distance between non-pure states. This suggests that the metric information encoded in formula (1.1) is not exhausted once one knows the distance between pure states.

The plan of the paper is the following. In Section 2 we recall some basics of transport theory and noncommutative geometry in order to establish – in Proposition 2.1 – the equality between d_D and W for any complete spin manifold $\mathcal{M}$. We also discuss various definitions of the spectral distance, characterize its connected components and emphasize the importance of the completeness condition. In Section 3 we provide some lower and upper bounds for the distance. Specializing to $\mathcal{M} = \mathbb{R}^n$, we explicitly compute the distance between a class of states generalizing Gaussian wave-packets. Section 4 deals with noncommutative examples. We show that on the truncations of the Moyal plane introduced in [7], the Wasserstein distance W_D with cost d_D defined on $\mathcal{P}(\mathcal{A})$ does not coincide with the spectral distance on $\mathcal{S}(\mathcal{A})$. But on almost-commutative geometries, including the standard model of elementary particles, the spectral distance between certain classes of states may be recovered as a Wasserstein distance W' with cost d' defined on a subset of $\mathcal{S}(\mathcal{A})$ containing $\mathcal{P}(\mathcal{A})$. We also point out a reformulation of the spectral distance on $\mathcal{P}(\mathcal{A})$ in term of the minimal work W_I associated to a cost c_I non-vanishing on the diagonal (i.e. $c_I(x,x) \neq 0$).

2 Spectral distance as Wasserstein distance of order 1

2.1 Spectral distance from Kantorovich duality

For any locally compact Hausdorff topological space $\mathcal{X}$, states $\varphi \in \mathcal{S}(C_0(\mathcal{X}))$ are given by Borel probability measures μ on $\mathcal{X}$, via the rule

$$\varphi(f) \doteq \int_{\mathcal{X}} f \mathrm{d}\mu \qquad \forall\, f \in \mathcal{A}. \tag{2.1}$$

This is a simple application of Riesz representation and Hahn–Banach theorems together with the assumption that $\mathcal{X}$ is σ-compact in order to avoid regularity problems. Any such μ defines a state since f vanishing at infinity (hence being bounded) guarantees that (2.1) is finite. Pure states correspond to Dirac-delta measures. To provide some physical intuition, one can view φ as a wave-packet and imagine that it describes the probability distribution of a bunch of particles. Strictly speaking a wave-packet is a square root of the Radon–Nikodym derivative of $\mathrm{d}\mu$ with respect to some fixed σ-finite positive measure $\mathrm{d}x$ on $\mathcal{X}$ (it is a square-integrable function, almost everywhere defined and unique modulo a phase, whenever $\mathrm{d}\mu$ is absolutely continuous with respect to $\mathrm{d}x$). For instance in quantum mechanics, with $\mathcal{X} = \mathbb{R}^n$ and $\mathrm{d}x$ the Lebesgue measure, a wave-packet is a function $\phi \in L^2(\mathbb{R}^n)$ and the corresponding measure is $\mathrm{d}\mu = |\phi(x)|^2 \mathrm{d}x$.

Assuming $\mathcal{X}$ is a metric space with distance function d, there is a natural way to measure how much two states φ_1 and φ_2 differ, which is the expectation value of the distance between the two corresponding distributions

$$\mathbb{E}(d; \mu_1 \times \mu_2) = \int_{\mathcal{X}\times\mathcal{X}} d(x,y) \mathrm{d}\mu_1(x) \mathrm{d}\mu_2(y),$$

where μ_i is associated to φ_i via (2.1). Other ways are suggested by transport theory (all material here is taken from [1, 43, 44, 6], where an extensive bibliography can be found. Following [43] we assume from now on that $\mathcal{X}$ has a countable basis, so that it is a Polish space). Assume there exists a positive real function $c(x,y)$ – the "cost function" – that represents the work needed to move from x to y. A good measure on how much the φ_i's differ is given by the minimal work W required to move the bunch of particles from the configuration φ_1 to the configuration φ_2, namely

$$W(\varphi_1,\varphi_2) \doteq \inf_\pi \int_{\mathcal{X}\times\mathcal{X}} c(x,y)\mathrm{d}\pi, \tag{2.2}$$

where the infimum is over all measures π on $\mathcal{X}\times\mathcal{X}$ with marginals μ_1, μ_2 (i.e. the push-forwards of π through the projections $\mathbb{X},\mathbb{Y}:\mathcal{X}\times\mathcal{X}\to\mathcal{X}$, $\mathbb{X}(x,y)\doteq x$, $\mathbb{Y}(x,y)\doteq y$, are $\mathbb{X}_*(\pi)=\mu_1$ and $\mathbb{Y}_*(\pi)=\mu_2$). Such measures are called *transportation plans*. Finding the optimal transportation plan, that is the one which minimizes W, is a non-trivial question known as the Monge–Kantorovich problem. This is a generalization of Monge [36] "déblais et remblais" problem, where one considers only those transportation plans that are supported on the graph of a *transportation map*, i.e. a map $T:\mathcal{X}\to\mathcal{X}$ such that $T_*\mu_1=\mu_2$. Namely,

$$W_{\text{Monge}}(\varphi_1,\varphi_2) \doteq \inf_T \int_{\mathcal{X}} c(x,T(x))\mathrm{d}\mu_1(x). \tag{2.3}$$

One of the interests of Kantorovich's generalization [28] is that the infimum in (2.3) is not always a minimum: an optimal transportation map may not exist. On the contrary the infimum in (2.2) is a minimum and always coincides with Monge infimum, even when the optimal transportation map does not exist. Moreover when the cost function c is a distance d, (2.2) is in fact a distance on the space of states – with the infinite value allowed – called the Kantorovich–Rubinstein distance (this case was first studied in [29]). To be sure it remains finite (see (2.13) below), it is convenient to restrict to the set $\mathcal{S}_1(C_0(\mathcal{X}))$ of states whose moment of order 1 is finite, that is those distributions μ such that

$$\mathbb{E}(d(x_0,\circ);\mu) = \int_{\mathcal{X}} d(x_0,x)\mathrm{d}\mu(x) < +\infty, \tag{2.4}$$

where x_0 is an arbitrary but fixed point in $\mathcal{X}$. Note that as soon as $\mathbb{E}(d(x_0,\circ);\mu)$ is finite for x_0, then by the triangle inequality $\mathbb{E}(d(x,\circ);\mu)\leq\mathbb{E}(d(x_0,\circ);\mu)+d(x_0,x)$ is finite for any $x\in\mathcal{X}$ so that $S_1(C_0(\mathcal{X}))$ is independent on the choice of x_0. The Kantorovich–Rubinstein distance is also known as the *Wasserstein distance of order* 1. The distance of order p is given by a similar formula with $\mathbb{E}(d^p;\mu_1\otimes\mu_2)^{1/p}$, $1\leq p<\infty$, but in this paper we are interested only in the distance of order one and we shall simply call W the Wasserstein distance.

The link with non-commutative geometry, which seems to have been first noted for $\mathcal{M}$ compact in [39], is the following: when $\mathcal{X}$ is a spin manifold $\mathcal{M}$, the Wasserstein distance with cost function the geodesic distance d is nothing but the spectral distance (1.1) associated to the canonical spectral triple (1.2). This is *a priori* not difficult to see: On one side a central result of transport theory, Kantorovich duality, provides a dual formulation of the Wasserstein distance as a supremum instead of an infimum, namely (cf. Theorem 5.10 and equation (5.11) of [44])

$$W(\varphi_1,\varphi_2) = \sup_{||f||_{\text{Lip}}\leq 1,\, f\in L^1(\mu_1)\cap L^1(\mu_2)} \left(\int_{\mathcal{X}} f\mathrm{d}\mu_1 - \int_{\mathcal{X}} f\mathrm{d}\mu_2\right) \tag{2.5}$$

for any pair of states in $S(C_0(\mathcal{X}))$ such that the right-hand side in the above expression is finite. The supremum is on all real $\mu_{i=1,2}$-integrable functions f that are 1-Lipschitz, that is to say

$$|f(x)-f(y)|\leq d(x,y) \qquad \forall\, x,y\in\mathcal{X}.$$

On the other side, the commutator $-i[\gamma^\mu \partial_\mu, f]$ (where we use Einstein summation convention and sum over repeated indices) acts on $L^2(\mathcal{M}, \mathcal{S})$ as multiplication by $-i\gamma^\mu \partial_\mu f$. Moreover the supremum in (1.1) can be searched on self-adjoint elements [26], that for $\mathcal{A} = C_0^\infty(\mathcal{M})$ simply means real functions f. Thus

$$\begin{aligned} \|[D,f]\|_{\mathrm{op}}^2 &= \|\gamma^\mu \partial_\mu f\|_{\mathrm{op}}^2 = \|(\gamma^\mu \partial_\mu f)(\gamma^\nu \partial_\nu f)\|_{\mathrm{op}} = \|\tfrac{1}{2}(\gamma^\mu\gamma^\nu + \gamma^\nu\gamma^\mu)\partial_\mu f \partial_\nu f\|_{\mathrm{op}} \\ &= \|\mathbb{I} g^{\mu\nu}\partial_\mu f \partial_\nu f\|_{\mathrm{op}} = \|g^{\mu\nu}\partial_\mu f \partial_\nu f\|_\infty = \|f\|_{\mathrm{Lip}}^2, \end{aligned} \tag{2.6}$$

where we used: the C^*-property of the norm, that $\partial_\mu f$ commutes with $\partial_\nu f$ and the γ's matrices, equation (1.3), that $\sqrt{g^{\mu\nu}\partial_\mu f \partial_\nu f}$ evaluated at x is the norm of the gradient $\nabla f(x)$ and $\sup_{x\in\mathcal{M}} \|\nabla f(x)\|_{T_xM} = \|f\|_{\mathrm{Lip}}$ by Cauchy's mean value theorem. Consequently the commutator-norm condition in the spectral distance formula yields on f the condition required in Kantorovich's dual formula. However one has to be careful that, although the φ_i's on the l.h.s. of (2.5) denote states of $C_0^\infty(\mathcal{M})$, the supremum on the r.h.s. includes functions non-vanishing at infinity. Therefore (1.1) equals (2.5) if and only if the supremum on 1-Lipschitz smooth functions vanishing at infinity is the same as the supremum on 1-Lipschitz continuous functions non-necessarily vanishing at infinity. That $C_0^\infty(\mathcal{M})$ is dense within $C_0(\mathcal{M})$ is well known, but it might be less known that continuous K-Lipschitz functions can be approximated by smooth K-Lipschitz functions. In the following we use this result – proved for finite-dimensional manifolds in [24] and extended to infinite dimension in [2] – in order to to prove Rieffel's remark, generalized to complete locally compact manifolds (e.g. complete finite-dimensional manifolds).

Proposition 2.1. *For any $\varphi_1, \varphi_2 \in \mathcal{S}(C_0^\infty(\mathcal{M}))$ with $\mathcal{M}$ a (complete, Riemannian, finite dimensional, connected, without boundary) spin manifold, one has*

$$W(\varphi_1, \varphi_2) = d_D(\varphi_1, \varphi_2).$$

Proof. i) It is well known [2] that on $\mathbb{R}^n$ K-Lipschitz functions are the uniform limit of smooth K-Lipschitz functions (for any $K \geq 0$). It may be less known that the same is true for any (finite-dimensional) Riemannian manifold $\mathcal{M}$, and for K-Lipschitz functions vanishing at infinity. Let us give a short proof of this result.

According to Theorem 1 in [2] (that is valid for separable Riemannian manifolds, so in particular for finite-dimensional manifolds), given a Lipschitz function f, for any continuous function $\varepsilon : \mathcal{M} \to \mathbb{R}^+$ and for any $r > 0$ there exists a smooth function $g_{\varepsilon,r} : \mathcal{M} \to \mathbb{R}$ such that

$$\|g_{\varepsilon,r}\|_{\mathrm{Lip}} \leq \|f\|_{\mathrm{Lip}} + r \qquad \text{and} \qquad |f(x) - g_{\varepsilon,r}(x)| \leq \varepsilon(x) \qquad \forall\, x \in \mathcal{M}.$$

As a corollary, if f is a K-Lipschitz function vanishing at infinity, we can fix a sequence ε_n of continuous functions vanishing at infinity and uniformly converging to zero, and a sequence of positive numbers r_n converging to zero in order to get a sequence of smooth functions $g_n : \mathcal{M} \to \mathbb{R}$ such that

$$\|g_n\|_{\mathrm{Lip}} \leq K + r_n \qquad \text{and} \qquad |f(x) - g_n(x)| \leq \varepsilon_n(x) \qquad \forall\, x \in \mathcal{M}.$$

Obviously g_n vanishes at infinity (since both f and ε_n vanish at infinity). Let us call $f_n := K(K + r_n)^{-1} g_n$. The f_n's are the required smooth K-Lipschitz functions vanishing at infinity that converge uniformly to f. Indeed

$$f - f_n = \frac{K}{K + r_n}(f - g_n) + \frac{r_n}{K + r_n} f$$

implies

$$\sup_x |f(x) - f_n(x)| \leq \frac{K}{K + r_n} \sup_x |\varepsilon_n(x)| + \frac{r_n}{K + r_n} \sup_x |f(x)|,$$

and the right hand side goes to zero for $n \to \infty$ since $\sup_x |\varepsilon_n(x)| \to 0$ and $r_n \to 0$ by assumption, while $\sup_x |f(x)|$ is finite since f is continuous vanishing at infinity.

ii) By (2.6) the supremum in (1.1) is on 1-Lipschitz smooth functions vanishing at infinity. By i) above any 1-Lipschitz function f in $C_0(\mathcal{M})$ can be uniformly approximated by smooth 1-Lipschitz functions f_n in $C_0^\infty(\mathcal{M})$. Since any state φ of $C_0(\mathcal{M})$ is continuous [3] with respect to the sup-norm, namely

$$\lim_{n\to+\infty} ||f - f_n||_\infty = 0 \quad \Longrightarrow \quad \lim_{n\to+\infty} \varphi(f_n) = \varphi(f),$$

the supremum in (1.1) can be equivalently searched on continuous functions and the spectral distance writes

$$d_D(\varphi_1, \varphi_2) = \sup_{f\in C_0(\mathcal{M},\mathbb{R}),\, ||f||_{\mathrm{Lip}}\leq 1} \left(\int_{\mathcal{M}} f \mathrm{d}\mu_1 - \int_{\mathcal{M}} f \mathrm{d}\mu_2 \right). \tag{2.7}$$

iii) In case $\mathcal{M}$ is compact $C_0(\mathcal{M},\mathbb{R})$ coincides with $C(\mathcal{M},\mathbb{R})$ and (2.7) equals (2.5), hence the result. In case $\mathcal{M}$ is only locally compact, $C_0(\mathcal{M},\mathbb{R}) \subset C(\mathcal{M},\mathbb{R})$ so that $d_D(\varphi_1,\varphi_2) \leq W(\varphi_1,\varphi_2)$. To get the equality, it is sufficient to show that to any 1-Lipschitz μ_i-integrable function $f \in C(\mathcal{M},\mathbb{R})$ is associated a sequence of functions $f_n \in C_0(\mathcal{M},\mathbb{R})$ such that

$$||f_n||_{\mathrm{Lip}} \leq 1 \tag{2.8}$$

and

$$\lim_{n\to+\infty} \Delta(f_n) = \Delta(f), \tag{2.9}$$

where $\Delta(f) \doteq \varphi_1(f) - \varphi_2(f)$. We claim that such a sequence is given by

$$f_n(x) \doteq f(x) e^{-d(x_0,x)/n}, \qquad n \in \mathbb{N}, \tag{2.10}$$

where x_0 is any fixed point. Indeed since $\mathcal{M}$ is complete, $d(x_0,x)$ diverges at infinity as explained in Section 2.3 below; by the 1-Lipschitz condition $|f(x)| \leq |f(x_0)| + d(x_0,x)$, and this proves that $|f_n(x)| \leq \big(|f(x_0)| + d(x_0,x)\big) e^{-d(x_0,x)/n}$ vanishes at infinity.

To obtain (2.8), one first notices that $\Delta(f+C) = \Delta(f)$ for any $C \in \mathbb{R}$, so that we can assume without loss of generality that $f(x_0) = 0$, that is to say

$$|f(x)| = |f(x) - f(x_0)| \leq d(x_0,x).$$

Second, from $\nabla f_n = \big(\nabla f - n^{-1} f \nabla d(x_0,\circ)\big)\, e^{-d(x_0,\circ)/n}$ and remembering that both f and $d(x_0,\circ)$ are 1-Lipschitz, one gets

$$|\nabla f_n| \leq \big(1 + n^{-1} d(x_0,\circ)\big) e^{-n^{-1} d(x_0,\circ)}.$$

The inequality (2.8) then follows from $(1+\xi)e^{-\xi} \leq 1 \ \forall\ \xi > 0$. (2.9) comes from Lebesgue's dominated convergence theorem: $|f_n(x)| \leq |f(x)| \ \forall\ x,\ n$ and $|f|$ is μ_i integrable by hypothesis, so that $\lim\limits_{n\to\infty} \int_{\mathcal{M}} f_n \mathrm{d}\mu_i = \int_{\mathcal{M}} f \mathrm{d}\mu_i$. ■

2.2 Alternative definitions

There exist several equivalent formulations of the spectral distance. First one may consider continuous instead of smooth functions. Indeed, as explained in [11], for any measurable bounded function f one can view $[D, f]$ as the bilinear form

$$\xi, \eta \to \langle D\xi, f\eta\rangle - \langle f^*\xi, D\eta\rangle,$$

well defined on the domain of D (a dense subset of $L^2(\mathcal{M},\mathcal{S})$). Therefore $[D,f]$ makes sense also when f is not smooth and one can define [12] the spectral distance as the supremum on all continuous functions $f\in C_0(M)$ with $||[D,f]||\leq 1$, that is the set of 1-Lipschitz functions, obtaining thus directly (2.7). In the literature one finds both definitions: supremum on continuous functions [12, 23] or on smooth functions [13, 16].

Second, one may be puzzled by the use of the spin structure to recover the Riemannian metric. In fact instead of the Dirac operator one can equivalently use, as explained in [14], the signature operator $d+d^\dagger$ acting on the Hilbert space $L^2(\mathcal{M},\wedge)$ of square-integrable differential forms, or the de-Rham Laplacian $\Delta = dd^\dagger + d^\dagger d$ acting on $L^2(\mathcal{M})$. Here d is the exterior derivative and $d^\dagger$ its adjoint with respect to the inner product [23]

$$\langle\omega,\omega'\rangle = \int_{\mathcal{M}} (\omega,\omega')\nu_g \qquad \forall\, \omega,\omega'\in L^2(\mathcal{M},\wedge)$$

with ν_g the volume form and the inner product on k-form given by

$$\left(dx^{\alpha_1}\wedge\cdots\wedge dx^{\alpha_k}, dx^{\beta_1}\wedge\cdots\wedge dx^{\beta_{k'}}\right) = \delta_{kk'}\det\left(g^{\alpha_i\beta_j}\right) \qquad 1\leq i,j\leq k.$$

The action of both operators only depends on the Riemannian structure and suitable commutators yield on self-adjoint elements $f=f^*$ in $C_0^\infty(\mathcal{M})$ the same semi-norm as the commutator with D. Explicitly[3]

$$||[d+d^\dagger,f]^2||_{\rm op} = \tfrac{1}{2}||[[\Delta,f],f]||_{\rm op} = ||[D,f]||^2_{\rm op}. \tag{2.11}$$

To show these equalities, let us note that on $L^2(\mathcal{M},\wedge)$, $[d,f]=\epsilon(df)$ where ϵ denotes the wedge multiplication,

$$\epsilon(df)\omega \doteq df\wedge\omega.$$

Therefore $[d,f]$ commutes with 0-form, and the same is true for its adjoint $[d,f]^\dagger$. Assuming $f=f^*$, that is $[d^\dagger,f]=-[d,f]^\dagger$, few manipulations with commutators yield

$$\begin{aligned}
[[\Delta,f],f] &= [[d,f]d^\dagger,f] + [d[d^\dagger,f],f] + [[d^\dagger,f]d,f] + [d^\dagger[d,f],f] && \text{(2.12a)}\\
&= 2[d,f][d^\dagger,f] + 2[d^\dagger,f][d,f] && \text{(2.12b)}\\
&= 2[d+d^\dagger,f]^2, && \text{(2.12c)}
\end{aligned}$$

where $[d,f]^2=[d^\dagger,f]^2=0$ due to the graded commutativity of the wedge product. Remembering that $\epsilon(dx^\mu)^\dagger = g^{\mu\nu}\iota(\partial_\nu)$ where ι is the contraction (see e.g. [23]), one gets $[d,f]^\dagger = \epsilon(df)^\dagger = ((\partial_\mu f)\epsilon(dx^\mu))^\dagger = (g^{\mu\nu}\partial_\nu f)\iota(\partial_\mu)$ so that (2.12b) becomes

$$[[\Delta,f],f] = -2\partial_\rho f(g^{\mu\nu}\partial_\nu f)\left(\epsilon(dx^\rho)\iota(\partial_\mu) + \iota(\partial_\mu)\epsilon(dx^\rho)\right) = -2g^{\nu\rho}\partial_\rho f\partial_\nu f,$$

where we used the fact that $\epsilon(dx^\rho)\iota(\partial_\mu)+\iota(\partial_\mu)\epsilon(dx^\rho) = \delta^\rho_\mu$. With (2.12c) this shows that $[d+d^\dagger,f]^2 = \frac{1}{2}[[\Delta,f],f]$ is the operator of point-wise multiplication by the function $g^{\mu\nu}\partial_\mu f\partial_\nu f$, whose operator norm is $||g^{\mu\nu}\partial_\mu f\partial_\nu f||_\infty$. (2.11) then follows from (2.6).

Notice that instead of the Laplacian one could use any other 2nd order differential operator with the same principal symbol.

[3] Although we use the same symbol, one should keep in mind that the three operator norms in (2.11) correspond to actions on different Hilbert spaces: $L^2(\mathcal{M},\wedge)$, $L^2(\mathcal{M})$, $L^2(\mathcal{M},\mathcal{S})$.

2.3 On the importance of being complete

At point iii in the proof of Proposition 2.1 it is crucial that $\mathcal{M}$ is complete. By the Hopf–Rinow theorem a complete finite-dimensional Riemannian manifold is a proper metric space, hence the geodesic distance from any fixed point x_0, $x \mapsto d(x_0, x)$, is a proper map [42]. In particular for non-compact $\mathcal{M}$ this means that $d(x_0, x)$ diverges at infinity, so that the functions f_n in (2.10) vanish at infinity.

When $\mathcal{M}$ is not complete, not only the f_n do not vanish at infinity which spoils the proof, but also the definition of the Wasserstein distance requires more attention. Indeed in [43] Kantorovich duality is proved assuming $\mathcal{X}$ is complete. It is not clear to the authors whether the duality holds in the non-complete case. However one can still take (2.5) as a definition of W, letting aside whether this is the same quantity as (2.2) or not. Then it is easy to see that on non-complete $\mathcal{M}$ the spectral distance and W are not necessarily equal. Suppose indeed that $\mathcal{N}$ is compact, and $\mathcal{M}$ is obtained from $\mathcal{N}$ by removing a point x_0, so that $\mathcal{M}$ is locally compact and *not* complete. The spectral distances $d_D^{\mathcal{M}}$ and $d_D^{\mathcal{N}}$ computed on $\mathcal{M}$ and $\mathcal{N}$ are equal. Indeed by (2.7), which holds also in the non-complete case, we can compute d_D as supremum over continuous functions instead of smooth ones; in the computation of the spectral distance on $\mathcal{N}$, it is equivalent to take the supremum over 1-Lipschitz $f \in C(\mathcal{N})$ or over 1-Lipschitz $f' = f - f(x_0) \in C(\mathcal{N})$ vanishing at x_0 (since $\varphi_1(f) - \varphi_2(f) = \varphi_1(f') - \varphi_2(f')$ for any two states φ_1, φ_2); therefore

$$\begin{aligned} d_D^{\mathcal{N}}(\varphi_1, \varphi_2) &= \sup_{f \in C(\mathcal{N})} \left\{ \varphi_1(f) - \varphi_2(f);\ ||f||_{\text{Lip}} \leq 1 \right\} \\ &= \sup_{f \in C(\mathcal{N}), f(x_0)=0} \left\{ \varphi_1(f) - \varphi_2(f);\ ||f||_{\text{Lip}} \leq 1 \right\} = d_D^{\mathcal{M}}(\varphi_1, \varphi_2), \end{aligned}$$

where in last equality we used $C_0(\mathcal{M}) = \{f \in C(\mathcal{N}), f(x_0) = 0\}$.

On the contrary, W computed between pure states coincide with the geodesic distance, and the latest may or may not be the same on $\mathcal{N}$ and $\mathcal{M}$. For instance one does not modify the geodesic distance by removing a point from the two sphere. But taking for $\mathcal{N}$ the circle, thought of as the closed interval $[0, 1]$ with 0 and 1 identified, and $\mathcal{M}$ the open interval $(0, 1)$, one gets $W_{\mathcal{N}}(x, y) = \min\{|x - y|, 1 - |x - y|\}$ whereas $W_{\mathcal{M}}(x, y) = |x - y|$. Thus on $\mathcal{M} = (0, 1)$

$$d_D^{\mathcal{M}}(x, y) = d_D^{\mathcal{N}}(x, y) = W_{\mathcal{N}}(x, y) \neq W_{\mathcal{M}}(x, y).$$

2.4 On the hypothesis of finite moment of order 1

Restricting to the states with finite moment of order one (cf. (2.4)) guarantees that Wasserstein distance is finite, since by (2.5)

$$\begin{aligned} W(\varphi_1, \varphi_2) &= \sup_{||f||_{\text{Lip}} \leq 1} \left(\int_{\mathcal{X}} \big(f(x) - f(x_0)\big) \mathrm{d}\mu_1(x) - \int_{\mathcal{X}} \big(f(x) - f(x_0)\big) \mathrm{d}\mu_2(x) \right) \\ &\leq \sup_{||f||_{\text{Lip}} \leq 1} \int_{\mathcal{X}} \big|f(x) - f(x_0)\big| \mathrm{d}\mu_1(x) + \sup_{||f||_{\text{Lip}} \leq 1} \int_{\mathcal{X}} \big|f(x) - f(x_0)\big| \mathrm{d}\mu_2(x) \\ &\leq \int_{\mathcal{X}} d(x, x_0) \mathrm{d}\mu_1(x) + \int_{\mathcal{X}} d(x, x_0) \mathrm{d}\mu_2(x) < \infty. \end{aligned} \tag{2.13}$$

An obvious upper-bound is then obtained by choosing $\pi = \mu_1 \times \mu_2$ in (2.2):

$$d_D(\varphi_1, \varphi_2) = W(\varphi_1, \varphi_2) \leq \mathbb{E}(d; \mu_1 \times \mu_2). \tag{2.14}$$

When at least one of the states is pure, this upper bound is an exact result, even outside $\mathcal{S}_1(C_0^\infty(\mathcal{M}))$.

Proposition 2.2. *For any $x \in \mathcal{M}$ and any state $\varphi \in \mathcal{S}(C_0^\infty(\mathcal{M}))$,*

$$d_D(\varphi, \delta_x) = W(\varphi, \delta_x) = \mathbb{E}\big(d(x, \circ); \mu\big).$$

Proof. By Proposition 2.1,

$$\begin{aligned} d_D(\varphi, \delta_x) = W(\varphi, \delta_x) &= \sup_{||f||_{\mathrm{Lip}} \leq 1, f \in L^1(\mu)} \left(\int_{\mathcal{M}} f \mathrm{d}\mu - f(x) \right) \\ &= \sup_{||f||_{\mathrm{Lip}} \leq 1, f \in L^1(\mu)} \left(\int_{\mathcal{M}} (f(y) - f(x)) \, \mathrm{d}\mu(y) \right) \\ &\leq \sup_{||f||_{\mathrm{Lip}} \leq 1, f \in L^1(\mu)} \int_{\mathcal{M}} |f(x) - f(y)| \, \mathrm{d}\mu(y) \\ &\leq \int_{\mathcal{M}} d(x, y) \, \mathrm{d}\mu(y) = \mathbb{E}\big(d(x, \circ); \mu\big). \end{aligned}$$

This supremum is attained on the 1-Lipschitz functions $f(y) \doteq d(x, y)$ in case $\mu \in \mathcal{S}_1(C_0^\infty(\mathcal{M}))$, or is obtained by the sequence f_n as defined in (2.10) in case μ has not a finite moment of order 1. ■

Notice that this proposition does not rely on the finiteness of W nor d_D, and makes sense since $\mathbb{E}\big(d(x, \circ); \mu\big)$ is either infinite or convergent, the integrand in (2.4) being a positive function. When the distributions are both localized around two points, x, y, transportation maps are simply paths from x to y and the minimal work coincides with the cost $c(x, y)$ to move from x to y, i.e. with the geodesic distance. In other words

$$d_D(\delta_x, \delta_y) = W(\delta_x, \delta_y) = d(x, y) \tag{2.15}$$

and one retrieves (1.5). Proposition 2.2 also yields an alternative definition of $S_1(C_0^\infty(\mathcal{M}))$.

Corollary 2.1. $\varphi \in \mathcal{S}_1(C_0^\infty(\mathcal{M}))$ *if and only if* φ *is at finite spectral-Wasserstein distance from any pure state.*

Proof. By the triangle inequality, the moment of order one of φ is finite for a fixed $x = x_0$ if and only if it is finite for all $x \in \mathcal{M}$. ■

Let us conclude this section with some topological remarks.

Definition 2.1. Given an arbitrary spectral triple $(\mathcal{A}, \mathcal{H}, D)$, we call [33]

$$\mathrm{Con}(\varphi) \doteq \{\varphi' \in S(\mathcal{A}), \; d_D(\varphi, \varphi') < +\infty\}.$$

Notice that connected components in $\mathcal{S}(\mathcal{A})$ for the topology induced by d_D coincide with sets of states at finite distance from each other, thus justifying the name $\mathrm{Con}(\varphi)$. Indeed $\mathrm{Con}(\varphi)$ is path-connected since for any $\varphi_0, \varphi_1 \in \mathrm{Con}(\varphi)$ the map

$$[0, 1] \ni t \mapsto \varphi_t = (1 - t)\varphi_0 + t\varphi_1 \in \mathrm{Con}(\varphi)$$

is continuous (for all $\epsilon > 0$ called $\delta_\epsilon = \epsilon / d(\varphi_0, \varphi_1)$ from (1.9) we get $|t - s| < \delta_\epsilon \Rightarrow d_D(\varphi_t, \varphi_s) < \epsilon$). That $\mathrm{Con}(\varphi)$ is maximal – i.e. there is no connected component containing it properly – can be easily seen: for any $\varphi' \in \mathrm{Con}(\varphi)$ any open ball centered at φ' is contained in $\mathrm{Con}(\varphi)$, so that $\mathrm{Con}(\varphi)$ is open; by the triangle inequality the same is true for the complementary set, proving that $\mathrm{Con}(\varphi)$ is also closed. Therefore, any set containing properly $\mathrm{Con}(\varphi)$ is not connected since it contains a subset that is both open and closed.

For the Wasserstein distance, the set of states with finite moment of order one is a connected component.

Corollary 2.2. *For any* $\varphi \in \mathcal{S}_1(C_0^\infty(\mathcal{M}))$, $\mathrm{Con}(\varphi) = \mathcal{S}_1(C_0^\infty(\mathcal{M}))$.

Proof. Let $\varphi \in \mathcal{S}_1(C_0^\infty(\mathcal{M}))$. By the triangle inequality,

$$d_D(\varphi, \varphi') \leq d_D(\varphi, \delta_x) + d_D(\delta_x, \varphi')$$

is finite by Corollary 2.1 as soon as $\varphi' \in \mathcal{S}_1(C_0^\infty(\mathcal{M}))$. Similarly

$$d_D(\delta_x, \varphi') \leq d_D(\delta_x, \varphi) + d_D(\varphi, \varphi')$$

implies that $d_D(\varphi, \varphi')$ is infinite when $\varphi' \notin \mathcal{S}_1(C_0^\infty(\mathcal{M}))$. ■

Notice that considering only pure states, the set

$$\widetilde{\mathrm{Con}}(\varphi) \doteq \mathrm{Con}(\varphi) \cap \mathcal{P}(\mathcal{A}), \qquad \varphi \in \mathcal{P}(\mathcal{A}) \tag{2.16}$$

is not necessarily (path)-connected in $\mathcal{P}(\mathcal{A})$. In the example (1.7) $\widetilde{\mathrm{Con}}(\varphi)$ is indeed a connected component of $\mathcal{P}(\mathcal{A})$; but in the standard model (see Section 4.2) $\widetilde{\mathrm{Con}}(\varphi) = \mathcal{P}(\mathcal{A})$ and contains two disjoint connected components $(\mathcal{M}, \delta_{\mathbb{C}}), (\mathcal{M}, \delta_{\mathbb{H}})$.

3 Bounds for the distance and (partial) explicit results

Having in mind that noncommutative geometry furnishes a description of the full standard model of the electro-weak and strong interactions minimally coupled to Euclidean general relativity [8], computing the spectral distance could be a way to obtain a “picture” of spacetime at the scale of unification. Regarding pure states, this picture has been worked out in [35] and is recalled in Section 4: one finds that the connected component of d_D is the disjoint union of two copies of a spin manifold $\mathcal{M}$, with distance between the copies coming from the Higgs field. Extending this picture to non-pure states is far from trivial since already on Euclidean space explicit computations of the Wasserstein distance are very few. This does not seem to be the most interesting issue in optimal transport, where one is rather interesting in determining the optimal plan than computing W. On the contrary from our perspective computing d_D is of most interest, while finding the optimal plan (i.e. the element that reaches the supremum in the distance formula) is not an aim in itself. In this section, we collect various results on the Wasserstein-spectral distance in the commutative case $\mathcal{A} = C_0^\infty(\mathcal{M})$: upper and lower bounds for the distance on any spin manifold $\mathcal{M}$, and explicit result for a certain class of states in case $\mathcal{M} = \mathbb{R}^n$. Some of these results might be known from optimal transport theory, but we believe it is still interesting to present them from our perspective.

We postpone to the next section a discussion of the noncommutative case. In all this section, $d_D = W$ and to avoid repetition we simply call it the spectral distance.

3.1 Upper and lower bounds on any spin manifold

On a spin manifold, a lower bound on the spectral distance between states with finite moment of order 1 is given by the distance between their mean points. For $\mathcal{M} = \mathbb{R}^m$ the mean point of $\varphi \in \mathcal{S}_1(C_0^\infty(\mathcal{M}))$ is the barycenter of μ, namely

$$\bar{x} \doteq (\bar{x}^\alpha) \qquad \text{with} \quad \bar{x}^\alpha = \mathbb{E}(x^\alpha; \mu) = \int_{\mathbb{R}^m} x^\alpha \mathrm{d}\mu, \tag{3.1}$$

where

$$x^\alpha : \ \mathbb{R}^m \to \mathbb{R}, \qquad x \mapsto x^\alpha(x),$$

with $\alpha = 1, \dots, m$ denote a set of Cartesian coordinate functions on $\mathbb{R}^m$. For $\mathcal{M}$ that is not the Euclidean space, the mean point can be defined through Nash embedding, that is an isometric embedding of $\mathcal{M}$ onto a subset $\widetilde{\mathcal{M}}$ of the Euclidean space $\mathbb{R}^n$ for some $n \geq m$ [37, 38]. We say that the embedding $N : \mathcal{M} \to \widetilde{\mathcal{M}}$ is convex if $\widetilde{\mathcal{M}}$ is a convex subset of $\mathbb{R}^n$. In that case, the barycenter $\tilde{x}$ of $\tilde{\mu} \doteq N_* \mu$ is in $\widetilde{\mathcal{M}}$, and

$$\bar{x} \doteq N^{-1}(\tilde{x}) \in \mathcal{M}$$

is a well defined generalization of (3.1) to $\mathcal{M}$.

Before showing in Proposition 3.1 that the distance between mean points bounds from below the spectral distance, let us recall two properties of Nash embedding that will be useful in the following.

Lemma 3.1. *Let $\mathcal{M}$ be a Riemannian manifold admitting a convex isometric embedding $N : \mathcal{M} \to \widetilde{\mathcal{M}} \subset \mathbb{R}^n$. Then*

$$d(x, y) = |N(x) - N(y)|, \tag{3.2}$$

where $| \cdot |$ is the Euclidean distance on $\mathbb{R}^n$. Moreover, if f is a 1-Lipschitz function on the metric space $(\mathcal{M}, d)$ then $\tilde{f} \doteq f \circ N^{-1}$ is 1-Lipschitz on $(\widetilde{\mathcal{M}}, | \cdot |)$.

Proof. An isometry $N : \mathcal{M} \to \widetilde{\mathcal{M}}$ preserves the distance function and sends geodesics to geodesics (cf. e.g. [9, page 61]). If $\widetilde{\mathcal{M}}$ is convex, since the metric on $\widetilde{\mathcal{M}}$ is the restriction of the Euclidean metric of $\mathbb{R}^n$, geodesics are straight lines and the distance is the Euclidean one. This proves (3.2).

As a consequence, if f is a 1-Lipschitz function on $\mathcal{M}$, for any $\tilde{x}, \tilde{y} \in \widetilde{\mathcal{M}}$ we have

$$|\tilde{f}(\tilde{x}) - \tilde{f}(\tilde{y})| = |f(x) - f(y)| \leq d(x, y) = |\tilde{x} - \tilde{y}|,$$

where $x = N^{-1}(\tilde{x})$ and $y = N^{-1}(\tilde{y})$. This means that $\tilde{f}$ is 1-Lipschitz on $\widetilde{\mathcal{M}}$. ■

Proposition 3.1. *Let $\mathcal{M}$ be a Riemannian spin manifold that admits a convex isometric embedding $\mathcal{M} \hookrightarrow \mathbb{R}^n$. For any states φ_1, φ_2 in $S(C_0(\mathcal{M}))$ with mean points $\bar{x}_1$, $\bar{x}_2$,*

$$d(\bar{x}_1, \bar{x}_2) \leq d_D(\varphi_1, \varphi_2) \leq \mathbb{E}\big(d; \mu_1 \times \mu_2\big).$$

Proof. (2.14) holds true, so we need to prove only the lower bound. Let us fix a basis of the vector space $\mathbb{R}^n$ such that

$$\tilde{x}_1 - \tilde{x}_2 = \big(|\tilde{x}_1 - \tilde{x}_2|, 0, \dots, 0\big), \tag{3.3}$$

where $\tilde{x}_i = N(\bar{x}_i)$, $i = 1, 2$, has Cartesian coordinates

$$\int_{\widetilde{\mathcal{M}}} x^\beta \, \mathrm{d}\tilde{\mu}_i, \qquad \beta = 1, \dots, n.$$

(3.3) is equivalent to

$$\int_{\widetilde{\mathcal{M}}} x^1 \mathrm{d}\tilde{\mu}_1 - \int_{\widetilde{\mathcal{M}}} x^1 \mathrm{d}\tilde{\mu}_2 = |\tilde{x}_1 - \tilde{x}_2| \quad \text{and} \quad \int_{\widetilde{\mathcal{M}}} x^\beta \mathrm{d}\tilde{\mu}_1 - \int_{\widetilde{\mathcal{M}}} x^\beta \mathrm{d}\tilde{\mu}_2 = 0 \quad \text{for} \quad \beta \in [2, n] \, .$$

Therefore

$$|\tilde{x}_1 - \tilde{x}_2| = \left| \int_{\widetilde{\mathcal{M}}} x^1 \mathrm{d}\tilde{\mu}_1 - \int_{\widetilde{\mathcal{M}}} x^1 \mathrm{d}\tilde{\mu}_2 \right| \leq \sup_{||\tilde{f}||_{\mathrm{Lip} \leq 1}} \left| \int_{\widetilde{\mathcal{M}}} \tilde{f} \mathrm{d}\tilde{\mu}_1 - \int_{\widetilde{\mathcal{M}}} \tilde{f} \mathrm{d}\tilde{\mu}_2 \right|$$

where we noticed that x^1 is Lipschitz in $\widetilde{\mathcal{M}}$ with constant 1 since

$$x^1(\tilde{x}) - x^1(\tilde{y}) \leq \sqrt{(x^1(\tilde{x}) - x^1(\tilde{y}))^2 + \sum\nolimits_{\alpha=2}^{n} (x^\alpha(\tilde{x}) - y^\alpha(\tilde{x}))^2} = |\tilde{x} - \tilde{y}|.$$

Using $\int_{\widetilde{\mathcal{M}}} \tilde{f} \mathrm{d}\tilde{\mu} = \int_{\mathcal{M}} f \mathrm{d}\mu$ and Lemma 3.1,

$$|\tilde{x}_1 - \tilde{x}_2| \leq \sup_{||f||_{\mathrm{Lip}\leq 1}} \left| \int_{\mathcal{M}} f \mathrm{d}\mu_1 - \int_{\mathcal{M}} f \mathrm{d}\mu_2 \right| = d_D(\varphi_1, \varphi_2).$$

Proposition follows from (3.2), i.e. $|\tilde{x}_1 - \tilde{x}_2| = d(\bar{x}_1, \bar{x}_2)$. ■

Note that when $\mathcal{M}$ does not admit a convex embedding, Proposition 3.1 still holds with $|\tilde{x}_1 - \tilde{x}_2|$ instead of $d(\bar{x}_1, \bar{x}_2)$. However this might not be the most interesting lower bound since it involves a distance on $\mathbb{R}^n$ that is not the push-forward of the one on $\mathcal{M}$ (the points $\tilde{x}_i$ may *not* be in $\widetilde{\mathcal{M}}$, and the Euclidean distance, even when restricted to $\widetilde{\mathcal{M}}$, is *not* the push-forward of the geodesic distance on $\mathcal{M}$).

3.2 Spectral distance in the Euclidean space

In this section we study the spectral distance between states of $\mathcal{S}_1(C_0^\infty(\mathcal{M}))$ in the case $\mathcal{M} = \mathbb{R}^m$. To a given density probability $\psi \in L^1(\mathbb{R}^m)$ with finite moment of order 1, e.g. a Gaussian $\psi(x) = \pi^{-\frac{m}{2}} e^{-|x|^2}$, one can associate a state $\Psi_{\sigma,x}$ given by

$$\Psi_{\sigma,x}(f) \doteq \frac{1}{\sigma^m} \int_{\mathbb{R}^m} f(\xi) \psi(\tfrac{\xi - x}{\sigma}) \mathrm{d}^m \xi,$$

for any $x \in \mathbb{R}^m$ and $\sigma \in \mathbb{R}^+$. This becomes the pure state (the point) δ_x in the $\sigma \to 0^+$ limit since $\lim\limits_{\sigma \to 0^+} \psi_{\sigma,x}(f) = f(x)$. In this sense $\Psi_{\sigma,x}$ can be viewed as a "fuzzy" point, that is to say a wave-packet – characterized by a shape ψ and a width σ – describing the uncertainty in the localization around the point x.

The spectral distance between wave packets with the same shape is easily calculated.

Proposition 3.2. *The distance between two states $\Psi_{\sigma,x}$ and $\Psi_{\sigma',y}$ is*

$$d_D(\Psi_{\sigma,x}, \Psi_{\sigma',y}) = \int |x - y + (\sigma - \sigma')\xi| \psi(\xi) \mathrm{d}^m \xi. \tag{3.4}$$

In particular, for $\sigma = \sigma'$ the distance does not depend on the shape ψ:

$$d_D(\Psi_{\sigma,x}, \Psi_{\sigma,y}) = |x - y|.$$

Proof. Since $f(z) - f(w) \leq |z - w|$ for any 1-Lipschitz function f, we have

$$\begin{aligned}
\Psi_{\sigma,x}(f) - \Psi_{\sigma',y}(f) &= \int \big(f(\sigma\xi + x) - f(\sigma'\xi + y)\big)\, \psi(\xi) \mathrm{d}^m \xi \\
&\leq \int |x - y + (\sigma - \sigma')\xi|\, \psi(\xi)\, \mathrm{d}^m \xi.
\end{aligned}$$

When $\sigma = \sigma'$ the upper bound is attained by the function

$$h(z) = z \cdot \frac{x - y}{|x - y|},$$

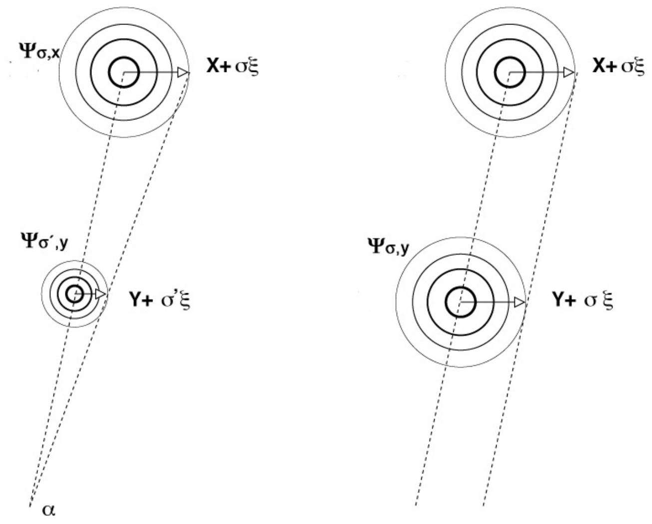

Figure 1. The function h that attains the supremum, in case $\sigma > \sigma'$ and $\sigma = \sigma'$. Dot lines are tangent to the gradient of h.

while for $\sigma \neq \sigma'$ it is attained by the function

$$h(z) = |z - \alpha| \,,$$

where

$$\alpha \doteq \frac{\sigma' x - \sigma y}{\sigma' - \sigma} \tag{3.5}$$

as shown in Fig. 1. ■

In case $\sigma \neq \sigma'$ the function h that attains the supremum measures the geodesic distance between $z \in \mathbb{R}^m$ and the point α defined in (3.5). Geometrically the latest is the intersection of (x, y) with $(x + \sigma\xi, y + \sigma'\xi)$, and is independent of ξ. In case $\sigma = \sigma'$, α is send to infinity and h measures the length of the projection of z on the (x, y) axis (see Fig. 1). The picture is still valid for pure states, i.e. $\sigma = \sigma' = 0$: h can be taken either as the geodesic distance to any point on (x, y) outside the segment $[x, y]$, or as the distance to any axis perpendicular to (x, y) that does not intersect $[x, y]$.

For Gaussian shape, Proposition 3.2 can be confronted with the Wasserstein distance of order 2 between Gaussians computed in [21]. The final expression for the distance of order 2 is quite simpler, since it is an algebraic function of the means and covariance matrices. Here, computing the integral in (3.4) for ψ a Gaussian with the help of a symbolic computation software, one finds complicated expressions involving Bessel functions.

The distance between two arbitrary states on $\mathbb{R}^m$ is less easily computable. Let us first recall what happens in the one dimensional case (cf. e.g. [19]).

Proposition 3.3. *For any two states φ_1 and φ_2, whose corresponding probability measures μ_1 and μ_2 have no singular continuous part,*

$$d_D(\varphi_1, \varphi_2) = \int_{-\infty}^{+\infty} \left| \int_{-\infty}^{z} \big(\mathrm{d}\mu_1(\xi) - \mathrm{d}\mu_2(\xi)\big) \right| \mathrm{d}z.$$

Proof. Integrating by parts one finds

$$\varphi_1(f) - \varphi_2(f) = -\int_{\mathbb{R}} f'(z)\Delta(z)\mathrm{d}z,$$

where

$$\Delta(z) \doteq \int_{-\infty}^{z} \big(\mathrm{d}\mu_1(\xi) - \mathrm{d}\mu_2(\xi)\big) \tag{3.6}$$

is the *cumulative distribution* of the measure $\mu_1 - \mu_2$. We are assuming that μ_i are the sum of a pure point part and a part that is absolutely continuous with respect to the Lebesgue measure. In this case $\Delta(z)$ is piecewise continuous, its sign is a piecewise continuous function, and the primitive $h(z)$ of the sign of $\Delta(z)$ is a Lipschitz continuous function. Now, the 1-Lipschitz condition says that a.e. $|f'(z)| \leq 1$ on $\mathbb{R}$, hence

$$d_D(\varphi_1, \varphi_2) \leq \int_{\mathbb{R}} |\Delta(z)| \mathrm{d}z.$$

This upper bound is attained by any 1-Lipschitz function h such that

$$h'(z) = 1 \quad \text{when} \quad \Delta(z) \geq 0, \qquad h'(z) = -1 \quad \text{otherwise}.$$

By the above consideration, such a 1-Lipschitz function (at least one) always exists. ■

In a quantum context, one can view the cumulative distributions $c_i = \int_{\infty}^{z} \mathrm{d}\mu_i(\xi)$ as the probability to find the particle on the half-line $(-\infty, z]$ before the transport ($i = 1$) or after the transport ($i = 2$). $\Delta(z)$ in (3.6) measures the probability flow across z, and the Wasserstein distance is the integral on $\mathbb{R}$ of the modulus of this probability flow.

On $\mathbb{R}^m$ with $m > 1$ there is no such an explicit result. If ψ_1, ψ_2 are bounded Lipschitz functions with compact support, we know from [19] that there exist two (a.e. unique) bounded measurable Lipschitz functions $a, u : \mathbb{R}^m \to \mathbb{R}$ such that $a \geq 0$,

$$-\nabla(a\nabla u) = \psi_1 - \psi_2$$

in the weak sense, and $|\nabla u| = 1$ almost everywhere on the set where $a > 0$. We have then

$$d_D(\varphi_1, \varphi_2) = \int_{\mathbb{R}^m} a(x) \mathrm{d}x.$$

Indeed integrating by parts, we can write

$$\varphi_1(f) - \varphi_2(f) = \int_{\mathbb{R}^m} (\nabla f) \cdot (a\nabla u) \mathrm{d}x$$

and since $|a\nabla u| = a$ we have

$$d_D(\varphi_1, \varphi_2) \leq \int_{\mathbb{R}^m} a(x) \mathrm{d}x \ .$$

The sup is attained by the function $f = u$.

4 Noncommutative examples

At the light of Proposition 2.1 one may wonder if the analogy between the spectral and the Wasserstein distances still makes sense in a non-commutative framework. In other terms, for $\mathcal{A}$ noncommutative is the distance on $\mathcal{S}(\mathcal{A})$ computed by (1.1) related to some Wasserstein distance?

The most obvious answer, based on Gelfand's identification (1.4) between points and pure states, would be to consider the Wasserstein distance W_D on the metric space $(\mathcal{P}(\mathcal{A}), d_D)$, and question whether W_D on the set $\mathrm{Prob}(\mathcal{P}(\mathcal{A}))$ of probabilities distributions on $\mathcal{P}(\mathcal{A})$ coincides with d_D on $\mathcal{S}(\mathcal{A})$. This is obviously true for pure states since by (2.15)

$$W_D(\omega_1,\omega_2) = c(\omega_1,\omega_2) = d_D(\omega_1,\omega_2) \qquad \forall\, \omega_1,\omega_2 \in \mathcal{P}(\mathcal{A}).$$

For non-pure states however this is usually not true. The reason is that even if

$$\mathcal{S}(\mathcal{A}) \subset \mathrm{Prob}(\mathcal{P}(\mathcal{A}))$$

for commutative and almost commutative C^*-algebras (see the definition in the next paragraph), there is not a 1-to-1 correspondence between the two sets (except in the commutative case). For instance, as recalled in Section 4.1, $S(M_2(\mathbb{C}))$ is a non-trivial quotient of $\mathrm{Prob}(\mathcal{P}(M_2(\mathbb{C})))$, meaning that two probability distributions $\phi_1 \neq \phi_2$ may give the same state $\varphi_1 = \varphi_2$. Thus, given a spectral triple on $M_2(\mathbb{C})$ as the one studied in [7], one has $d_D(\varphi_1,\varphi_2) = 0$ while $W_D(\phi_1,\phi_2) \neq 0$ since the cost being a distance (and assuming $\mathcal{P}(\mathcal{A})$ is a polish space), W is a distance and vanishes if and only if $\phi_1 = \phi_2$.

However this does not mean that the optimal transport interpretation of the spectral distance loses all interest in the noncommutative framework. As explained in Section 4.2, there are interesting analogies between the spectral distance and some Wasserstein distances other than W_D in almost-commutative geometries. The latest are spectral triples $(\mathcal{A},\mathcal{H},D)$ obtained as the product of a spin manifold by a finite-dimensional spectral triple $(\mathcal{A}_I,\mathcal{H}_I,D_I)$, namely

$$\mathcal{A} = C_0^\infty(\mathcal{M}) \otimes \mathcal{A}_I, \qquad \mathcal{H} = L_2(\mathcal{M},\mathcal{S}) \otimes \mathcal{H}_I, \qquad D = -i\gamma^\mu\partial_\mu \otimes \mathbb{I}_I + \gamma^{m+1} \otimes D_I, \tag{4.1}$$

where $\mathbb{I}_I$ is the identity of $\mathcal{H}_I$ and γ^{m+1} is the product of the gamma matrices in case m is even, or the identity in case m is odd. $\mathcal{A}_I$ being finite-dimensional (or, equivalently, $C_0^\infty(\mathcal{M})$ being Abelian) one has [27]:

$$\mathcal{P}(\mathcal{A}) = \mathcal{P}(C_0^\infty(\mathcal{M})) \times \mathcal{P}(\mathcal{A}_I). \tag{4.2}$$

For simplicity we discuss here the case $\mathcal{A}_I = \mathbb{C}^2$ acting on $\mathcal{H}_I = \mathbb{C}^2$, while in Section 4.2 we will focus on the finite-dimensional spectral triple describing the internal degrees of freedom (hence the subscript I) of the standard model of particle physics. Since $\mathbb{C}^2$ has two pure states – $\delta_0(z_0,z_1) = z_0$, $\delta_1(z_0,z_1) = z_1$ $\forall\,(z_0,z_1) \in \mathbb{C}^2$ – from (4.2)

$$\mathcal{P}(\mathcal{A}) = \mathcal{M} \times \{0,1\}$$

This is Connes' idea of "product of the continuum by the discrete" [11] seen at the level of pure states: through the product by a finite-dimensional spectral triple, the points of the manifold acquire a $\mathbb{Z}_2$ internal discrete structure, namely to a point $x = \delta_x$ in the commutative case corresponds two pure states in $\mathcal{P}(\mathcal{A})$,

$$x_0 = (\delta_x,\delta_0), \qquad x_1 = (\delta_x,\delta_1).$$

Moreover, although $\mathcal{P}(\mathcal{A})$ is the disjoint union of two copies of $\mathcal{M}$, points on distinct copies need not to be at infinite spectral distance from one another. Indeed one shows that [35]

$$d_D(x_0,x_1) = d_{D_I}(\delta_0,\delta_1), \tag{4.3}$$

where d_{D_I} denotes the spectral distance associated to $(\mathcal{A}_I,\mathcal{H}_I,D_I)$. $\mathcal{A}_I$ is represented on $\mathcal{H}_I$ by diagonal matrices but $D_I \in M_2(\mathbb{C})$ needs not to be diagonal. Especially, if D_I has non-zero off-diagonal terms then $||[D,a]|| \leq 1$ is a non-trivial constraint which guarantees that (4.3) is finite. Note that if one takes the direct sum of spectral triples instead of a product, namely $C_0^\infty(\mathcal{M}) \oplus C_0^\infty(\mathcal{M})$ acting on $L_2(\mathcal{M},\mathcal{S}) \oplus L_2(\mathcal{M},\mathcal{S})$, one gets the same pure states as above but

$D = -i\gamma^\mu\partial\mu \oplus -i\gamma^\mu\partial\mu$ does not have off-diagonal terms and the two copies of $\mathcal{M}$ are at infinite distance from one another (points have not enough space to "talk to each other" through the off-diagonal terms of D).

From an optimal transport point of view, the non-vanishing of (4.3) can be interpreted as the fact that "staying at a point", which is costless in the commutative case since $c(x,x) = d_D(x,x) = 0$, may have a cost

$$d_D(x_0, x_1) \neq 0 \tag{4.4}$$

in almost-commutative geometries, corresponding to the "internal jump" from x_0 to x_1. We investigate this idea in Section 4.2, showing in Proposition 4.2 that the spectral distance between pure states is the minimal work W_I associated to a cost c_I.

4.1 The Moyal plane

The Moyal algebra $\mathcal{A}_\theta$ is the noncommutative deformation of the non-unital Schwartz algebra $S(\mathbb{R}^2)$ with point-wise product

$$(f \star g)(x) = \frac{1}{(\pi\theta)^2}\int d^2y d^2z\, f(x+y)g(x+z)e^{-i2y\Theta^{-1}z} \qquad \forall\, f, g \in S(\mathbb{R}^2),$$

where $y\Theta^{-1}z \equiv y^\mu\Theta^{-1}_{\mu\nu}z^\nu$ and

$$\Theta_{\mu\nu} = \theta\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$$

with $\theta \in \mathbb{R}$, $\theta \neq 0$. In [7] we studied the spectral distance associated to the spectral triple built in [20] around the action of $\mathcal{A}_\theta$ on $L_2(\mathbb{R}^2)$ and the usual Dirac operator on $\mathbb{R}^2$. We found that the topology induced by the spectral distance on $\mathcal{S}(\mathcal{A}_\theta)$ is not the weak*-topology (a condition required by Rieffel [39, 40, 41] in the unital case in order to define compact quantum metric spaces, and adapted by [30] to the non-unital case). However by viewing $\mathcal{A}_\theta$ as an algebra of infinite-dimensional matrices, we proposed some finite-dimensional truncations of the Moyal spectral triple, based on the algebra $M_n(\mathbb{C})$, $n \in \mathbb{N}$, that makes it a quantum metric space. Explicitly, for $n = 2$ $\mathcal{P}(M_2(\mathbb{C}))$ is homeomorphic to the Euclidean 2-sphere,

$$\xi \in \mathcal{P}(M_2(\mathbb{C})) \longrightarrow \mathbf{x}_\xi = (x_\xi, y_\xi, z_\xi) \in S^2,$$

so that a non-pure state φ is determined by a probability distribution ϕ on S^2. Its evaluation

$$\varphi(a) = \int_{S^2}\phi(\mathbf{x}_\xi)\,\mathrm{Tr}(s_\xi a)\,d\mathbf{x}_\xi = \mathrm{Tr}\left(\left(\int_{S^2}\phi(\mathbf{x}_\xi)s_\xi\,d\mathbf{x}_\xi\right)a\right) \qquad \forall\, a \in M_2(\mathbb{C}),$$

with $s_\xi \in M_2(\mathbb{C})$ the support of the pure state $\mu^{-1}(\mathbf{x}_\xi)$ and $d\mathbf{x}_\xi$ the $SU(2)$ invariant measure on S^2, only depends on the barycenter of ϕ,

$$\bar{\mathbf{x}}_\phi = (\bar{x}_\phi, \bar{y}_\phi, \bar{y}_\phi) \qquad \text{with} \quad \bar{x}_\phi := \int_{S^2}\phi(\mathbf{x}_\xi)x_\xi d\mathbf{x}_\xi$$

and similar notation for $\bar{y}_\phi$, $\bar{z}_\phi$. With the equivalence relation $\phi \sim \phi' \iff \bar{\mathbf{x}}_\phi = \bar{\mathbf{x}}_{\phi'}$ one gets that

$$\mathcal{S}(\mathcal{M}_2(\mathbb{C})) = \mathcal{S}(C(S^2))/\sim$$

is homeomorphic to the Euclidean 2-ball:

$$\varphi \xrightarrow{\mu} \bar{\mathbf{x}}_\phi \in \mathcal{B}^2.$$

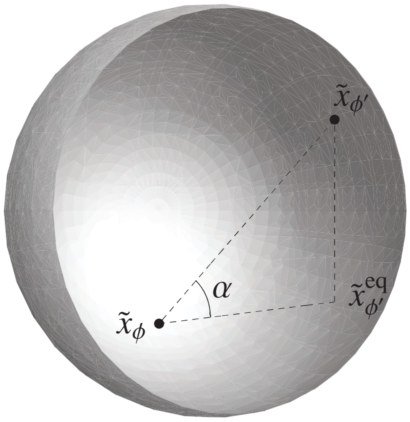

Figure 2. The vertical plane containing $\bar{\mathbf{x}}_\phi$, $\bar{\mathbf{x}}_{\phi'}$.

In [7] we computed

$$d_D(\bar{\mathbf{x}}_\phi, \bar{\mathbf{x}}_{\phi'}) = \sqrt{\frac{\theta}{2}} \times \begin{cases} \cos\alpha \, d_{Ec}(\tilde{\mathbf{x}}_\phi, \tilde{\mathbf{x}}_{\phi'}) & \text{when } \alpha \leq \dfrac{\pi}{4}, \\ \dfrac{1}{2\sin\alpha} \, d_{Ec}(\tilde{\mathbf{x}}_\phi, \tilde{\mathbf{x}}_{\phi'}) & \text{when } \alpha \geq \dfrac{\pi}{4}, \end{cases}$$

where $d_{Ec}(\bar{\mathbf{x}}_\phi, \bar{\mathbf{x}}_{\phi'}) = |\bar{\mathbf{x}}_\phi - \bar{\mathbf{x}}_{\phi'}|$ is the Euclidean distance and α is the angle between the segment $[\bar{\mathbf{x}}_\phi, \bar{\mathbf{x}}_{\phi'}]$ and the horizontal plane $z_\xi = \text{const}$ (see Fig. 2).

The state space of the truncated Moyal algebra is not the space of distributions on S^2, but a quotient of it. Therefore for $\phi \sim \phi', \phi \neq \phi'$, one has $d_D(\varphi, \varphi') = 0$ while $W_D(\varphi, \varphi') \neq 0$. In this example there seems to be no Wasserstein distance naturally associated to the spectral distance. In the next section, we exhibit another noncommutative example for which there exists a Wasserstein distance W', with cost d' defined on a set bigger than $\mathcal{P}(\mathcal{A})$, that coincides with d_D for some non-pure states.

4.2 The standard model

The spectral triple describing the standard model of elementary particles is the product (4.1) of a manifold with the finite-dimensional algebra $\mathbb{C}\oplus\mathbb{H}\oplus M_3(\mathbb{C})$ acting on $\mathcal{H}_I = \mathbb{C}^{96}$. D_I is a 96×96 matrix with entries the masses of the elementary fermions, the Cabibbo–Kobayashi–Maskawa matrix and the neutrino mixing-angles. The choice of the algebra is dictated by physics [8] (its unitaries group gives back the gauge group of the standard model) and 96 is the number of elementary fermions[4]. The pure states of $M_3(\mathbb{C})$ turn out to be at infinite distance from one another and from any other pure state [35] so that, from the metric point of view, the interesting part is the product $(\mathcal{A}, \mathcal{H}, D)$ of a manifold by $(\mathcal{A}_I \doteq \mathbb{C}\oplus\mathbb{H}, \mathcal{H}_I, D_I)$. By (4.2) the pure-states space is

$$\mathcal{P}(\mathcal{A}) = \mathcal{M} \times \{0, 1\}$$

since both $\mathbb{C}$ and the algebra of quaternion $\mathbb{H}$ have one single pure state[5]

$$\delta_{\mathbb{C}}(z) = \Re(z) \quad \forall\, z \in \mathbb{C}, \qquad \delta_{\mathbb{H}}(h) = \frac{1}{2}\mathrm{Tr}\, h \quad \forall\, h = \begin{pmatrix} \alpha & \beta \\ -\bar{\beta} & \bar{\alpha} \end{pmatrix} \in \mathbb{H}, \alpha, \beta \in \mathbb{C}.$$

[4](6 leptons + 6 quarks × 3 colors) × 2 chiralities = 48, to which are added 48 antiparticles.

[5]$\mathbb{H}$ is a real, but not a complex algebra. Real C^*-algebras are defined similarly to complex ones [22], except that one imposes that $1 + a^*a$ is invertible – which comes as a consequence in the complex case. A state φ is then defined as a real, real-linear, positive form such that $\varphi(\mathbb{I}) = 1$ and $\varphi(a^*) = \varphi(a)$. Hence $\mathbb{H}$ has only one state, given by half the trace. Viewing $\mathbb{C}$ and $C_0^\infty(\mathcal{M})$ as real C^*-algebras, their states are obtained by taking the real part of their "usual" complex states. See [31, page 42].

The space underlying the standard model is thus the disjoint union of two copies of the manifold, and with some computation one shows [35] that the spectral distance is finite on $\mathcal{P}(\mathcal{A})$ and coincides with the geodesic distance d' on the manifold $\mathcal{M}' = \mathcal{M} \times \mathcal{I}$, with $\mathcal{I} := [0,1]$ the closed unit interval, with metric

$$\begin{pmatrix} g^{\mu\nu}(x) & 0 \\ 0 & ||D_I||^2 \end{pmatrix}, \tag{4.5}$$

where g is the Riemannian metric on $\mathcal{M}$. Assuming that $\mathcal{M}$ is complete, and writing $t \in \mathcal{I}$ the extra-coordinate, this can be restated as: the spectral distance between pure states in the standard model coincides with the Wasserstein distance W' on the metric space $(\mathcal{M}', d')$ restricted to the two hyperplanes $t = 0$, $t = 1$. This reformulation of the main result of [35] allows to determine the spectral distance between a certain class of non-pure states, namely those localized on one of the copies of $\mathcal{M}$.

Explicitly, $\mathcal{S}(\mathcal{A})$ is the set of couples of measures $\varphi = (\mu, \nu)$ on $\mathcal{M}$, normalized to

$$\int_{\mathcal{M}} d\mu + \int_{\mathcal{M}} d\nu = 1,$$

whose evaluation on

$$\mathcal{A} \ni a = f \oplus \begin{pmatrix} g & b \\ -\bar{b} & \bar{g} \end{pmatrix},$$

where f, g, b are in $C_0^\infty(\mathcal{M})$, is given by

$$\varphi(a) = \int_{\mathcal{M}} \Re(f)\, d\mu + \int_{\mathcal{M}} \Re(g)\, d\nu.$$

Two states φ_1, φ_2 are localized on the same copy of $\mathcal{M}$ if $\varphi_1 = (0, \nu_1)$, $\varphi_2 = (0, \nu_2)$; or $\varphi_1 = (\mu_1, 0)$, $\varphi_2 = (\mu_2, 0)$. Such states can be viewed as elements of $\mathcal{S}(\mathcal{A})$, $\mathcal{S}(C_0^\infty(\mathcal{M}'))$ and $\mathcal{S}(C_0^\infty(\mathcal{M}))$.

Proposition 4.1. *For two states φ_1, φ_2 localized on the same copy of $\mathcal{M}$,*

$$d_D(\varphi_1, \varphi_2) = W(\varphi_1, \varphi_2) = W'(\varphi_1, \varphi_2). \tag{4.6}$$

Proof. To fix notation we assume that $\varphi_1 = (0, \nu_1)$, $\varphi_2 = (0, \nu_2)$ are localized on the $\delta_{\mathbb{H}}$ copy of $\mathcal{M}$, that we write $\mathcal{M}_{\mathbb{H}}$ and associate to the value $t = 1$ of the extra-parameter $t \in \mathcal{I}$. The evaluation of $\varphi_1 - \varphi_2$ on $a \in \mathcal{A}$ only depends on $\mathrm{Re}(g)$. Moreover for $a = a*$, $||[D, a]|| \leq 1$ implies that g is 1-Lipschitz [35, equation (16)], so that

$$d_D(\varphi_1, \varphi_2) \leq W(\varphi_1, \varphi_2). \tag{4.7}$$

The equality is attained by considering $a = g_w \otimes \mathbb{I}_3$ where g_w is the real 1-Lipschitz function that attains the supremum in the computation of W. Hence the first equation in (4.6).

To show that $W(\varphi_1, \varphi_2) = W'(\varphi_1, \varphi_2)$, one first notices that (4.5) being block-diagonal implies that $d'((x, 1), (y, 1)) = d(x, y)$. Hence the restriction $g(y) \doteq g'(y, 1)$ on $\mathcal{M}_{\mathbb{H}}$ of any 1-Lipschitz function g' in $C_0^\infty(\mathcal{M}')$ is 1-Lipschitz,

$$|g'(x, t_0) - g'(y, t_1)| \leq d'((x, t_0), (y, t_1)) \Rightarrow |g(x) - g(y)| \leq d(x, y).$$

Conversely to any 1-Lipschitz function $g \in C_0^\infty(\mathcal{M})$ one associates

$$g'(x, t) \doteq g(x) \qquad \forall\, x \in \mathcal{M}, t \in \mathcal{I}$$

which is 1-Lipschitz in $C_0^\infty(\mathcal{M}')$ and takes the same value as g on $\mathcal{M}_{\mathbb{H}}$. Therefore

$$\sup_{g' \in C_0^\infty(\mathcal{M}'), ||g'||_{\mathrm{Lip}=1}} |\varphi_1(g') - \varphi_2(g')| = \sup_{g \in C_0^\infty(\mathcal{M}), ||g||_{\mathrm{Lip}=1}} |\varphi_1(g) - \varphi_2(g)|$$

which, together with (4.7), yields the result. ∎

It is tempting to postulate that $d_D(\varphi_1, \varphi_2) = W'(\varphi_1, \varphi_2)$ for states localized on different copies, or states that are not localized on any copy (i.e. $\varphi = (\mu, \nu)$ with μ, ν both non-zero). This point is under investigation.

A disturbing point in the computation of the spectral distance in the standard model is the appearance of a compact extra-dimension $\mathcal{I}$ while the internal structure is discrete. In [35] $\mathcal{I}$ came out more as a computational artifact than a requirement of the model. From the Wasserstein distance point of view, the introduction of the extra-dimension can be seen in the following proper inclusions:

$$\mathcal{P}(\mathcal{A}) \subset \mathcal{M}' \subset \mathcal{S}(\mathcal{A}).$$

Namely W' is associated to the metric space $(\mathcal{M}', d')$ which is bigger than $\mathcal{P}(\mathcal{A})$ (the points between the sheets are not pure states of $\mathcal{A}$) and smaller than $\mathcal{S}(\mathcal{A})$ (non-pure states localized on a copy are not in $\mathcal{M}'$). To get rid of the extra-dimension, one could consider the metric space $(\mathcal{P}(\mathcal{A}), d_D)$ with associated Wasserstein distance W_D, but this would be of poor interest since the definition of $d_D = d'$ requires the knowledge of $\mathcal{M}'$. Alternatively one could look for a cost defined solely on $\mathcal{M}$. For states (pure or not) localized on the same copy, this cost is simply the geodesic distance on $\mathcal{M}$, as shown in Proposition 4.1. For pure states on distinct copies such a cost c_I also exists and is given by

$$c_I(x, y) \doteq \sqrt{d(x, y)^2 + \frac{1}{||D_I||^2}}. \tag{4.8}$$

Note that c_I is not a distance since it does not vanish on the diagonal,

$$c_I(x, x) = \frac{1}{||D_I||} \qquad \forall\, x \in \mathcal{M},$$

but gives back the jump-cost (4.4) (by (4.5), $d(x_0, x_1) = \frac{1}{||D_I||}$). However c_I satisfies the condition required to proved Kantorovich's duality (namely [43] it is lower semicontinuous and satisfies $c_I(x, y) \geq a(x) + b(y)$ for some real-valued upper semicontinous μ_i-integrable functions). Hence it makes sense to consider the minimal work W_I associated to c_I, whose formula is given by (2.5) with the Lipschitz condition replaced by $|f(x) - f(y)| \leq c_I(x, y)$.

Proposition 4.2. *The spectral distance between pure states $x_0 \doteq (\delta_x, \delta_{\mathbb{C}})$, $y_1 \doteq (\delta_y, \delta_{\mathbb{H}})$ on distinct copies is*

$$d_D(x_0, y_1) = W_I(x, y).$$

Proof. For pure states the minimal work is the cost itself: $W_I(x, y) = c_I(x, y)$ for any $x, y \in \mathcal{M}$. Since $d_D(x_0, y_1) = d'(x_0, y_1)$ as recalled above (4.5), the result follows if one proves that

$$d'^2(x_0, y_1) = d^2(x, y) + \frac{1}{||D_I||^2}. \tag{4.9}$$

This has been shown in [35] but we briefly restate the argument here for sake of completeness. Let us write $x^a = (x^\mu \in \mathcal{M}, t \in [0, 1])$ a point of $\mathcal{M}'$ and $g^{tt} = g_{tt}^{-1} = ||D_I||^2$ the extra-metric component. The Christoffel symbols involving t are

$$\Gamma^t_{t\mu} = \Gamma^t_{\mu t} = \frac{1}{2} g^t \partial_\mu g_t, \qquad \Gamma^\mu_t = -\frac{1}{2} g^{\mu\nu} \partial_\nu g_{tt}, \qquad \Gamma^\mu_{0\nu} = \Gamma^\mu_{\nu t} = \Gamma^t_{tt} = \Gamma^t_{\mu\nu} = 0.$$

The geodesic equation $\ddot{x}^a + \Gamma^a_{bc} \dot{x}^b \dot{x}^c$ writes

$$\ddot{x}^t + g^{tt}(\partial_\mu g_{tt}) \dot{x}^t \dot{x}^\mu = 0, \tag{4.10a}$$

$$\ddot{x}^\mu - \frac{1}{2}g^{\mu\nu}(\partial_\nu g_{tt})(\dot{t})^2 + \Gamma^\mu_{\lambda\rho}\dot{x}^\lambda\dot{x}^\rho = 0 \tag{4.10b}$$

and, because g_{tt} is a constant, they simplify to

$$\dot{t} = \text{const} = K, \tag{4.11a}$$

$$\ddot{x}^\mu + \Gamma^\mu_{\lambda\rho}\dot{x}^\lambda\dot{x}^\rho = 0. \tag{4.11b}$$

The first geodesic equation indicates that the proper length τ of a geodesic $\mathcal{G}' = x(\tau)$ in $\mathcal{M}'$ between x_0 and y_1 is proportional to t,

$$dt = K d\tau,$$

as well as to the line element $ds = \sqrt{g_{\nu\nu}dx^\mu dx^\nu}$ of $\mathcal{M}$ since

$$1 = ||\dot{x}|| = g_{ab}\dot{x}^a\dot{x}^b = g_{\mu\nu}\frac{dx^\mu}{d\tau}\frac{dx^\nu}{d\tau} + g_{tt}K^2 = \frac{ds^2}{d\tau^2} + g_{tt}K^2, \tag{4.12}$$

so that, assuming $g_{tt}K^2 \neq 1$ (which from (4.12) amounts to take $x \neq y$),

$$d\tau = \frac{ds}{\sqrt{1 - g_{tt}K^2}}\,.$$

The second geodesic equation (4.11b) shows that the projection on $\mathcal{M}$ of $\mathcal{G}'$ is a geodesic $\mathcal{G}$ of $\mathcal{M}$. Therefore

$$d'(x_0, y_1) = \int_{\mathcal{G}'} d\tau = \int_{\mathcal{G}} \frac{d\tau}{ds} ds = \frac{1}{\sqrt{1 - g_{tt}K^2}} \int_{\mathcal{G}} ds = \frac{1}{\sqrt{1 - g_{tt}K^2}} d(x, y),$$

that is to say

$$d'^2(x_0, y_1) = d^2(x, y) + g_{tt}K^2 d'^2(x_0, y_1).$$

Writing $K^2 d'^2(x_0, y_1)$ as

$$\left(\int_{\mathcal{G}'} K d\tau\right)^2 = \left(\int_{\mathcal{G}'} dt\right)^2 = (t((y, 1) - t((x, 0)))^2 = 1$$

one obtains (4.9), and the result. ■

Let us underline an interesting feature of this proposition, namely that a cost which is not a distance $\mathcal{M}$ can be seen as a distance on $\mathcal{M} \times \mathcal{M}$.

Proposition 4.2 can also be rewritten using a cost vanishing on the diagonal, namely

$$c'_I(x, y) \doteq c_I(x, y) - \frac{1}{||D_I||^2} = \sqrt{d(x, y)^2 + \frac{1}{||D_I||^2}} - \frac{1}{||D_I||}.$$

Writing W'_I the associated minimal work, one has

$$d_D(x_0, y_1) = W_I(x, y) + \frac{1}{||D_I||^2}.$$

Quite remarkably, the cost (4.8) is similar to the cost (32) introduced in [6] in the framework of the relativistic heat equation.

Proposition 4.2 relies on the fact that the jump-cost (4.4) is constant. From a physics point of view, this means that one does not take into account the Higgs field. In almost commutative geometries, the latest is obtained by *inner fluctuation of the metric* [13] that substitute D with a covariant Dirac operator. From the metric point of view this amounts to replacing (see [35] for details) (4.5) by

$$\begin{pmatrix} g^{\mu\nu}(x) & 0 \\ 0 & ||D_I + H(x)||^2 \end{pmatrix},$$

where

$$H(x) = \left(|1 + h_1(x)|^2 + |h_2(x)|^2\right) m_t^2$$

where (h_1, h_2) is the (complex) Higgs doublet and m_t is the mass of the quark top. Instead of (4.3) one has $d_D(x_0, x_1) = d_{D_I+H(x)}(\delta_{\mathbb{C}}, \delta_{\mathbb{H}})$ (the jump-cost is no longer constant). In analogy with (4.8), one could define the cost

$$\tilde{c}_I(x, y) = \sqrt{d(x, y)^2 + d_{D_I+H(x)}(\delta_{\mathbb{C}}, \delta_{\mathbb{H}})^2}$$

which allows to avoid the introduction of the extra-dimension. However the projection of a geodesic in $\mathcal{M}'$ is no longer a geodesic of $\mathcal{M}$ ((4.10b) does not simplify to (4.11b)) so there is no way to express $d'(x_0, y_1)$ as a function of $d(x, y)$, and $d_D(x_0, y_1) = d'(x_0, y_1)$ no longer equals $W_I(x_0, y_1)$.

5 Conclusion

The spectral distance between states on a complete Riemannian spin manifold coincides with the Wasserstein distance of order 1. In the noncommutative case the analogies between the spectral distance and various Wasserstein distances, although still not fully understood, shed an interesting light on the interpretation of the distance formula in noncommutative geometry. In physics, defining the distance as a supremum rather than an infimum is useful since, at small scale, quantum mechanics indicates that notions as "paths between points" no longer make sense, so that the classical definition of distance as the length of the shortest path loses any operational meaning. An interesting feature of noncommutative geometry is to provide a notion of distance overcoming this difficulty. In transport theory the interpretation of Kantorovich formula has an interpretation in economics rather than in quantum mechanics: while Monge formulation corresponds to the minimization of a cost, Kantorovich dual formula corresponds to the maximization of a profit. To repeat a classical example found in the literature: assume that the distribution of flour-producers on a given territory $\mathcal{M}$ is given by μ_1 and the distribution of bakeries by μ_2. Consider a transport-consortium whose job consists in buying the flour at factories and selling it to bakers. The consortium fixes the value $f(x)$ of the flour at the point x (it buys the flour at the price $f(x)$ if there is a factory, or it sells it at a price $f(x)$ if there is a bakery). The Wasserstein distance is the maximum profit the consortium may hope, under the constraint of staying competitive, that means not selling the flour to a price higher than the bakeries would pay if they were doing the transport by themselves (i.e. $f(x) \leq f(y) + c(x, y)$ for all x, y). This raises an interesting question: in a quantum context, what is the physical meaning of this "profit" that one is maximizing while computing the distance? If one view the states on $\mathcal{M}$ as wave functions, is the distance related to the minimum work required to transform one wave into the second? More specifically, for $\mathcal{M} = \mathbb{R}^m$ what physical quantity represents (3.4), and what is the meaning of the point α?

Acknowledgments

We would like to thank Hanfeng Li and anonymous referees for their valuable comments.

References

[1] Ambrosio L., Lecture notes on optimal transport problems, in Mathematical Aspects of Evolving Interfaces (Funchal, 2000), *Lecture Notes in Math.*, Vol. 1812, Springer, Berlin, 2003, 1–52.

[2] Azagra D., Ferrera J., López-Mesas F., Rangel Y., Smooth approximation of Lipschitz functions on Riemannian manifolds, *J. Math. Anal. Appl.* **326** (2007), 1370–1378, math.DG/0602051.

[3] Bratteli O., Robinson D.W., Operator algebras and quantum statistical mechanics. 1. C^*- and W^*-algebras, symmetry groups, decomposition of states, 2nd ed., *Texts and Monographs in Physics*, Springer-Verlag, New York, 1987.

[4] Biane P., Voiculescu D., A free probability analogue of the Wasserstein metric on the trace-state space, *Geom. Funct. Anal.* **11** (2001), 1125–1138, math.OA/0006044.

[5] Bimonte G., Lizzi F., Sparano G., Distances on a lattice from non-commutative geometry, *Phys. Lett. B* **341** (1994), 139–146, hep-lat/9404007.

[6] Brenier Y., Extended Monge–Kantorovich theory, in Optimal Transportation and Applications (Martina Franca, 2001), *Lecture Notes in Math.*, Vol. 1813, Springer, Berlin, 2003, 91–121.

[7] Cagnache E., D'Andrea F., Martinetti P., Wallet J.C., The Spectral distance in the Moyal plane, arXiv:0912.0906.

[8] Chamseddine A.H., Connes A., Marcolli M., Gravity and the standard model with neutrino mixing, *Adv. Theor. Math. Phys.* **11** (2007), 991–1089, hep-th/0610241.

[9] Chavel I., Riemannian geometry – a modern introduction, *Cambridge Tracts in Mathematics*, Vol. 108, Cambridge University Press, Cambridge, 1993.

[10] Connes A., Compact metric spaces, Fredholm modules, and hyperfiniteness, *Ergodic Theory Dynam. Systems* **9** (1989), 207–220.

[11] Connes A., Noncommutative geometry, Academic Press, Inc., San Diego, CA, 1994.

[12] Connes A., Noncommutative geometry and reality, *J. Math. Phys.* **36** (1995), 6194–6231.

[13] Connes A., Gravity coupled with matter and the foundation of non-commutative geometry, *Comm. Math. Phys.* **182** (1996), 155–176.

[14] Connes A., A unitary invariant in Riemannian geometry, *Int. J. Geom. Methods Mod. Phys.* **5** (2008), 1215–1242, arXiv:0810.2091.

[15] Connes A., Lott J., The metric aspect of noncommutative geometry, in New symmetry principles in quantum field theory (Cargèse, 1991), *NATO Adv. Sci. Inst. Ser. B Phys.*, Vol. 295, Plenum, New York, 1992, 53–93.

[16] Connes A., Marcolli M., Noncommutative geometry, quantum fields and motives, *American Mathematical Society Colloquium Publications*, Vol. 55, American Mathematical Society, Providence, RI; Hindustan Book Agency, New Delhi, 2008.

[17] Doplicher S., Fredenhagen K., Roberts J.E., Spacetime quantization induced by classical gravity, *Phys. Lett. B* **331** (1994), 39–44.

[18] Doplicher S., Fredenhagen K., Roberts J.E., The quantum structure of spacetime at the Planck scale and quantum fields, *Comm. Math. Phys.* **172** (1995), 187–220, hep-th/0303037.

[19] Evans L.C., Gangbo W., Differential equations methods for the Monge–Kantorevich mass transfer problem, *Mem. Amer. Math. Soc.* **137** (1999), no. 653.

[20] Gayral V., Gracia-Bondía J.M., Iochum B., Schücker T., Varilly J.C., Moyal planes are spectral triples, *Comm. Math. Phys.* **246** (2004), 569–623, hep-th/0307241.

[21] Givens C.R., Shortt R.M., A class of Wasserstein metrics for probability distributions, *Michigan Math. J.* **31** (1984), 231–240.

[22] Goodearl K.R., Notes on real and complex C^*-algebras, *Shiva Mathematics Series*, Vol. 5, Shiva Publishing Ltd., Nantwich, 1982.

[23] Gracia-Bondía J.M., Varilly J.C., Figueroa H., Elements of noncommutative geometry, *Birkhäuser Advanced Texts: Basler Lehrbücher*, Birkhäuser Boston, Inc., Boston, MA, 2001.

[24] Greene R.E., Wu H., C^∞ approximations of convex, subharmonic, and plurisubharmonic functions, *Ann. Sci. École Norm. Sup. (4)* **12** (1979), 47–84.

[25] Gromov M., Metric structures for Riemannian and non-Riemannian spaces, *Progress in Mathematics*, Vol. 152, Birkhäuser Boston, Inc., Boston, MA, 1999.

[26] Iochum B., Krajewski T., Martinetti P., Distances in finite spaces from noncommutative geometry, *J. Geom. Phys.* **37** (2001), 100–125, hep-th/9912217.

[27] Kadison R.V., Ringrose J.R., Fundamentals of the theory of operator algebras. Vol. II. Advanced theory, *Pure and Applied Mathematics*, Vol. 100, Academic Press, Inc., Orlando, FL, 1986.

[28] Kantorovich L.V., On the transfer of masses, *Dokl. Akad. Nauk. SSSR* **37** (1942), 227–229.

[29] Kantorovich L.V., Rubinstein G.S., On a space of totally additive functions, *Vestnik Leningrad. Univ.* **13** (1958), no. 7, 52–59.

[30] Latrémolière F., Bounded-Lipschitz distances on the state space of a C^*-algebra, *Taiwanese J. Math.* **11** (2007), 447–469.

[31] Martinetti P., Distances en géométrie non-commutative, PhD Thesis, math-ph/0112038.

[32] Martinetti P., Carnot–Carathéodory metric and gauge fluctuations in noncommutative geometry, *Comm. Math. Phys.* **265** (2006), 585–616, hep-th/0506147.

[33] Martinetti P., Spectral distance on the circle, *J. Funct. Anal.* **255** (2008), 1575–1612, math.OA/0703586.

[34] Martinetti P., Smoother than a circle or how noncommutative geometry provides the torus with an egocentred metric, in Modern Trends in Geometry and Topology (Deva, 2005), Cluj Univ. Press, Cluj-Napoca, 2006, 283–293, hep-th/0603051.

[35] Martinetti P., Wulkenhaar R., Discrete Kaluza–Klein from scalar fluctuations in noncommutative geometry, *J. Math. Phys.* **43** (2002), 182–204, hep-th/0104108.

[36] Monge G., Mémoire sur la Théorie des Déblais et des Remblais, Histoire de l'Acad. des Sciences de Paris, 1781.

[37] Nash J., C^1-isometric imbeddings, *Ann. of Math. (2)* **60** (1954), 383–396.

[38] Nash J., The imbedding problem for Riemannian manifolds, *Ann. of Math. (2)* **63** (1956), 20–63.

[39] Rieffel M.A., Metric on state spaces, *Doc. Math.* **4** (1999), 559–600, math.OA/9906151.

[40] Rieffel M.A., Compact quantum metric spaces, in Operator Algebras, Quantization, and Noncommutative Geometry, *Contemp. Math.*, Vol. 365, Amer. Math. Soc., Providence, RI, 2004, 315–330, math.OA/0308207.

[41] Rieffel M.A., Gromov–Hausdorff distance for quantum metric spaces, *Mem. Amer. Math. Soc.* **168** (2004), no. 796, 1–65, math.OA/0011063.

[42] Roe J., Index theory, coarse geometry, and topology of manifolds, *CBMS Regional Conference Series in Mathematics*, Vol. 90, American Mathematical Society, Providence, RI, 1996.

[43] Villani C., Topics in optimal transportation, *Graduate Studies in Mathematics*, Vol. 58, American Mathematical Society, Providence, RI, 2003.

[44] Villani C., Optimal transport. Old and new, *Grundlehren der Mathematischen Wissenschaften*, Vol. 338, Springer-Verlag, Berlin, 2009.