Rev. Mat. Iberoamericana **22** (2006), no. 3, 833–849

A geometry on the space of probabilities II. Projective spaces and exponential families

Henryk Gzyl and Lázaro Recht

Abstract

In this note we continue a theme taken up in part I, see [8], namely to provide a geometric interpretation of exponential families as end points of geodesics of a non-metric connection in a function space. For that we characterize the space of probability densities as a projective space in the class of strictly positive functions, and these will be regarded as a homogeneous reductive space in the class of all bounded complex valued functions. We shall develop everything in a generic $\mathcal{C}^*$-algebra setting, but shall have the function space model in mind.

1. Preliminaries

As we mentioned in [8], exponential families of probability densities have been very much in use in Statistics and Information Theory, see Barndorff-Nielsesn's [3], Kullback's [12] or [13] and Vajda's [19] for example. In [18] and [17], Pistone and Sempi and Pistone and Rogantin, examine a geometric (manifold) structure on the class of probability densities of probabilities equivalent to a given one. Here we shall examine another geometric structure on the space of probabilities. For us, probabilities with densities will be described as representatives of equivalence classes of a projective structure on the class of positive, invertible elements in a special complex Banach algebra. We should also mention the pioneering results by Amari [1] and [2], as well as Efron's [6], where a geometric structure on the class of exponential families is studied, and trace back from [9] and [7] the enormous literature

2000 Mathematics Subject Classification: Primary: 46L05, 53C05, 53C56, 60B99, 60E05. Secondary: 53C30, 32M99, 62A25, 94A17.
Keywords: C^*-algebra, reductive homogeneous space, lifting of geodesics, exponential families, maximum entropy method.

on a topic with which this line of work will eventually connect, namely, that of non-commutative random variables and quantum probability.

To make this note self-contained, we shall recall some results obtained by Corach, Porta and Recht in [4], [15] and [16], and we shall adapt some of what we did in [8]. Our case is simpler than the theory developed by Corach, Porta and Recht, because all the Banach algebras we deal with here are commutative. In section 2, the basic algebraic structure is explained, the connection (its associated distribution of horizontal spaces) and its geodesics are studied. We devote section 3 to the geometry on the class of probability densities regarded as representatives of rays in a projective space. The rays correspond to the strictly positive functions when the $\mathcal{C}^*$-algebra is a function algebra.

Although we present our results in the framework of a generic abstract $\mathcal{C}^*$-algebra $\mathcal{A}$, we urge the reader to keep in mind the standard algebras of function type. For the time being the direct application of our results to convolution algebras seems cumbersome, a fact that could be avoided bringing Fourier analysis into play, but this makes a verbatim application of our results impossible.

Before continuing, we should mention that given a probability space $(\Omega, \mathcal{F}, P)$ and any random variable X such that $E[e^X] < \infty$, clearly $\frac{e^X}{E[e^X]}$ is a positive probability density, and every positive density Z can be so written. The whole point of this note is to understand this fact from a geometric point of view. To paraphrase [18]: "…nothing is so practical as a good theory". We should also add that the $\mathcal{C}^*$-algebra setting is a bit less general that the setting examined in [18] and [17], but in this setting many analytical details are much simpler, and the extension to algebras other than function algebras is included. We mention as well that the characterization of densities in terms of projective structures may allow us to explore connections between projective geometry and statistical inference.

2. The basic $\mathcal{C}^*$ algebra and its properties

In this section we recall the basic facts about the geometry on a commutative $\mathcal{C}^*$ algebra $\mathcal{A}$ with a unit. In particular we define a special connection, describe its geodesics and the parallel transport along them, as well as the resulting geometries on the positive elements and the basic properties of the projective spaces in the set of positive elements in $\mathcal{A}$.

The typical examples of commutative $\mathcal{C}^*$-algebras that come up to mind easily are function algebras and convolution algebras. Let us list a couple of each, but mention that below we have in mind function algebras, even though we state results in an abstract setting.

Example 1 The class of complex valued functions on a probability or finite measure space $(\Omega, \mathcal{F}, Q)$. The conjugation operation $X \to X^*$ is just the standard complex conjugation. The (multiplicative) unit, obviously denoted by 1, regarded as a function. The norm being the usual (essential) supremum norm.

Example 2 When Ω is compact, $\mathcal{C}(\Omega)$, the class of continuous, complex valued functions comprise a $\mathcal{C}^*$-algebra.

Example 3 Let $a < b$ be two real numbers and denote by $\mathcal{C}^n([a,b])$ the class of all n-times continuously differentiable, complex valued functions. The algebra operations are defined pointwise, and the norm is defined by

$$\|X\| = \sup\Big\{ \sum_k \frac{|X^{(k)}(t)|}{k!} \,\Big|\, t \in [a,b] \Big\}.$$

Example 4 Let Ω be either $\mathbb{Z}$ or $\mathbb{N}$ and let

$$\mathcal{A} = \Big\{ X : \Omega \to \mathbb{C} \mid \sum_k |X(k)| < \infty \Big\}.$$

As product operation we consider $X * Y(k) = \sum_{n=a}^{k} X(n)Y(k-n)$, where $a = -\infty$ or $a = 0$ depending on the range of k. The unit is the sequence $1(k) = \delta_{(0,k)}$.

Example 5 The previous example can be generalized considerably. Let Ω denote a commutative topological group, and let $\mathcal{A}$ denote the class of all (necessarily) bounded, σ-finite, complex valued measures on $(\Omega, \mathcal{B}(\Omega))$. Given X, $Y \in \mathcal{A}$, we define their (convolution) product

$$X * Y(B) = \int_\Omega X(B-t)Y(dt); \quad \forall\, B \in \mathcal{B}(\Omega).$$

The unit for this product is the Dirac-point mass measure concentrated at $0 \in \Omega$ and the norm on the space is the total variation measure $\|X\| = |X|(\Omega)$.

In the all of these examples it is clear which the invertible (with respect to the product in the algebra) and the positive elements are. In the last two, the invertibility has to be characterized in terms of Fourier transforms. Also, the conjugation operation $^* : \mathcal{A} \to \mathcal{A}$ is just the ordinary complex conjugation and we shall write $\mathcal{A} = \mathcal{A}^s \oplus i\mathcal{A}^s$. Clearly, in the examples above $\mathcal{A}^s$ denotes the underlying real part of the algebra. Nevertheless, we direct the reader to Pedersen's [14] for details about $\mathcal{C}^*$-algebras, in particular for the definition of the exponential, arbitrary powers and logarithms. But for algebras of function type, this is inherited from the field of complex numbers.

2.1. The basic reductive homogeneous space

In this algebra the set of invertible vectors $G = \{X \in \mathcal{A} \,|\, X^{-1} \text{ exists}\}$ is a (commutative) group and the class $G^+ \subset G$ denotes the class of positive invertible elements. Also, it is an standard result that G is an open set in $\mathcal{A}$ and that the inversion operation is continuously differentiable. This allows us to provide G with a manifold structure modeled on $\mathcal{A}$ regarded as a Banach algebra. In several of the examples mentioned above the invertible elements are just the non-identically zero functions.

Now, to begin with, note that G^+ is a homogeneous space for the group action defined by

$$L_g : G^+ \to G^+; \quad L_g(a) = (g^*)^{-1} a g^{-1}, \;\; \forall\, a \in G^+.$$

for any $g \in G$. Since the product is commutative, $L_g(a) = |g|^{-2}a$. An intuitive way of understanding that mapping is to realize that every $a \in G^+$ defines a scalar product on $\mathcal{H}_a \equiv L_2(aP)$ by

$$< X, Y >_a = E[XYa] = \int XYadP.$$

Now we may interpret the group action as an isometry $\mathcal{H}_a \to \mathcal{H}_{L_g(a)}$.

Now, let us fix some arbitrary $a_0 \in G^+$, and define the projection operator $\pi_{a_0} : G \to G^+$ by means of

$$\pi_{a_0}(g) = L_g(a_0)$$

and notice right away that the fiber (isotropy group) over a_0 is defined by

$$I_{a_0} = \{g \in G \mid \pi_{a_0}(g) = a_0\} = \{g \in G \mid g^* g = 1\},$$

and notice as well that when $\mathcal{A}$ is a function algebra, I_{a_0} is the class of functions taking values in the circle, that is, an infinite dimensional torus. When the product in the algebra is of convolution type, things are not that simple, but loosely speaking, if a notion of Fourier transform exists, the Fourier transforms of the elements of the group are functions taking values in the unit circle.

Since G is an open subset in $\mathcal{A}$, its tangent space at any point is $\mathcal{A}$, i.e.

$$(TG)_1 = \mathcal{A},$$

and it is easy to see that

$$(TI_{a_0})_1 = V_1 = \{0\} \oplus i\mathcal{A}^s\,,$$

which in the non-commutative case corresponds to the anti-hermitian elements in $\mathcal{A}$.

The derivative $(D\pi_{a_0})_1(X)$ of π_{a_0} at 1 in the direction of $X \in \mathcal{A}$ is easy to compute, and it is given by

$$(D\pi_{a_0})_1(X) = -a_0(X + X^*)\,.$$

Clearly

$$(D\pi_{a_0})_1 : \mathcal{A} \to (TG^+)_{a_0} \equiv \mathcal{A}^s \oplus \{0\}.$$

We shall define the horizontal space at $1 \in G$ as

$$H_1 \equiv \{X \in \mathcal{A} \,|\, (a_0)^{-1}X^*a_0 = X\}\,,$$

which can be written as

$$H_1 = \{X \in \mathcal{A} \,|\, X^* = X\} = \mathcal{A}^s \oplus \{0\}$$

and we have the obvious splitting

$$\mathcal{A} = H_1 \oplus V_1.$$

Not only that, the map $(D\pi_{a_0})_1$ is invertible from the left. That is, there exists a mapping $\kappa_{a_0} : (TG^+)_{a_0} \to (TG)_1$, given by

$$\kappa_{a_0}(z) \equiv -\frac{a_0^{-1}}{2}z$$

such that $(TG^+)_{a_0} \xrightarrow{\kappa} (TG)_1 \xrightarrow{(D\pi_{a_0})} (TG^+)_{a_0}$ is the identity mapping.

With these ingredients, a smooth distribution of horizontal and vertical spaces, and a connection can be constructed as follows. For any $a \in G$, $\mathcal{A} = a\mathcal{A} = aH \oplus aV \equiv H_a \oplus V_a$, where $H_a = a\mathcal{A}^s \oplus \{0\}$ and $V_a = \{0\} \oplus ia\mathcal{A}^s$, and we have

Lemma 2.1 *With the notations introduced above*

i) $\mathcal{A} = (TG)_1 a = H_a \oplus V_a$.

ii) $H_{ah} = H_a h$; *for any* $a \in G$ *and any* $h \in I_a$.

To define the connection we set

Definition 2.1 *The* $\mathcal{A}$*-valued 1-form* κ *defined by* $\kappa : G \to \mathcal{A}$

$$\kappa(a) = \kappa_a = \kappa_{a_0} \circ (D\mathcal{L}_g)^{-1} \qquad \text{if } \mathcal{L}_g(a_0) = a\,.$$

Note that κ is **equivariant**, that is $\kappa_a \circ (D\mathcal{L}_g) = \kappa_{a_0}$.

The following is also easy:

Lemma 2.2 *i) κ_a is independent of $g \in G$ such that $\mathcal{L}_g(a_0) = a$.*

ii) κ being equivariant implies

$$H_a = \kappa_a((TG^+)_a) = H_1$$

whenever $\mathcal{L}_g(a_0) = a$, and also

iii) $(D\pi_a) \circ \kappa_a : (TG^+)_a \to (TG^+)_a$ is the identity mapping.

iv) $(D\pi_a)_1$ leaves H_a invariant and $(D\pi_a)_1(H_a) = H_a$.

The $\mathcal{A}$-valued linear mapping κ defined on the tangent bundle (TG^+) is called the **structure 1-form** of the homogeneous space G^+. All the geometry on G^+ comes from κ. This whole setup is part of what Kobayashi and Nomizu call reductive homogeneous structure. See [11] for full details.

2.2. Lifting curves from G^+ to G and parallel transport

Let us begin with a basic lemma. Here the reason of being of the connection κ will become apparent.

Lemma 2.3 *Let $a(t) : [0,1] \to G^+$ be a differentiable curve in G^+. There exists a curve $g(t) : [0,1] \to G$, (called **the lifting of** $a(t)$ **to** G)such that*

$$L_{g(t)}(a_0) = \pi_{a_0}(g(t)) = a(t) \tag{2.1}$$

where the identification $a(0) = a_0$ will be used from now on.

Proof. Let us verify that the solution to the (transport) equation

$$\dot{g}(t) = \kappa_{a(t)}(\dot{a}(t))g((t) \tag{2.2}$$

satisfies (2.1). Here the commutativity makes things really simple. Equation (2.2), explicitly spelled out, is

$$\dot{g}(t) = -\frac{\dot{a}(t)}{2a(t)}g(t); \quad g(0) = 1\,;$$

which can easily be solved to yield

$$g(t) = \left(\frac{a(0)}{a(t)}\right)^{1/2}. \tag{2.3}$$

Note that

$$\pi_{a_0}(g(t)) = \left(\frac{a(t)}{a(0)}\right)^{1/2} a(0) \left(\frac{a(t)}{a(0)}\right)^{1/2} = a(t).$$

∎

The **parallel transport** along the continuous curve $a(t)$ (from $a(0)$ to $a(1)$) is defined in

Definition 2.2 *Let $a(t)$ be a curve in G^+ and let $g(t)$ be its lifting to G. The parallel transport along $a(.)$ is the mapping $\tau(a(.)) : (TG^+)_{a(0)} \to (TG^+)_{a(1)}$ defined by*

$$\tau(a(.))(X) = L_{g(1)}(X) \tag{2.4}$$

We may now say that a differentiable curve $a(t)$ is a **geodesic** if $\dot{a}(0)$ is transported onto $\dot{a}(t)$ by means of the (time) rescaled curve $b(s) := a(st)$, $s \in [0,1]$. From (2.3)–(2.4) it is clear that this amounts to

$$\dot{a}(t) = \frac{a(t)}{a(0)}\dot{a}(0) \quad \Longleftrightarrow \quad \frac{\dot{a}(t)}{a(t)} = \frac{\dot{a}(0)}{a(0)} \equiv X.$$

Or equivalently,

Lemma 2.4 *The curve $a(t)$ is a geodesic if and only if there exists a (real) vector X such that*

$$a(t) = a(0)e^{tX}.$$

Comment 2.1 *This means that the lifted geodesic is given by*

$$g(t) = a(0)^{1/2}e^{-tX/2}.$$

Observe that if we specify the initial and final points of the geodesic, the vector X is automatically determined:

$$a(1) = a(0)e^X \Rightarrow X = \ln\Big(\frac{a(1)}{a(0)}\Big),$$

and the equation of the geodesic can be rewritten as

$$a(t) = a(0)^{1-t}a(1)^t. \tag{2.5}$$

2.3. Semi-norms and Kullback's "distance"

The notion of geodesic introduced above is not related to a metric notion but to a notion of parallel transport. In order to have distance along geodesics, one need to introduce norms in the tangent bundle. What we do below is a bit less: we define a pseudo-metric starting from a pseudo norm, which will lead us to the famous Kullback distance used in Information Theory. Recall that $\mathcal{A}$ is a Banach algebra, and thus comes provided with a norm, which if needed will be denoted by a symbol different that the one we introduce next. For the next definition we shall need the following

Assumption 1 There exists at least one positive linear, continuous functional on $\mathcal{A}$ denoted by $E[Z]$. That is, if $Z \geq 0$ then $E[Z] \geq 0$.

Comment 2.2 *Actually, given any positive, linear functional $E : \mathcal{A} \to \mathbb{C}$, for any $p \in G^+$ we can define $E_p[X] \equiv E[pX]$, and clearly E_p is a linear positive functional on $\mathcal{A}$.*

Let us examine a few candidates for the list of examples given above. In the first example we already have a measure at hand, so $E[Z] = \int Z dQ$. For the next two examples, just consider any finite measure Q on the underlying space, and do as above. A good candidate for example 4 is given by the pairing

$$E[X] \equiv < 1, X > \equiv \sum X(n),$$

and similarly, for example 5,

$$E[X] \equiv < 1, X > \equiv X(1) = \int 1 dX.$$

For reasons that will be clear below, we shall consider the following semi-norm: For any $Z \in \mathcal{A}$ define

$$\|Z\| = |E[Z]|.$$

To define a metric on TG^+, we begin by defining it at $(TG^+)_1$ by $\|X\|_1 \equiv \|X\|$, and transporting it to any other $(TG^+)_{a_0}$ by means of the group action: that is, we set

$$\|X\|_{a_0} = |E[a_0^{-1} X]|.$$

It is easy to verify that this is a consistent definition. Note now that if $a(t)$ is a geodesic joining $a(0)$ to $a(1)$ in G^+, then the "length" of the velocity vector along the geodesic is constant:

$$\|\dot{a}(t)\|_{a(t)} = |E[a(t)^{-1} a(t) X] = \|X\|_1 = \Big\| \ln \Big(a(1)/a(0) \Big) \Big\|$$

and therefore, the geodesic distance from $a(0)$ to $a(1)$ is given by

$$d(a(0), a(1)) = \int_0^1 \|\dot{a}(t)\|_{a(t)} dt = \|X\|_1 = \Big\| \ln \Big(a(1)/a(0) \Big) \Big\|.$$

Comment 2.3 *Note that once we have a positive linear functional, we can rapidly come up with variations on the theme. For example, for a fixed $p \in G^+$ we might have defined*

$$\|X\|_{(p), a(0)} := |E_p[a(0)^{-1} X]| = |E[pa(0)^{-1} X]|$$

and we would have ended up with

$$\|\dot{a}(t)\|_{(p) a(t)} = |E[pa(t)^{-1} a(t) X] = \|X\|_{(p)} = \|p \ln \Big(a(1)/a(0) \Big) \|,$$

which is still symmetric in $a(0)$, $a(1)$. *If we chose* $p = a(1)$, *we would end up with*

$$\|X\|_{a(1)} = \left|E[a_1 \ln\left(a_1/a_0\right)]\right|$$

which is not symmetric anymore (the norm depends on the final point of the trajectory). It is a simple application of Jensen's inequality to verify that when a_1 *and* a_0 *are densities (with respect to the same probability), that is* $E[a_1] = E[a_0] = 1$, *then* $K(a_1, a_0) \equiv E[a_1 \ln\left(a_1/a_0\right)] \geq 0$. $K(a_1, a_0)$ *is called the Kullback distance between* a_1 *and* a_0.

In [12] it is proved that $a_1 = a_0$ if and only if $K(a_1, a_0) = 0$. But this is neither a symmetric function nor does it satisfy a triangular inequality. So it is not a true distance even when restricted to the class of densities. But it has nevertheless proven enormously useful in Information Theory, Statistics and Inverse Problems.

In the context of our first three examples, that is when $\mathcal{A} = L_\infty(\Omega, \mathcal{F}, Q)$, and $E[X] = \int X dQ$, then for positive a_1 and a_0 we have the obvious

$$\|X\|_{(p)} = \int p \ln\left(\frac{a_1}{a_0}\right) dQ$$

where as above, $X \equiv \ln\left(\frac{a_1}{a_0}\right)$.

2.4. Conditional expectations and additive and multiplicative decompositions

Consider a sub-algebra $\mathcal{B}$ of the algebra $\mathcal{A}$, that is a linear subspace, closed in the topology induced by the original norm, and closed with respect to the multiplicative and conjugation operations as well. In the setup of examples (1)-(3), think for instance on the functions measurable with respect to a sub-σ-algebra $\mathcal{G}$ of the given σ-algebra $\mathcal{F}$ on Ω. Let us now make the following

Assumption 2 We shall assume the existence of a positive, linear, orthogonal projection $E_{\mathcal{B}} : \mathcal{A} \to \mathcal{B}$ such that $E_{\mathcal{B}}[XY] = Y E_{\mathcal{B}}[X]$, $\forall X \in \mathcal{A}$ and $\forall Y \in \mathcal{B}$.

Note that when $\mathcal{B} = \mathbb{C}$, then any E_p such that $E_p[1] = 1$ satisfies the definition. In the setup of examples (1)-(3), for any fixed probability measure Q on Ω, denote by $E_{Q,\mathcal{G}}$ the usual conditional expectation. This is an orthogonal projection of $\mathcal{A}$ onto $\mathcal{B}$ satisfying $E_{Q,\mathcal{G}}[XY] = E_{Q,\mathcal{G}}[X]Y$ for any Y in $\mathcal{B}$.

This conditional expectation induces a decomposition $\mathcal{A} = \mathcal{B} \oplus K$, where to simplify notation we set $K \equiv \ker(E_{\mathcal{B}})$. It induces also a similar decomposition on $\mathcal{A}^s \simeq (TG^+)_1$ given by $\mathcal{A}^s = \mathcal{B}^s \oplus K^s$.

The interesting thing is that this decomposition can be lifted to the set of positive vectors, i.e., to G^+ by means of the exponential map. That is, if $q \in G^+$, and $q = e^X$ for $X \in A^s$, then

$$q = e^X = e^{E_{\mathcal{B}}[X]} e^{X - E_{\mathcal{B}}[X]}$$

and, in the case that will be important for us below, when $\mathcal{B} = \mathbb{C}$, which in our standing class of examples, corresponds to the case in which the σ-algebra defining $\mathcal{B}$ is trivial, and $E_{\mathcal{B}} = E_p$ is as above, the last decomposition will look like

$$q = e^X = e^{E_p[X]} e^{X - E_p[X]} . \tag{2.6}$$

2.5. The $\mathcal{B}$-projective structure on the class G^+

We want to define equivalence classes (modulo $\mathcal{B}$) in such a way that they are preserved under the action of $G_{\mathcal{B}}$, the group of invertible elements in $\mathcal{B}$, that is, under the action of the mapping $a \rightarrow L_g(a) = (g^*)^{-1} a g^{-1}$, but for $g \in G_{\mathcal{B}}$. In particular we want the relation, denoted by $\sim_{\mathcal{B}}$, to be such that, if $a(t)$ is a curve in G^+, then $\tilde{a}(t) = L_g(a(t)) \sim_{\mathcal{B}} a(t)$, that is, the resulting projective structure is to be preserved under the action of L_g. In particular, since we shall be transporting tangent vector fields, we will want the tangent

$$\tilde{X} = \dot{\tilde{a}}(0) = \frac{1}{|g|^2}\Big(X - a\Big(\frac{V}{g} + \frac{V^*}{g^*}\Big)\Big)$$

to be somehow equivalent to X. Here, $V = \dot{g}(0)$. For that, note that the previous identity can be rewritten as

$$\frac{\tilde{X}}{\tilde{a}} = \frac{X}{a} + W ,$$

where $W = -(\frac{V}{g} + \frac{V^*}{g^*})$ is a symmetric element in $\mathcal{B}$. Now we state

Definition 2.3 *With the notations introduced above, we say that a and $\tilde{a}$, both in G^+, are $\sim_{\mathcal{B}}$ if and only if*

$$\frac{\tilde{a}}{a} \in G^+_{\mathcal{B}} .$$

Comment 2.4 *Notice that if $\frac{\tilde{a}(t)}{a(t)} = h(t) \in G^+_{\mathcal{B}}$ and $g(t) \in G_{\mathcal{B}}$ is any square root of $h(t)^{-1/2}$, then, taking logarithms and differentiating at $t = 0$, we obtain*

$$\frac{\tilde{X}}{\tilde{a}} = \frac{X}{a} - (\frac{V}{g} + \frac{V^*}{g^*}).$$

That is, the equivalence relation may be lifted to $G^+ \times \mathcal{A}^+$, which should be regarded as a trivial tangent bundle.

We can form the quotient space $G^+/\sim_{\mathcal{B}}$ and verify that $G^+/\sim_{\mathcal{B}}\simeq \mathbb{P}_{\mathcal{B}} \equiv \{\alpha \in G^+ \mid E_{\mathcal{B}}[\alpha] = 1\}$, where the equivalence is brought about by the mapping $\Phi_p : G^+ \to \mathbb{P}_{\mathcal{B}}$ and the following

Lemma 2.5 *With the notations introduced above, $\tilde{a} \sim_{\mathcal{B}} a$ if and only if $\Phi_p(\tilde{a}) = \Phi_p(a)$, where $\Phi_p(a) = \frac{a}{E_{\mathcal{B}}[a]}$.*

Proof. Let $\tilde{a} = ah$, where $h \in G^+_{\mathcal{B}}$. Therefore, $E_{\mathcal{B}}[\tilde{a}] = hE_{\mathcal{B}}[a]$ and $\Phi_p(\tilde{a}) = \Phi_p(a)$. Conversely, if $\Phi_p(\tilde{a}) = \Phi_p(a)$, then

$$\tilde{a} = a\frac{E_{\mathcal{B}}[\tilde{a}]}{E_{\mathcal{B}}[a]},$$

thus $\frac{\tilde{a}}{a} \in G^+_{\mathcal{B}}$. ∎

Comment 2.5 *When $\mathcal{B} = \mathbb{C}$ is the trivial sub-algebra then $G^+_{\mathcal{B}} = [0,\infty)$, and if E_p is the projection onto $\mathcal{B}$, then we have*

$$\Phi_p(a) = \frac{a}{E_p[a]}.$$

In this case we write $\mathbb{P}_{\mathcal{B}} = \mathbb{P}_p$, and $\mathbb{P}_p$ can be regarded as the class of probabilities equivalent to E_p. Also, in this case we shall put $\sim$ instead of $\sim_{\mathcal{B}}$ to simplify the notation.

To define the action of G on $G^+/\sim_{\mathcal{B}}$ we proceed as usual: if $[a]$ denotes the equivalence class of $a \in G^+$, then we put $L_g[a] = [L_g a]$, and we have the following simple result:

Lemma 2.6 *Let $g \in G$ and let $[a] = [b]$. Then $[L_g(a)] = [L_g(b)]$.*

Proof. Invoking Lemma 2.5, it suffices to see that $\Phi_p([L_g(a)]) = \Phi_p([L_g(b)])$. For that it is enough to note that $b = ha$, where $h \in G^+_{\mathcal{B}}$, from which the desired conclusion follows. ∎

Comment 2.6 *The notion of $\mathcal{B}$-equivalence can be related to the notion of* ***sufficiency****. As a matter of fact, the analogue of Theorem 3.2 in chapter 3 of [13] asserts that $\mathcal{B}$ is sufficient for a family $\{a_t\} \subset \mathcal{A}$ if and only if there exists $b \in \mathcal{A}$ such that $a_t \sim_{\mathcal{B}} b$.*

2.6. Geometry on $\mathbb{P}_{\mathcal{B}}$

In the finite dimensional case, one can either introduce homogeneous coordinates, or work with a well chosen class of representatives for $G^+/\sim_{\mathcal{B}}$. As a matter of fact, we did that in [8]. Let us repeat what is needed here. To define the geodesic curves, we proceed as above and examine what the isotropy group is and how to describe the vertical and horizontal spaces.

We already know how does G act on $\mathbb{P}_{\mathcal{B}}$. Let $\alpha \in \mathbb{P}_{\mathcal{B}}$ and let us consider the mapping $\hat{\pi}_\alpha : G \to G$ and define the isotropy group of this action by $\hat{I}_\alpha = \{g \in G \,|\, \hat{\pi}_\alpha(g) = \alpha\}$.

Clearly, the tangent space to $\mathbb{P}_{\mathcal{B}}$ is $\{X \in \mathcal{A}^s \,|\, E_{\mathcal{B}}[X] = 0\} = K^s$. Note that if $g(t)$ is any curve in G such that $g(0) = 1$ and $\dot{g}(0) = X$, then

$$(D\hat{\pi})_1(X) = -\alpha(X + X^*) + \alpha E_{\mathcal{B}}[(X + X^*)\alpha] \in K^s$$

(This is easy to see differentiating

$$\frac{|g(t)|^{-2}\alpha}{E_{\mathcal{B}}[|g(t)|^{-2}\alpha]}$$

at $t = 0$). Note also that if $X \in \mathcal{B}$, then $(D\hat{\pi})_1(X) = 0$. Note as well that the tangent space $(T\hat{I}_\alpha)_1 = K^a$, *i.e.*, it consists of those antisymmetric elements X of $\mathcal{A}$ that have zero trace ($E_{\mathcal{B}}[X] = 0$).

Therefore $\mathcal{A} = \mathcal{B} \oplus K^s \oplus K^a$. Starting from this (which is a decomposition of the tangent space to G at $g = 1$) we can define a distribution of horizontal spaces by $H_g = \mathcal{B} \oplus \{gX \,|\, X \in K^s\}$. Again, to lift curves in $\mathbb{P}_{\mathcal{B}}$, we need a connection, this time defined as follows: for $\alpha \in \mathbb{P}_{\mathcal{B}}$ and $Y \in T_\alpha \mathbb{P}_{\mathcal{B}}$, we put

$$\kappa_\alpha(Y) = -\frac{1}{2}\alpha^{-1}Y.$$

Clearly, for $Y \in \mathbb{P}_{\mathcal{B}}$ we have $(D\hat{\pi})_1(\kappa_\alpha(Y)) = Y$. Now, to lift curves and define geodesics we can proceed verbatim as above.

3. Exponential coordinates on $\mathbb{P}_p$

We saw above that we can identify rays in G^+ with points in $\mathbb{P}$ via an equivalence relation. That is, we identify lines in G^+ with the point they intersect at $\mathbb{P}_p$; that is, $\mathbb{P}$ can be regarded as a projective space. In other words, as the quotient space $G^+/\sim$, where $\sim$ is the equivalence relation of Definition 2.3. In particular, we saw in Lemma 2.5 that

$$a(1) \sim a(2) \quad \text{whenever} \quad \frac{a(1)}{E_p[a(1)]} = \frac{a(1)}{E_p[a(2)]}.$$

Just before (2.5) we saw that given two points a_0 and a_1, there is vector field $X = \ln(\frac{a_1}{a_0})$ such that $\gamma(t) = a_0 e^{tX}$ is the geodesic joining a_0 to a_1 in G^+ (and G).

Note that the trace on $\mathbb{P}_p$ of a geodesic given by (2.5), or if you prefer, the equivalence class of each point of the geodesic, is given by

$$\gamma(t) = \frac{q(t)}{E_p[q(t)]} = \frac{q(0)^{1-t}q(1)^t}{E_p[q(0)^{1-t}q(1)^t]} \tag{3.1}$$

The geometric interpretation of (3.1) as the representative in $\mathbb{P}_p$ of the rays through the geodesic described in (2.5) is clear now. For earlier appearances of these curves in the context of Information Theory see chapter 3 of [12], where a guide to earlier references is provided.

Note as well that Lemma 2.6 asserts that collinear points stay collinear under the action L_g. Consider now fixed $p \in G^+$ and $q_0 \in \mathbb{P}_p \subset G^+$. Since $L_g : G^+ \to G^+$ is bijective, then there exists $g \in G$ such that

$$q = \frac{|g|^{-2} q_0}{E_p[|g|^{-2} q_0]}.$$

That is, all probabilities equivalent to E_p can be obtained from any given q_0 by means of the group action. While this is clear in the examples (1)-(3) when we consider probability laws with strictly positive densities, the geometric interpretation of the fact is nice.

3.1. Exponential Families

Let us now examine a bit further in what sense $\frac{e^X}{E_p[e^X]}$ is natural in our setup. Set $a(0) = 1$ and let $a(1)$ be any other point in G^+. We now know that there exists a real vector X, actually given by $X = \ln a(1)$, such that $a(t) = e^{tX}$ joins 1 geodesically to $a(1)$, and the trace on $\mathbb{P}_p$ of this geodesic is $q(t) = a(t)/E_p[a(t)]$ or if you prefer,

$$q(t) = \frac{e^{tX}}{E_p[e^{tX}]}.$$

That is, we have a correspondence between vectors in $\mathcal{A}$ regarded as tangent vectors to TG^+ and probabilities in $\mathbb{P}_p$, which we shall now explore further.

We shall consider the mapping

$$\Phi : (TG^+)_1 \simeq \mathcal{A} \to \mathbb{P}_p \simeq G^+/\sim$$

given by

$$X \to \Phi(X) = \frac{e^X}{E_p[e^X]} \tag{3.2}$$

and now we shall examine some of the basic properties of this map. Observe first that $\Phi(X) = \Phi(X + \alpha 1)$ for any $\alpha \in \mathbb{C}$. Thus Φ as defined cannot be a bijective map. Positive collinear vectors differ by a factor of $e^{\alpha 1}$ for appropriate α.

Recall that to understand this more algebraically we noted that $\mathcal{A}^s \simeq (TG^+)_1 = \mathcal{B} \oplus K^s$, where $K = \ker E_p[.]$ and K^s is the class of real, centered random variables, and for this we want to regard the expected value as a linear mapping from $\mathcal{A}^s$ onto a commutative algebra $\mathcal{B}^s$ (which in this case coincides with $\mathbb{R}$). This additive decomposition at the Lie algebra level induces a multiplicative decomposition at the group level. That is, we can write any positive element in $g \in G^+$ as

$$g = e^X = e^{E_p[X]} e^{X - E_p[X]}.$$

This establishes a mapping from $\mathcal{B} \times C$ where $C = \{e^Y \mid E[Y] = 0\}$ onto G^+. We thus obtain an approach to exponential families somewhat similar to that of Pistone and Sempi in [18] motivated by the work of Porta and Recht in [16].

Note now that the projection

$$g = e^{E_p[X]} e^{X - E_p[X]} \to \frac{e^{X - <1,X>}}{E[e^{X - <1,X>}]}$$

is independent of $e^{E_p[X]}$. This motivates the following: To make the map Φ a bijection, we have to restrict its domain. Thus if we define

$$\Phi_p : K^s \to \mathbb{P}_p; \quad Y \to \Phi_p(Y) = \frac{e^Y}{E_p[e^Y]} \tag{3.3}$$

we have a bijection, the inverse mapping being given by

$$\Phi_p^{-1} : q \to Y = \ln q - E_p[\ln q].$$

To conclude, we note that the special role played by the vector p can be done away as follows: We could have chosen any other positive $q \in G^+$ and we could have defined the standard $E_q[X]$. Observe that $a = qp^{-1} \in \mathbb{P}_p$ because $E_p[a] = 1$ and therefore $a = a/E_p[a]$.

Summing up, if we put $K_q \equiv (\ker E_p[.])^s$, or more explicitly $K_p = \{X \in \mathcal{A}_0 \mid E_q[X] = 0\}$, which is a linear subspace of $\mathcal{A}_0$ on which the following maps are defined

$$\begin{aligned} \Psi_p(Y) : K_p &\to \mathbb{R} \\ Y &\to \Psi_p(Y) = \ln E_p[e^Y], \\ \Phi_p(Y) : K_p &\to \mathbb{P}_p \\ Y &\to \Phi_p(Y) = e^{Y - \Psi_p(Y)} p = \frac{e^Y}{E_p[e^Y]} p. \end{aligned}$$

and every Φ_p maps K_p bijectively onto $\mathbb{P}_p$.

Now it is not hard to see that the collection $\{K_p, \Phi_p\}$ provide an atlas for $\mathbb{P}_p$ modeled on the (closed) linear subspace K_p of $\mathcal{A}_0$, the coordinate maps Φ_p are globally defined, and that the changes of coordinates $\Phi_q^{-1} \circ \Phi_p : K_p \to K_q$ are affine maps.

3.2. Some transformation properties

Let $\mathcal{A}^1$ and $\mathcal{A}^2$ be two $\mathcal{C}^*$-algebras. Let $T : \mathcal{A}^1 \to \mathcal{A}^2$ be a $\mathcal{C}^*$-algebra morphism, that is T is linear, $T(ab) = T(a)T(b)$ and $T(a^*) = T(a)^*$. Let $E : \mathcal{A}^2 \to \mathbb{C}$ be a linear, positive functional, such that $E[1] = 1$. Then $E^T : \mathcal{A}^1 \to \mathbb{C}$, defined by

$$E^T[a] = E[T(a)] \quad \forall\, a \in \mathcal{A}^1$$

clearly defines linear, positive functional, such that $E^T[1] = 1$.

In the context of the function algebra models, assume you are given a measurable mapping $T : (\Omega_2, \mathcal{F}_2) \to (\Omega_1, \mathcal{F}_1)$. By means of T we can pull back functions $X : \Omega_1 \to \mathbb{C}$ onto functions $T(X) \equiv X \circ T : \Omega_2 \to \mathcal{C}$. Given any measure Q on $(\Omega_2, \mathcal{F}_2)$, it can be pushed forward onto a measure on $(\Omega_1, \mathcal{F}_1)$. The issue to examine is what happens to the quotient spaces we considered above.

Lemma 3.1 *Let T be a positive and surjective $\mathcal{C}^*$-algebra morphism as above. For $p \in G_1^+$, denote by E_p^T and put $\Phi_p^T(a) = \frac{a}{E_p^T[a]}$. Then $T \circ \Phi_p^T(a) = \Phi_{Tp} \circ T(a)$.*

Proof. The proof is easy. Note that $E_p^T[a] = E^T[pa] = E[T(ap)] = E_{Tp}[Ta]$, and therefore

$$\Phi_{Tp} \circ T(a) = \Phi_{Tp}(Ta) = \frac{Ta}{E_{Tp}[Ta]} = \frac{Ta}{E_p^T[a]} = T(\frac{a}{E_p^T[a]}) = T \circ \Phi_p^T(a).$$

■

Comment 3.1 *Put $K_p^T = \{X \in \mathcal{A}_{1,0} \,|\, E_p^T[X] = 0\}$, then $TX \in K_{Tp} = \{Y \in \mathcal{A}_{2,0} \,|\, E_{Tp}[Y] = 0\}$, then we can replace e^X for a in the previous Lemma.*

4. Concluding remarks

We saw that we can define a homogeneous, reductive space structure upon the positive densities equivalent to a given one, such that the curves given by (2.5) or (3.1) are the geodesics of a non Riemannian connection defined on the class on non-vanishing, complex valued, bounded functions on $(\Omega, \mathcal{F}, Q)$. In this setup, the exponential families of densities appear as representatives of equivalence classes of appropriately defined equivalence relation associated to an expectation, and the entropy functional

$$K(p,q) = \int p_1(\omega)\ln\Big(\frac{p_1(\omega)}{p_0(\omega)}\Big)dQ(\omega)$$

is the norm of the velocity at the end point of the geodesic joining p_1 to p_0.

Aknowledgements. We want to thank the referee for his comments and for directing us to very interesting literature on related problems.

References

[1] Amari, S.: Differential geometry of curved exponential families. Curvature and information loss. *Ann. Statist.* **10** (1982), no. 2, 357–385.

[2] Amari, S. and Nagaoka, H.: *Methods of information geometry*. Translations of Mathematical Monographs **191**. American Mathematical Society, Providence, RI, 2000; Oxford University Press, Oxford, 2000.

[3] Barndorff-Nielsen, O.: *Information and exponential families in statistical theory.* Wiley Series in Probability and Mathematical Statistics. Wiley, Chichester, 1978.

[4] Corach, G.; Porta, H. and Recht, H.: The geometry of the space of selfadjoint invertible elements in a C^*-algebra. *Integral Equations Operator Theory* **16** (1993), no. 3, 335–359.

[5] Efron, B.: The geometry of exponential families. *Ann. Statist.* **6** (1975), no. 2, 362–376.

[6] Escher, F.: On the probability function in the collective theory of risk. *Scandinavisk Aktuarietidskrift* **15** (1932), 175–195.

[7] Gibilisco, P. and Isola, T.: Connections on statistical manifolds of density operators by geometry of non-commutative L_p spaces. *Infin. Dimens. Anal. Quantum Probab. Relat. Top.* **2** (1999), no. 1, 169–178.

[8] Gzyl, H. and Recht, L.: The geometry on the class of probabilities I: The finite dimensional case. *Rev. Mat. Iberoamericana* **22** (2006), 545–558.

[9] Grasselli,. M. R. and Streater, R. F.: On the uniqueness of the Chenstov metric in quantum information theory. *Infin. Dimens. Anal. Quantum Probab. Relat. Top.* **4** (2001), no. 2, 173–182.

[10] Jaynes, E.: Information Theory and Statistical Physics. *Phys. Rev.* **106** (1957), 620–630.

[11] Kobayashi, S. and Nomizu, K.: *Foundations of Differential Geometry*, vol. II. Interscience Tracts in Pure and Applied Mathematics **15**, vol. II. Interscience Publishers John Wiley & Sons, New York-London-Sydney, 1969.

[12] Kullback, S.: *Information Theory and Statistics.* Dover Pubs. New York, 1968.

[13] Kullback, S., Keegel, J. and Kullback, J.: *Topics in Statistical Information Theory*. Lecture Notes in Statistics **42**. Springer-Verlag, Berlin, 1987.

[14] Pedersen, G.: *C^*-Algebras and their Automorphism Groups*. London Mathematical Society Monographs **14**. Academic Press, London-New York, 1979.

[15] Porta, H. and Recht, L.: Conditional expectations and operator decompositions. *Ann. Global Anal. Geom.* **12** (1994), no. 4, 335–339.

[16] Porta, H. and Recht, L.: Exponential sets and their geometric motions. *J. Geom. Anal.* **6** (1996), no. 2, 277–285.

[17] Pistone, G. and Rogantin, M. P.: The exponential statistical manifold: mean parameters, orthogonality and space transformations. *Bernoulli* **5** (1999), no. 4, 721–760.

[18] Pistone, G. and Sempi, C.: An infinite dimensional geometric structure on the space of all probability measures equivalent to a given one. *Ann. Statist.* **23** (1995), no. 5, 1543–1561.

[19] Vajda, I.: *Theory of statistical inference and information*. Kluwer Acad. Pubs., Dordrecht, 1989.

Recibido: 21 de mayo de 2004
Revisado: 27 de junio de 2005

Henryk Gzyl
Departamento de Estadística
Universidad Carlos III de Madrid
C/ Madrid, 126, 28903-Getafe, Madrid (Spain)
and
Departamento de Estadística y Cómputo Científico
Universidad Simón Bolívar
AP 89000,Caracas 1080-A,Venezuela
hgzyl@est-econ.uc3m.es, hgzyl@usb.ve

Lázaro Recht
Departamento de Matemática
Universidad Simón Bolívar
AP 89000, Caracas 1080-A (Venezuela)
recht@usb.ve