

# INTRINSIC GEOMETRY ON THE CLASS OF PROBABILITY DENSITIES AND EXPONENTIAL FAMILIES

HENRYK GZYL AND LÁZARO RECHT

*Abstract*

---

We present a way of thinking of exponential families as geodesic surfaces in the class of positive functions considered as a (multiplicative) sub-group  $G^+$  of the group  $G$  of all invertible elements in the algebra  $\mathcal{A}$  of all complex bounded functions defined on a measurable space. For that we have to study a natural geometry on that algebra. The class  $\mathcal{D}$  of densities with respect to a given measure will happen to be representatives of equivalence classes defining a projective space in  $\mathcal{A}$ . The natural geometry is defined by an intrinsic group action which allows us to think of the class of positive, invertible functions  $G^+$  as a homogeneous space. Also, the parallel transport in  $G^+$  and  $\mathcal{D}$  will be given by the original group action. Besides studying some relationships among these constructions, we examine some Riemannian geometries and provide a geometric interpretation of Pinsker's and other classical inequalities. Also we provide a geometric reinterpretation of some relationships between polynomial sequences of convolution type, probability distributions on  $\mathbb{N}$  in terms of geodesics in the Banach space  $\ell_1(\alpha)$ .

---

## 1. Introduction and preliminaries

Exponential families of probability densities appear at least in three different ways: On one hand, they appear as parametric distributions for which unbiased estimators achieve their maximum efficiency (or minimum variance). Even though the result is a textbook matter, see [LM] or [W] for example, let us briefly describe it following the presentation in [S] (which together with [GP] are a good starting point for a fertile interaction between probability, physics and geometry): Let

---

2000 *Mathematics Subject Classification*. Primary: 46L05, 53C05, 53C56, 62B01, 60B99, 60E05; Secondary: 53C30, 51M05, 55M, 62A25, 22E, 33E.

*Key words*. Exponential families, projective geometry, parallel transport, sequences of convolution type.

$\mathbf{T} = (T_1, \dots, T_k)$  be an unbiased estimator of  $\Theta = (\theta_1, \dots, \theta_k)$ , unknown parameters of a family  $\Pi = \{p(x, \Theta)\}$  describing the probability densities of an  $(S, \mathcal{S})$ -valued random variable  $X$  with respect to a given probability measure  $m(dx)$ . If we put  $V_{ij} = E_{\Theta}[(T_i - \theta_i)(T_j - \theta_j)]$ , and under the usual regularity assumptions we set  $G_{ij} = E_{\Theta} \left[ \frac{\partial \ln p(x, \Theta)}{\partial \theta_i} \frac{\partial \ln p(x, \Theta)}{\partial \theta_j} \right]$ , then in matrix terms the Cramer-Rao inequality asserts that  $(V - G^{-1})$  is positive, and when  $VG = I$ , then the minimum variance or maximum efficiency is achieved. In this case, this amounts to saying that  $p(x, \Theta) = c(\Theta)^{-1} \exp\{\langle \Theta, \mathbf{T} \rangle\}$ .

The second way in which exponential families appear is related to the notion of sufficiency. To cite an old result on the relation of exponential families and sufficient statistics, consider the following result by Brown, [Br], improving on a result by Dynkin:

**Theorem 1.1.** *Let  $\Pi = \{p(x, \Theta)\}$  be a family of densities on an interval  $I$  such that every  $p(x, \Theta)$  is bounded away from zero and continuously differentiable on  $I$ . Suppose that there is a nontrivial sufficient statistics  $\mathbf{T}$  for  $\Theta$  on the basis of  $n$ -observations. Then  $\Pi$  is a  $k$ -parameter exponential family with  $k < n$ .*

For much more about this and properties of exponential families, their use in statistics and information theory, some classical references are the volumes by Barndorff-Nielsen [B], Kullback [K] and Vajda [V].

As we shall mention later on, this old, important theorem is recalled here for two reasons: on one hand it brings forth the relationship between exponential families and sufficiency, and on the other, it makes the special class of densities we deal with (namely the invertible elements in the algebra  $\mathcal{A}$  defined below), appear as the natural class to consider.

The third road to exponential families comes from the solution of moment (or generalized moment) problems by information theoretic or maximum entropy methods. To state it as simple as possible, consider the problem of finding a density  $p(x)$  on the measure space  $(S, \mathcal{S}, m)$  such that the expected value of an  $\mathbb{R}^K$ -valued random variable  $X$  is preassigned to be  $E_p[X] = \int X(x)p(x)m(dx) = \mu$ , where  $\mu \in \mathbb{R}^K$  is a preassigned vector. The problem is not trivial when  $K = 1$  only when  $m(S) = \infty$ . This problem seems to have been first solved by Esscher in [Es], although it was during the 1950's that the idea was systematized into a variational method to lay down the foundations of statistical mechanics by Jaynes in [J].

The interplay between geometry and probability has proceeded along two major directions. One described in [Ef], [A] and the collection

[**ABKLR**], which goes from a geometry induced on the space of parameters to statistical properties of the parameterized family of densities, and the other which can be traced back from [**GP**] and [**PRo**], where the authors examine a geometric (manifold) structure on the class of probability densities of probabilities equivalent to a given one. Regretfully there does not seem to be an easy connection between these approaches and the one we develop here.

Here we shall examine a completely different geometric structure on the space of probabilities, that is, an approach that does not bear any relationship to that described in the above mentioned references. For us, probability distributions with density with respect to a given measure (or probability) will be described by representatives of a projective structure on the class of positive, invertible elements in a special complex Banach algebra, namely the Banach algebra of complex, measurable functions defined on a given measure space.

To make this note self contained, instead of referring the reader to [**GR**] where the basics were outlined, we shall again recall some results obtained by Corach, Porta and Recht in [**CPR**], [**PR1**] and [**PR2**], from which we draw freely. Our case is simpler than the theory developed there, because all the Banach algebras with which we deal with here are commutative. This is done in Section 2. In that section we obtain some properties of the geodesics in the space of positive invertible functions and we examine two Riemannian structures on that space. There we provide a geometric way of understanding Pinsker's inequalities. In Section 3 we consider two different Riemannian structures on the class of probability densities thought of as representatives of rays in the class of positive invertible elements of the algebra, that is, as representatives of a natural equivalence relation. The geometry of that projective space is transported onto the class of densities which will be representatives of the rays. In Section 4 we provide the geometrical characterization of exponential families in terms of exponential surfaces in the intrinsic geometry.

A shortcoming in the above presentation, is that the  $C^*$ -algebra to which it readily applies is the algebra of bounded, measurable functions, and for probabilistic applications one is interested in more general algebras. In Section 5 we explore the geometry on the space  $\ell_1(\alpha)$ , which is a convolution algebra, but not a  $C^*$ -algebra. A motivation for looking into this example is that it is the simplest example of convolution algebra which contains the measures on a countable set. Surprisingly enough, a variety of connections between probability and polynomial convolution sequences acquire a new meaning in the geometric setup that we propose.

## 2. The fundamental $C^*$ algebra and its properties

In this section we recall the basic facts about the geometry on a commutative  $C^*$  algebra  $\mathcal{A}$  with a unit. The basic model we should have in mind is  $\mathcal{A} = \{f: S \rightarrow \mathbb{C} \mid \|f\| = \text{esssup}(|f|) < \infty\}$ , where  $(S, \mathcal{S}, m)$  is a given measure space. We shall assume that the measure  $m$  is either a probability or a finite measure. We will define a special connection, describe its geodesics and the parallel transport along them, as well as the resulting geometries on the positive elements of the algebra, as well as the basic properties of the projective spaces in the set of positive invertible elements in  $\mathcal{A}$ . In this algebra the set of invertible (with respect to the usual pointwise multiplication as the product on the algebra) vectors  $G = \{g \in \mathcal{A} \mid g^{-1} \text{ exists}\}$  is a (commutative) group and the class  $G^+ \subset G$  denotes the class of positive invertible elements. Also, it is 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 structured modeled on  $\mathcal{A}$  thought of as a Banach algebra and that the Lie algebra of  $G$ , i.e. the tangent space to  $G$  at 1, is  $\mathcal{A}$ .

### 2.1. The basic reductive homogeneous space.

We want to think of  $G^+$  both as a homogeneous space and as the base space for a principal bundle. For that, we first need an action of  $G$  on  $G^+$ . We define

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

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  $\mathcal{H}_a \equiv L_2(am)$  by  $\langle X, Y \rangle_a = \int XYa \, dm$ . Now we may interpret the group action as an isometry  $\mathcal{H}_a \rightarrow \mathcal{H}_{L_g(a)}$ . This action is transitive and it is easy to see that  $G^+ \simeq G/I_a$ , where  $I_a$  is defined a few lines below. There is a well known way to define connections in this setup, see Chapters 10 and 11 of [KN] for example.

To define the principal bundle, fix some  $a \in G^+$  and define the projection operator

$$\pi_a: G \longrightarrow G^+ \text{ by means of } \pi_a(g) = L_g(a)$$

and notice right away that the isotropy group of  $a$  defined by

$$I_a = \{g \in G \mid \pi_a(g) = a\} = \{g \in G \mid g^*g = 1\},$$

verifies that  $\pi_a^{-1}(a) = I_a$ . Moreover the fiber  $\pi_a^{-1}(b)$  over  $b \in G^+$  is given by  $I_a h$  for appropriate  $h \in G$ . Notice that when  $\mathcal{A}$  is a function algebra,  $I_a$  is the class of functions taking value in the circle.

We mentioned that  $G$  is an open subset in  $\mathcal{A}$ , its tangent space at 1 is  $\mathcal{A}$ , i.e.

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

and it is easy to see that

$$(TI_a)_1 = V_a = \{0\} \oplus i\mathcal{A}^s.$$

The group action can be used to move this splitting around, and the existence of such splitting is equivalent to the existence of connections. The derivative  $(D\pi_a)_1(X)$  of  $\pi_a$  at 1 in the direction of  $X \in \mathcal{A}$  is easy to compute, and it is given by

$$(D\pi_a)_1(X) = -a(X + X^*).$$

Clearly

$$(D\pi_a): \mathcal{A} \longrightarrow (TG^+)_a \equiv \mathcal{A}^s \oplus \{0\}.$$

We shall define the horizontal space at  $a$  by

$$H_a \equiv \{X \in \mathcal{A} \mid (a)^{-1}X^*a = X\} = \{X \in \mathcal{A} \mid X^* = X\} = \mathcal{A}^s \oplus \{0\}$$

and we have the obvious splitting

$$\mathcal{A} = H_a \oplus V_a.$$

Not only that, the map  $(D\pi_a)_1$  is invertible from the left. That is, there exists a mapping (actually a section of the bundle  $(G, G^+, \pi_a)$ ),  $\kappa_a: (TG^+)_a \rightarrow (TG)_1$ , given by

$$\kappa_a(z) \equiv -\frac{a^{-1}}{2}z$$

such that  $(TG^+)_a \xrightarrow{\kappa_a} (TG)_1 \xrightarrow{(D\pi_a)_1} (TG^+)_a$  is the identity mapping.

**Definition 2.1.** Define the  $\mathcal{A}$ -valued 1-form  $\kappa: G^+ \rightarrow L(TG^+, \mathcal{A}^s)$  by  $\kappa_b = \kappa_a \circ (DL_g)^{-1}$ , where  $L_g(a) = b$ .

*Comment.* From now on we shall use the shorthand and  $\tilde{L}_g$  for the tangent mapping  $DL_g$  and  $\tilde{\pi}_a$  for the tangent  $D\pi_a$ , etc.

The mapping  $\kappa$  is called the *structure 1-form* of the homogeneous space  $G^+$ . All the geometry on  $G^+$  comes from  $\kappa$ . The basic properties of these mappings are contained in the simple

**Lemma 2.1.** *With the notations above, and if  $L_g(a) = b$ , we have (i)  $\pi_b = L_g \circ \pi_a$ , (ii) or in differential form:  $\tilde{\pi}_b = \tilde{L}_g \circ \tilde{\pi}_a$ , (iii)  $\kappa$  is equivariant, that is,  $\kappa_b \circ \tilde{L}_g = \kappa_a$ , and (iv)  $\tilde{\pi}_b \circ \kappa_b: (TG^+)_b \rightarrow (TG^+)_b$  is the identity mapping.*

## 2.2. Geodesics in $G^+$ .

Let us now recall the basic construction leading to the definition of geodesics in  $G^+$ . We begin with

**Definition 2.2.** Let  $a(t)$  be a continuously differentiable curve in  $G^+$  such that  $a(0) = a$ . Let  $\gamma(t)$  be the solution to the *transport equation*

$$(1) \quad \dot{\gamma}(t) = \kappa_{a(t)}(\dot{a}(t))\gamma(t), \quad \gamma(0) = 1.$$

The curve  $\gamma(t)$  is called the horizontal lift of  $a(t)$ .

**Proposition 2.1.** *With the notations introduced above we have:*

(i)  $\pi_a(\gamma(t)) = a(t)$  and (ii)  $\dot{\gamma}(t) \in \mathcal{A}^s = \kappa_a((TG^+)_a)$ .

*Proof:* Just note that integrating (1) we obtain  $\gamma(t) = \left(\frac{a(0)}{a(t)}\right)^{1/2}$ .  $\square$

**Definition 2.3.** Let  $a(t)$  be a continuously differentiable curve in  $G^+$  such that  $a(0) = a$ , and let  $Y(t)$  be a vector field defined along  $a(t)$ . The covariant derivative of  $Y(t)$  along  $a(t)$  is defined to be

$$(2) \quad D_{\dot{a}(t)}Y(t) = \pi_{a(t)} \left( \frac{d}{dt} \kappa_{a(t)}(Y(t)) \right).$$

Observe now that if  $a(t)$  is a twice differentiable curve in  $G^+$ , by considering  $Y(t) = \dot{a}(t)$  in (2) above, a simple computation yields:

$$(3) \quad D_{\dot{a}(t)}\dot{a}(t) = -\frac{\dot{a}^2(t)}{a(t)} + \ddot{a}(t)$$

and if, as usual, we say that  $a(t)$  is a *geodesic* whenever  $D_{\dot{a}(t)}\dot{a}(t) = 0$ , we readily obtain that

$$(4) \quad a(t) = a(0)e^{tX}, \quad \text{where } X = \ln \frac{\dot{a}(0)}{a(0)}.$$

Note as well that the geodesic which goes from  $a_0$  to  $a_1$ , both in  $G^+$ , in one time unit is rapidly obtained from (4) to be given by

$$(5) \quad a(t) = a_0^{1-t}a_1^t.$$

*Comments.* Clearly we also have  $X = \ln \left(\frac{a_1}{a_0}\right)$ . To justify the name of the curves (5) we have to verify that they minimize some distance function. To begin with, at  $a \in G^+$  define the following norm. For  $x \in (TG^+)_a$  set  $\|x\|_a = \|a^{-1}x\|$ , where  $\|\cdot\|$  is the norm in the Banach algebra. Let now  $a(t)$  for  $t \in [0, 1]$ , be a continuously differentiable curve. Define its length by

$$l(a) = \int_0^1 \|\dot{a}(t)\|_{a(t)} dt.$$

Define now for given  $a_0, a_1 \in G^+$

$$(6) \quad d(a_0, a_1) = \inf\{l(c) \mid c \text{ continuously differentiable curve} \\ \text{joining } a_0 \text{ to } a_1\}.$$

It so happens that the curve (5) is the minimizer of (6), and therefore  $\kappa$  is the connection in the Finsler geometry associated to  $d(a_0, a_1)$ . We have

**Theorem 2.1.** *Let  $a_0, a_1 \in G^+$  and denote by  $g_{a_0, a_1}(t)$  the curve (5) and let  $c(t)$  be any other continuously differentiable curve joining  $a_0$  to  $a_1$  in a unit of time. Then  $l(c) \geq l(g_{a_0, a_1})$ .*

*Proof:* Consider

$$l(c) = \int_0^1 \|\dot{c}(t)\|_{c(t)} dt = \int_0^1 \|c^{-1}(t)\dot{c}(t)\| dt \\ \geq \left\| \int_0^1 d \ln c(t) dt \right\| = \left\| \ln \left( \frac{a_1}{a_0} \right) \right\| = d(a_0, a_1).$$

To close this section note that going from  $a_0$  to  $a_1$  along a geodesic can be realized by means of the group action:

$$a_1 = a_0 e^X = L_g(a_0), \quad \text{with } g = e^{-X/2},$$

which prompts the following definition for vector fields along curves in  $G^+$ .

**Definition 2.4.** Let  $Y(t)$  be a vector field defined along a geodesic curve  $a(t) \in G^+$ . We say that  $Y$  is parallel if  $Y(t) = L_{g(t)}(Y(0))$  for every  $t$  where  $g(t)$  is a group element taking  $a_0$  onto  $a_1$ .

Let us examine the basics of a Riemannian geometry on  $G^+$ .

**2.3. Two Riemannian structures on  $G^+$ .**

In this section we examine two different Riemannian structures on  $G^+$ . With respect to the first one, the geodesic transport is our old group action, which turns out to be self dual. With respect to the other, there are two different parallel transports in duality. We also provide a geometric interpretation for Pinsker's inequality and a converse to it.

**Definition 2.5.** We define the scalar product on the tangent bundle  $TG^+$  as follows: On the tangent space  $(TG^+)_1 = \mathcal{A}^s$  to  $G^+$  at 1, we can define the scalar product  $\langle X, Y \rangle_1 \equiv E_m[XY]$  which can be made equivariant by setting  $\langle X, Y \rangle_a = E_m[a^{-1}Xa^{-1}Y]$ .

To begin with, we have the simple

**Lemma 2.2.** *The parallel transport in Definition 2.5, is self dual with respect to this scalar product, that is, if  $L_g(1) = a$ ,  $X_a = L_g(X_1)$  and  $Y_a = L_g(Y_1)$ ,*

$$(7) \quad \langle X_a, Y_a \rangle_a = \langle X_1, Y_1 \rangle_1.$$

*Comment.* Duality of different parallel transports is studied in [A], [ABKLR] and [S]. Here we provide a parallel transport leading to a connection that is self dual. We can define the squared distance along curves for this Riemannian metric by

**Definition 2.6.** Let  $c(t)$  be a continuously differentiable curve in  $G^+$  joining  $c(0)$  to  $c(1)$ . The (squared) distance along  $c$  is defined by

$$(8) \quad d_1(c(0), c(1)) = \int_0^1 \langle \dot{c}(t), \dot{c}(t) \rangle_{(c)(t)} dt.$$

What is interesting and remarkable at this stage is that

**Proposition 2.2.** *The geodesics of (8) are the geodesics of the connection in  $G^+$  and are given by (5).*

*Proof:* Left for the reader. Just notice that the Euler-Lagrange equation  $\frac{d}{dt} \left( \frac{\partial \mathcal{L}}{\partial \dot{c}} \right) = \left( \frac{\partial \mathcal{L}}{\partial c} \right)$  is easy to obtain in the commutative case. Here  $\mathcal{L}(c)$  is given by the right hand side of (8). The solution is as claimed.  $\square$

A simple computation shows that along a geodesic  $a(t) = a_0 e^{tX}$  with  $X = \ln \left( \frac{a_1}{a_0} \right)$ , the distance is as given in Theorem 2.1, i.e.,

$$(9) \quad d_1(a_0, a_1) = \|X\| = \left\| \ln \left( \frac{a_1}{a_0} \right) \right\|.$$

We shall now introduce a different Riemannian structure on  $G^+$ .

**Definition 2.7.** Let us now define the scalar product in  $(TG^+)_a$  by setting  $\langle X, Y \rangle_a \equiv E_m[a^{-1}XY]$ .

The analogue of Lemma 2.2 is the following

**Lemma 2.3.** *Let  $X_a = L_g(X_1) = T_{1,a}^+(X_1)$  where  $a = L_g(1)$  and let  $Y = T_{1,a}^-(Y)$ , that is, the parallel transport from the ambient space in which parallel transport is assumed to be the identity mapping. Then*

$$(T_{1,a}^+(X_1), T_{1,a}^-(Y))_a = E_m[a^{-1}aXY] = \langle X, Y \rangle_1$$

*that is, the geodesic transport and the identical parallel transport are dual with respect to the  $(\cdot, \cdot)$  product on  $TG^+$ .*

We also have

**Proposition 2.3.** *Let  $c(t)$  be a differentiable curve in  $G^+$  joining  $a_0$  to  $a_1$  in a unit of time. Define*

$$d_2(a_0, a_1) = \int_0^1 (\dot{c}(t), \dot{c}(t))_{c(t)} dt.$$

*Then the geodesic in that metric is  $a^*(t) = (a_0^{1/2} + t(a_1^{1/2} - a_0^{1/2}))^2$ .*

*Proof:* Notice to begin with the Euler-Lagrange equations determining the geodesic are now  $2\left(\frac{\ddot{c}}{\dot{c}}\right) = \left(\frac{\dot{c}}{c}\right)^2$ , which can be integrated to yield the claimed result. Next a simple application of Jensen's inequality yields that for any other continuously differentiable curve  $\hat{c}$  with the same initial and final points

$$(10) \quad \int_0^1 (\dot{a}^*(t), \dot{a}^*(t))_{a^*(t)} \equiv 4E_m[(a_1^{1/2} - a_0^{1/2})^2] \leq \int_0^1 (\dot{\hat{c}}(t), \dot{\hat{c}}(t))_{\hat{c}(t)},$$

thus concluding the proof. □

*Comment.* Actually (10) connects us to

**Corollary 2.1.** *Let now  $c(t) = a_0 e^{tX}$  with  $X = \ln\left(\frac{a_1}{a_0}\right)$ . Then (10) implies that*

$$(11) \quad E_m \left[ (a_1 - a_0) \ln \left( \frac{a_1}{a_0} \right) \right] \geq 4E_m [(a_1^{1/2} - a_0^{1/2})^2].$$

Now, when  $E_m[a_0] = E_m[a_1] = 1$ , we obtain  $4E_m[(a_1^{1/2} - a_0^{1/2})^2] \geq (E_m[(a_1 - a_0)])^2$  after a simple application of the Cauchy-Schwarz inequality, which turns (11) into a simple extension of the famous Pinsker's inequality. We thus obtain a totally geometric proof of that famous inequality.

To obtain a kind of converse to that inequality using our geometric setup, compute the distance  $d_1(a_0, a_1)$  along the geodesic  $a^*(t) = (a_0^{1/2} + t(a_1^{1/2} - a_0^{1/2}))^2$  for  $d_2$  to obtain

**Proposition 2.4.** *With the notations introduced above the following inequality holds*

$$(12) \quad \left\| \ln \left( \frac{a_1}{a_0} \right) \right\| \leq \| (a_1^{1/2} - a_0^{1/2}) \|.$$

#### 2.4. A projective class in $G^+$ .

In this section we consider a simple projective space,  $\mathbb{P}^+$ , arising from an equivalence relation on  $G^+$ . For this it is convenient to think of constants as constant functions, and the class of constant functions as a sub-algebra  $\mathcal{B}$  of  $\mathcal{A}$ . A projection (or conditional expectation when  $m$  is a probability) of  $\mathcal{A}$  onto  $\mathcal{B}$  is given by  $X \rightarrow E_m[X] \equiv \int_S X(s) dm(s)$ . We shall denote respectively by  $G_{\mathcal{B}}$  and  $G_{\mathcal{B}}^+$  the group of invertible elements and the group of positive invertible elements in  $\mathcal{B}$ .

**Definition 2.8.** Let  $a$  and  $b$  be in  $G^+$ . We shall say that  $a \sim_{\mathcal{B}} b$  if and only if there exists an element  $g \in G_{\mathcal{B}}$  such that  $b = L_g(a)$ . Form  $\mathbb{P}^+ = G^+ / \sim_{\mathcal{B}}$  and denote by  $\tau$  the quotient map  $\tau: G^+ \rightarrow \mathbb{P}^+$ .

*Comment.* This clearly amounts to saying that  $a \sim_{\mathcal{B}} b$  if there is a positive real  $r$  such that  $a = rb$ .

To define the tangent bundle to  $\mathbb{P}^+$ , we have to describe the tangent space at every  $\alpha = \tau(a) \in \mathbb{P}^+$ .

**Definition 2.9.** We shall say that  $(a, X) \sim_{\mathcal{B}} (a', X')$  if and only if  $a \sim_{\mathcal{B}} a'$  and  $X'/a' - X/a \in \mathcal{A}^s$ . Here  $X \in (TG^+)_a$  and  $X' \in (TG^+)_{a'}$ .

Clearly the group action of  $G$  on  $G^+$  can be induced on  $\mathbb{P}^+$  in the obvious way:

$$L_g(\alpha) = \tau(L_g(a)) \quad \text{if } \alpha = \tau(a), g \in G.$$

To construct geodesics on  $\mathbb{P}^+$ , an easy way out is to present it as a homogeneous space. For that, note that if we set  $\alpha_1 = \tau(1)$ , the isotropy group of  $\alpha_1$  is

$$I_{\alpha_1} = \{g \in G \mid L_g(\alpha_1) = \alpha_1\},$$

that is, the collection of all  $g$ 's such that  $\|g\|^{-2} \sim_{\mathcal{B}} 1$ , or  $G_{\mathcal{B}}$  in this case, and we have

**Proposition 2.5.** *With the notations introduced above, we have the isomorphism*

$$\mathbb{P}^+ \simeq G/I_{\alpha_1}.$$

*Comment.* Hereby we present  $\mathbb{P}^+$  as a homogenous space. Thus the whole machinery developed in [KN] may be brought into play to define connections and geodesics. For the time being we want to concentrate in a specific class of representatives for  $\mathbb{P}^+$  and leave the study of  $\mathbb{P}^+$  and some associate constructions for a continuation of [GR].

**3. Geometry in the class of probability densities**

The main result of this section consists of verifying that geodesics in  $G^+$  having initial and final point in the class of densities, stay there all the time. We shall examine two Riemannian geometries on  $\mathcal{D}$  that admit a pair of dual parallel transports. To establish the notation consider the basic class of densities with respect to  $m$ ,

$$\mathcal{D} = \{\rho \in G^+ \mid E_m[\rho] = 1\}.$$

To go from  $G^+$  to  $\mathcal{D}$  consider the projection

$$\Phi: G^+ \longrightarrow \mathcal{D} \quad \text{given by} \quad a \longrightarrow \Phi(a) = \frac{a}{E_m[a]}.$$

*Comments.* Notice that this mapping is constant along rays, i.e., it provides systems of representatives for  $\mathbb{P}^+$ . Note as well that if  $\rho = \Phi(a)$ , the tangent map induced by  $\Phi$  is given by

$$(13) \quad \tilde{\Phi}: (TG^+)_a \longrightarrow (T\mathcal{D})_\rho \quad X \longrightarrow \tilde{\Phi}(X) = \rho X - \rho E_m[\rho X] = Q_\rho(\rho X),$$

where  $Q_\rho(Y) \equiv Y - \rho E_m(Y)$  is a projection operator. Note that it acts on vectors tangent to the tangent space  $(TG^+)_\rho$ .

With the help of (13) we can obtain a connection on  $\mathcal{D}$  as follows.

**Definition 3.1.** With notations introduced above, if  $\rho(t)$  is a curve in  $\mathcal{D}$  with tangent  $X$  and  $Y$  is tangent to  $\mathcal{D}$  along  $\rho$ , then the covariant derivative of  $Y$  along  $X$  is defined to be

$$(14) \quad \tilde{D}_X Y = Q_\rho(D_X Y) = D_X Y - \rho E_m[D_X Y].$$

And we also have

**Theorem 3.1.** Let  $a(t) = a(0)e^{tX}$  be a geodesic in  $G^+$  passing through  $a(0)$  at time  $t = 0$  with speed  $X$ . Then  $\rho(t) = \Phi(a(t))$  is a geodesic in  $\mathcal{D}$ , i.e., it satisfies  $\tilde{D}_\rho \dot{\rho} = 0$ .

*Proof:* That  $\rho(t)$  is fully contained in  $\mathcal{D}$  is clear. We have to verify that  $\rho(t)$  satisfies  $\tilde{D}_\rho \dot{\rho} = 0$ . All it takes is to compute two derivatives and then to verify that  $D_\rho \dot{\rho} = \ddot{\rho} - \dot{\rho}^2/\rho = -\rho V(X)$ , where  $V(X) = E_m[\rho X^2] - (E_m[\rho X])^2$ . Apply (14) and the result easily drops out.  $\square$

*Comment.* A curious geometric connection appears here between variances and covariant derivatives.

**Corollary 3.1.** Let  $\rho(0)$  and  $\rho(1)$  be two densities. Then

$\rho(t) = \frac{\rho(0)^{(1-t)}\rho(1)^t}{E_m[\rho(0)^{(1-t)}\rho(1)^t]}$  is a geodesic in  $\mathcal{D}$ , going from  $\rho(0)$  to  $\rho(1)$  in a unit of time.

*Proof:* We know that given  $\rho_0$  and  $\rho_1$  there is a geodesic in  $G^+$ , written as  $a(t) = \rho_0 e^{tY}$ , joining the two points. According to the theorem,  $\Phi(a(t)) = \rho_0 e^{tY} / E_M[\rho_0 e^{tY}]$  is a geodesic in  $\mathcal{D}$  joining the same two points. Now set  $\rho_1 = \Phi(a(1))$ , solve for  $e^Y$  and substitute in  $\rho(t) \equiv \Phi(a(t))$  to obtain the desired result.

On  $\mathcal{D}$  parallel motion along geodesics is realized by projecting the geodesic motion in  $G^+$ . That is,  $\rho(t) = \frac{\rho(0)^{(1-t)} \rho(1)^t}{E_m[\rho(0)^{(1-t)} \rho(1)^t]} = \Phi \circ L_{g(t)}(\rho(0))$  where  $g(t) = e^{-tX/2}$  and  $X = \ln\left(\frac{\rho(1)}{\rho(0)}\right)$ . Actually there is a little more to this

**Proposition 3.1.** *The projected group action on  $\mathcal{D}$  is again a group action, i.e., the mapping  $\hat{L}_g(\rho) = \Phi \circ L_g(\rho)$ , defines a group action on  $\mathcal{D}$ .*

*Proof:* Just notice that

$$\begin{aligned} \Phi(L_{g_1} \Phi(L_{g_2}(\rho))) &= \frac{|g_1|^{-2} |g_2|^{-2} \rho / E_m[|g_2|^{-2} \rho]}{E_m[|g_1|^{-2} |g_2|^{-2} \rho / E_m[|g_2|^{-2} \rho]]} \\ &= \frac{|g_1|^{-2} |g_2|^{-2} \rho}{E_m[|g_1|^{-2} |g_2|^{-2} \rho]} = \Phi(L_{g_1 g_2}(\rho)). \quad \square \end{aligned}$$

The analogue of Definition 2.4 is in this case

**Definition 3.2.** A tangent vector field along a curve  $\rho(t)$  in  $\mathcal{D}$  is parallel if

$$Y(t) = \tilde{\Phi}(L_{g(t)} Y(0)), \quad \text{where } \rho(t) = \hat{L}_{g(t)}(\rho(0)).$$

Let us denote the parallel transport induced by the above restricted group action by  $\tilde{T}_{g_1 g_2}^+$ , i.e., if  $\hat{L}_g(\rho_1) = \rho_2$ , then for  $X_{\rho_1} \in (TG^+)_{\rho_1}$  we set

$$\tilde{T}_{\rho_1 \rho_2}^+(X_{\rho_1}) = \tilde{\Phi}(L_g(X_{\rho_1})) = \left(\frac{\rho_2}{\rho_1}\right) X_{\rho_1} - \rho_2 E_m \left[ \left(\frac{\rho_2}{\rho_1}\right) X_{\rho_1} \right].$$

**Lemma 3.1.** *With the notations introduced above*

$$\tilde{T}_{\rho_2 \rho_3}^+ \circ \tilde{T}_{\rho_1 \rho_2}^+ = \tilde{T}_{\rho_1 \rho_3}^+$$

for any  $\rho_1, \rho_2$  and  $\rho_3$ .

To define a Riemannian structure on  $\mathcal{D}$  it is convenient to refer all vectors to  $T\mathcal{D}_1$ . (By the way, it is at this point where the assumption that the measure  $m$  is a probability measure simplifies the computations.) Since the group action  $\hat{L}$  on  $\mathcal{D}$  is transitive, any vector  $X_\rho \in T\mathcal{D}_\rho$  can be

obtained from a vector  $X_0$  at  $T\mathcal{D}_1$  by  $X_\rho = \tilde{T}_{1\rho}(X_0)^+ = \rho X_0 - \rho E_m[\rho X_0]$ , and we can set

**Definition 3.3.** We can restrict the  $(\bullet, \bullet)_g$  scalar product to  $\mathcal{D}$  by

$$(\tilde{T}_{1\rho}^+(X_0), \tilde{T}_{1\rho}^+(Y_0))_g = C_\rho(X_0, Y_0) = E_m[\rho X_0 Y_0] - E_m[\rho X_0] E_m[\rho Y_0].$$

And as above, we have the same duality result

**Lemma 3.2.** *With the notations introduced above, and again denoting by  $\tilde{T}^- : (TG^+)_1 \rightarrow (TG^+)_\rho$  by  $\tilde{T}^-(X_0) = X_0$ , then*

$$(\tilde{T}_{1\rho}^+(X_0), \tilde{T}^-(X_0))_g = (X_0, Y_0)_1.$$

#### 4. Geodesic surfaces and exponential families

The thrust of this section is to establish that natural exponential families of densities appear as geodesic surfaces, but keep in mind that geodesic surface here means the following

**Definition 4.1.** A  $k$ -differentiable surface of order  $d$  in  $\mathcal{D}$  is a  $k$ -differentiable mapping  $\Psi : \Theta \rightarrow \mathcal{D}$ , where  $\Theta$  is an open convex set in  $\mathbb{R}^d$ .

*Comment.* As usual it is convenient to think of the surface as the range of  $\Psi$ , i.e., as a family  $\Pi = \{\Psi(\theta) \mid \theta \in \Theta\}$ .

**Definition 4.2.** We shall say that a surface is geodesic if any two of its points can be joined by a geodesic contained within the surface.

**Definition 4.3.** We shall say that a family  $\{\rho(\theta)\}$  is a natural exponential family labeled by  $\Theta$  if there exist  $X_1, \dots, X_d$  in  $\mathcal{A}^s$  and  $q$  in  $G^+$  such that

$$\rho(\theta) = \Phi(qe^{\langle \theta, X \rangle}).$$

**Theorem 4.1.** *Let  $\Pi$  be a surface of order  $d$  in  $\mathcal{D}$  such that the vector fields  $X_i \equiv \frac{\partial \ln \Psi(\theta)}{\partial \theta_i}$  for  $i = 1, \dots, d$ , are linearly independent. Then the family  $\Pi$  is exponential if and only if the image of any segment  $\theta(s) = (1-s)\theta_0 + s\theta_1$  in  $\Theta$  is a geodesic joining  $\Psi(\theta_0)$  to  $\Psi(\theta_1)$  in  $\mathcal{D}$ .*

*Proof:* Assume that  $\Psi(\theta) = qe^{\langle \theta, X \rangle}$ , and let  $\rho(s) = \Phi(\Psi(\theta_0)^{1-s} \Psi(\theta_1)^s)$  be the geodesic joining  $\Psi(\theta_0)$  to  $\Psi(\theta_1)$ . Then clearly  $\rho(s) = \Psi((1-s)\theta_0 + s\theta_1)$ .

Assume next that  $\Phi(\Psi(\theta_0)^{1-s} \Psi(\theta_1)^s) = \Psi((1-s)\theta_0 + s\theta_1)$ , that is that geodesics in  $\mathcal{D}$  are the image of lines in  $\Theta$ . Fix  $\theta_0$  and rewrite the last identity as  $\Psi(\theta_0 + s(\theta_1 - \theta_0)) = \Phi(\Psi(\theta_0)e^{sY})$  as above. Now differentiate both sides at  $s = 0$  to obtain

$$\langle \Delta\theta, \nabla\Psi(\theta_0) \rangle = \Psi(\theta_0)Y - \Psi(\theta_0)E_m[\Psi(\theta_0)].$$

Now set  $s = 1$ , bring in the definition of  $X$  and change  $\theta_1$  to  $\theta$  above to rewrite this identity as  $\langle \Delta\theta, X \rangle = \Psi(\theta_0)Y - \Psi(\theta_0)E_m[\Psi(\theta_0)]$  and obtain

$$\Psi(\theta) = \Phi(\Psi(\theta_0)e^Y) = \Phi(\Psi(\theta_0)e^{Y-E_m[\Psi(\theta_0)]}) = \Phi(\Psi(\theta_0)e^{\langle \Delta\theta, X \rangle})$$

where the second step is valid for  $\Phi$  is constant along rays in  $G^+$ . Now set  $q = \Psi(\theta_0)e^{-\langle \theta_0, X \rangle}$  and we are through with the proof.  $\square$

## 5. Geometry on $\ell_1(a)$ , geodesic families and polynomial sequences

Even though the scheme developed above provides a nice geometrical framework in which to understand exponential families, it has an unsavory or at least inconvenient limitation, namely, that it only allows for densities that are bounded away from zero. The algebraic setup developed above can be modified to overcome that inconvenience, and the purpose of this section is to show how this can be done in the context of a specific example, namely when the algebra is  $\ell_1(a)$ . Some unexpected connections appear.

### 5.1. Some properties on convolution sequences.

Let us begin by recalling some basic properties of convolution sequences. Let  $\{\alpha_n \mid n \geq 0\}$  be a sequence of positive numbers such that  $\alpha_{n+m} \leq \alpha_n \alpha_m$  for all  $n, m \geq 0$  and let us put

$$\mathcal{A} = \ell_1(\alpha) = \left\{ \{x_n \in \mathbb{C} : 0 \leq n < \infty\} \mid \sum_0^\infty \alpha_n |x_n| < \infty \right\}.$$

We shall consider  $\|x\|_\alpha = \sum_0^\infty \alpha_n |x_n|$  to be the norm on  $\mathcal{A}$ . If we regard  $\alpha$  as a measure on  $\mathbb{N}$ , then  $E_\alpha[x] = \sum_n x_n \alpha_n$  may be thought of as an integral, and talking about densities with respect to  $\alpha$  makes sense. To make this measure finite it suffices to assume that  $\alpha_1 < 1$ . Addition and scalar multiplication are component wise, whereas the product  $x * y$  of two elements  $x$  and  $y$  in  $\mathcal{A}$  is the convolution product. As mentioned, this turns  $\mathcal{A}$  into a Banach algebra with involution (given by the complex conjugation) which we shall denote by  $x \rightarrow \bar{x}$  since we are using  $*$  for the product. Also, the unit element in  $\mathcal{A}$  will be denoted by  $1$  and is the sequence  $\delta_0 = \{1, 0, \dots\}$ . All the basic properties about commutative Banach algebras the we need here can be seen in [GRS] and a more modern background on convolution spaces can be seen in [BCR].

We set  $\rho = \lim_{n \rightarrow \infty} \frac{1}{n} \log \alpha_n$  and assume that  $\rho > -\infty$ . It is also known that the complex homomorphisms of  $\mathcal{A}$  (i.e., continuous, linear maps from  $\mathcal{A}$  to  $\mathbb{C}$  that preserve multiplication) are of the form  $\Lambda_z(x) = \sum_0^\infty x_n z_n$ , with  $|z| \leq e^\rho$ . A basic property of the convolution operation is contained in

**Lemma 5.1.** *Let  $\{x_n\}$  be a sequence of complex numbers such that  $x_0 = 0$ . Then*

- a)  $x_n^{k*} = 0$  whenever  $k > n$ .
- b)  $x_n^{n*} = (x_1)^n$  for all  $n > 0$ .
- c)  $x_n^{k*}$  is a polynomial in  $x_1, \dots, x_k$  for all  $2 \leq k \leq n$  and  $k, n \geq 0$ .

Here  $x_n^{k*} = (x * \dots * x)_n$ .

The proof is simple and detailed in [dB]. An interesting consequence of this, also proved there, is

**Lemma 5.2.** *Let  $g_0, g_1, \dots, g_N$  be complex numbers and let  $h_0(t), \dots, h_N(t)$  be complex valued functions defined on  $\mathbb{R}$  or  $[0, \infty)$ , such that*

$$h_n(t + s) = \sum_0^n h_k(t) h_{n-k}(s) \quad \text{for } n = 0, \dots, N.$$

Define

$$f_n(t) = \sum_0^n g_n^{k*} h_k(t) \quad \text{for } n = 0, \dots, N,$$

then

$$(15) \quad f_n(t + s) = \sum_0^n f_k(t) f_{n-k}(s) \quad \text{for } n = 0, \dots, N.$$

**Definition 5.1.** A sequence of functions  $\{f_n(t) : n \geq 0\}$ , defined on  $\mathbb{R}$  or  $[0, \infty)$ , such that (15) holds, is said to be of convolution type.

The following result contain an analogue to the first lemma

**Lemma 5.3.** *Let  $\{f_n(t) : n \geq 0\}$  be a sequence of convolution type. Then*

- a) if  $f_0 = 0$ , then  $f_n(t) = 0$  for all  $n > 0$ ,
- b) if  $f_0(t_0) = 0$  for some  $t_0$ , then  $f_n = 0$  for all  $n \geq 0$ .

The following is some sort of converse to (15). It is also proved in [dB].

**Theorem 5.1.** Consider a convolution algebra  $\mathcal{A}$  of functions such that

- a) the only non zero elements  $f$  in the algebra that solve  $f(t + s) = f(t) + f(t)$  are  $f(t) = ct$ , with  $c \in \mathbb{C}$ ,
- b) the only non zero elements  $f$  in the algebra that solve  $f(t + s) = f(t)f(t)$  are  $f(t) = e^{at}$  with  $a \in \mathbb{C}$ .

Then  $\{f_n(t) : n \geq 0\}$  is a sequence of convolution type if and only if there exists  $a \in \mathbb{C}$  and a sequence  $\{g_n \in \mathcal{A}\}$ , with  $g_0 \neq 0$  such that

$$(16) \quad f_n(t) = e^{at} \sum_0^n g_n^{k*} \left( \frac{t^k}{k!} \right).$$

In the geometric setup an important role is played by (the connected component of) the group  $G$  of all invertible (with respect to the convolution product) elements of  $\mathcal{A} = \ell_1(\alpha)$ . Let us introduce the notation  $\mathcal{E} = \{e^x \mid x \in \mathcal{A}\}$ , where  $e^x \equiv \sum_k x^{k*}$ , then

**Theorem 5.2.** With the notations introduced above,  $G = \mathcal{E}$ .

The proof of this result and the next are detailed in [dB]. In terms of the generic automorphisms introduced above, we also have

**Theorem 5.3.** With the notations introduced above

$$G = \mathcal{E} = \left\{ x \in \mathcal{A} \mid \Lambda_z(x) = \sum_n x_n z^n \neq 0, \forall |z| \leq e^\rho \right\}.$$

We mention in passing that for  $|z| \leq e^\rho$ , the sets  $\mathcal{M}_z = \{x \in \mathcal{A} \mid \Lambda_z(x) = 0\}$  are maximal ideals in  $\mathcal{A}$ , and we have the mapping

$$\Lambda : \mathcal{A} \longrightarrow A(e^\rho) \quad x \longrightarrow \Lambda_\bullet(x)$$

where  $A(r) = \{g : \{|z| \leq r\} \rightarrow \mathbb{C} \mid g \text{ analytic for } |z| < r \text{ and continuous for } |z| \leq r\}$ .

**5.2. Geometry on  $G^+$ .**

The constructions developed in Section 2 above have to be transported to this example with care. This is due to the fact that in this case positivity is now relative to the convolution product, that is,  $a \in \ell(a)$  is positive if there exists  $b \in \ell(\alpha)$  such that  $a = b * \bar{b}$ . And it is important to keep in mind that positivity does not imply pointwise positivity. Consider  $b = (1, -2, 0, 0, \dots)$ , then  $b * b = (1, -4, 4, 0, \dots)$  which is positive but not pointwise positive.

As above, an action is defined for  $g \in G$  and  $a \in G^+$  by

$$L_g(a) = \bar{g}^{-1} * a * g^{-1} = (\bar{g} * g)^{-1} * a,$$

and it is clear that the group  $G$  acts on  $G^+$  transitively. The isotropy group of  $a \in G^+$  is defined by  $I_a = \{g \in G \mid L_g(a) = a\}$ , and clearly is independent of  $a$  and we have

**Lemma 5.4.** *With the notations above we have  $I_1 = \exp(\ell_1^a(\alpha))$ , where*

$$\ell_1^a(\alpha) = \{X \in \ell_1(\alpha) \mid \bar{X} + X = 0\}$$

*the antisymmetric elements in  $\ell_1(\alpha)$ .*

*Comment.* Since  $G$  is open in  $\mathcal{A}$  its tangent space at 1 is  $\mathcal{A}$ , and when we want to think of the elements of  $\mathcal{A}$  as tangent vectors, we shall denote them by capitals. Also, for  $a_1, a_2 \in G^+$ , clearly  $g = (a_1^{-1} * a_2)^{1/2} * u$  with  $u \in I_1$  satisfies  $L_g(a_1) = a_2$ .

*Proof:* Commutativity implies that  $I_1 = \{g \in G \mid \bar{g} * g = 1\}$ . Since  $G = \mathcal{E}$  we know that  $g = e^{X*}$  for some  $X$ . Since  $e^{X*} * e^{\bar{X}*} = e^0$  we are through.  $\square$

**Lemma 5.5.** *With the notations introduced above, the following identities hold.  $G^+ = G/I$ ,  $(TG)_1 = \ell_1(\alpha)$ ,  $\mathcal{I} \equiv (TI)_1 = \ell_1^a(\alpha)$  and for any  $a \in G^+$ ,  $(G^+)_a = \ell_1^s(\alpha) = \{X \in \mathcal{A} \mid X = \bar{X}\}$ .*

*Proof:* Let us verify the third assertion. Let  $u(t)$  be a smooth curve in  $I_1$  such that  $u(0) = 1$  and  $\dot{u}(0) = X$ . Differentiate both sides of  $\bar{u}(t) * u(t) = 1$  at  $t = 0$ .  $\square$

**5.2.1. Connection and geodesics on  $G^+$ .**

Again we direct the reader to Section 2 above or to [KN], [CPR], or [GR] for full details. Once we know that  $G^+$  is a homogeneous bundle, the procedure to define a geometry on it is standard. For  $a \in G^+$ , the bundle map  $\pi_a: G \rightarrow G^+$  is again  $\pi_a(g) = L_g(a)$ .

Notice now that  $\mathcal{A} = (TG)_1 = \ell_1(\alpha) = \ell_1^s(\alpha) + \ell_1^a(\alpha) = (TG^+)_a \oplus \mathcal{I}$ . This splitting can be transported everywhere by means of the group action. Here  $(TG^+)_a$  will play the role of the horizontal space and  $\mathcal{I}$  will be the vertical space at every point. The 1-form connection  $\kappa_a: (TG^+)_a \rightarrow (TG)_a$  is now  $\kappa_a(X) = -\frac{1}{2}a^{-1} * X$ . Is easy to verify that if  $D\pi_a$  denotes de derivative (tangent of push forward map) of  $\pi_a$ . Then,  $D\pi_a \circ \kappa_a: (TG^+)_a \rightarrow (TG)_a$  is the identity mapping.

Again, the lifting of a continuously differentiable curve  $a(t)$  in  $G^+$  with  $a(0) = a$  is the curve  $g(t) = (a(0) * a(t)^{-1})^{1/2}$  in  $G$ . Once we have this lifting, we can use the group action to define parallel transport and geodesics as in Section 2. Adapting what we did to this case we have

**Proposition 5.1.** *The geodesics in  $G^+$  are the curves  $a(t) = a(0) * e^{tX*}$  with  $X = \dot{a}(0) * a(0)^{-1}$ .*

*Proof:* It is easy to verify that the equation for the geodesics  $\ddot{a}(t) - a(t)^{-1} * \dot{a}^2(t) = 0$ , and then it is simpler to verify that  $a(0) * e^{tX^*}$  satisfies the equation.  $\square$

*Comments.* For the sake of emphasis,  $e^{tX^*}$  is the element in  $\mathcal{A}$  with components

$$(e^{tX^*})_n = \sum_{k=0}^n t^k X_n^{k*} / k!$$

As in Section 2, to justify the calling  $a(0) * e^{tX^*}$  a geodesic, we should verify that it minimizes some distance function. To begin with, at  $a \in G^+$  define the following norm. For  $X \in (TG^+)_a$  set  $\|X\|_{\alpha, a} = \|a^{-1}X\|_{\alpha}$ , where  $\|\cdot\|_{\alpha}$  is the norm in the Banach algebra. Let now  $a(t)$  for  $t \in [0, 1]$ , be a continuously differentiable curve. Define its length by

$$l(a) = \int_0^1 \|\dot{a}(t)\|_{\alpha, a(t)} dt.$$

Define now for given  $a_0, a_1 \in G^+$

$$(17) \quad d(a_0, a_1) = \inf\{l(c) \mid c \text{ continuously differentiable curve joining } a_0 \text{ to } a_1\}.$$

It so happens that the curve  $a(0) * e^{tX^*}$  with  $X = \dot{a}(0)a^{-1}(0) = \ln(a(1) * a(0)^{-1})$  is the minimizer of (17), and therefore  $\kappa$  is the connection in the Finsler geometry associated to  $d(a_0, a_1)$ . We have

**Theorem 5.4.** *Let  $a_0, a_1 \in G^+$  and denote by  $g_{a_0, a_1}(t)$  the curve  $a(0) * e^{tX^*}$  and let  $c(t)$  be any other continuously differentiable curve joining  $a_0$  to  $a_1$  in a unit of time. Then  $l(c) \geq l(g_{a_0, a_1})$ .*

*Proof:* As in Section 2, consider

$$\begin{aligned} l(c) &= \int_0^1 \|\dot{c}(t)\|_{\alpha, c(t)} dt = \int_0^1 \|c^{-1}(t) * \dot{c}(t)\| dt \\ &\geq \left\| \int_0^1 d \ln c(t) dt \right\|_{\alpha} = \left\| \ln \left( \frac{a_1}{a_0} \right) \right\|_{\alpha} = d(a_0, a_1). \end{aligned}$$

To close this section note that going from  $a_0$  to  $a_1$  along a geodesic can be realized by means of the group action:

$$a_1 = a_0 * e^{X^*} = L_g(a_0), \quad \text{with } g = e^{-X/2^*},$$

and a vector field  $Y$  defined along a geodesic curve  $a(t) \in G^+$  is parallel if  $Y(t) = L_{g(t)}(Y(0))$  for every  $t$  where  $g(t)$  is a group element taking  $a_0$  onto  $a_1$ .

**5.3. Sequences of convolution type and geodesics in  $G^+$ .**

Let us begin by recalling two results from [dB].

**Theorem 5.5.** *Let  $q(t) \equiv \{q_n(t) : n \geq 0\}$  be a sequence of polynomials of convolution type, with coefficient sequence  $\{g_n : n \geq 0\}$  (i.e.  $q_n(t) = \sum_{k=0}^n g_n^{k*} t^k / k!$ ), and let  $\{\alpha_n\}$  be as above. Then the following are equivalent:*

- a)  $\{g_n\} \in \ell_1(\alpha)$ ,
- b) there exists  $M > 0$  such that  $\|q(t)\|_{1,\alpha} \leq e^{|t|^M}$  for all  $t \in \mathbb{C}$ ,
- c)  $\lim_{t \downarrow 0} \|q(t)\|_{1,\alpha} = 1$ ,
- d)  $\limsup_{t \downarrow 0} \|q(t)\|_{1,\alpha} < 2$ ,
- e) there exist  $\delta > 0$  and  $t_0 \in (0, \delta)$  such that  $q(t) \in \ell_1(\alpha)$  for all  $t \in (0, \delta)$  and  $\Psi(t_0, z) \equiv \Lambda_z(q(t_0)) \neq 0$  if  $|z| = e^\rho$ ,
- f) there exists  $t_0 \in \mathbb{C}$  such that  $q(t_0) \in \ell_1(\alpha)$  and  $\Lambda_z(q(t_0)) \neq 0$  if  $|z| \leq e^\rho$ ,
- g) there exists  $t_0 \in \mathbb{C}$  such that  $q(t_0) \in \ell_1(\alpha)$  and  $q(-t_0) \in \ell_1(\alpha)$ .

Moreover, if one of these conditions holds, then  $\sum_n q_n(t) z^n / n! = e^{tg(z)}$  where  $g(z) = \sum_n g_n z^n$  and both series converge in  $A(e^\rho)$  and all  $t \in \mathbb{C}$ .

From parts (a) and (b) we can safely begin manipulating our polynomial sequences without worrying much about convergence issues. And of interest here is other result from [dB].

**Theorem 5.6.** *Let  $q(t) \equiv \{q_n(t) : n \geq 0\}$  be a sequence of polynomials of convolution type with coefficient sequence  $\{g_n \mid n \geq 0\}$ . Then*

$$\Lambda_z(q(t)) = \sum_n q_n(t) z^n = e^{tg(z)}$$

where  $g(z) = \sum_n g_n z^n$ . Also  $q_0 = 1$  as well as  $q_n(0) = 0$  for all  $n \geq 1$ .

Let us now go backwards. Consider a geodesic  $a(t) = e^{tX^*}$  in  $G^+$ , (see Proposition 5.1) with  $X \in \ell_1(\alpha)$  that joins  $a(0) = 1$  to  $a(1) = e^X$ . We have

**Lemma 5.6.** *With the notations above, the family  $a(t) = \{a_n(t) \mid n \geq 0\}$  is of convolution type with coefficient sequence  $\{X_n \mid n \geq 0\}$ .*

*Proof:* The claim follows from identifying the  $n$ -th coordinate of  $a(t + s) = a(t) * a(s)$ . □

**Lemma 5.7.** *With the notations introduced above, we also have*

$$\Lambda_z(e^{tX^*}) = e^{t\Lambda_z(X)}.$$

*Proof:* Just a simple computation.  $\square$

As in Section 4, we can define a  $k$ -dimensional geodesic surface in  $G^+$  by

**Definition 5.2.** A geodesic surface generated by  $\mathbf{X} = (X_1, \dots, X_k)$  is a mapping  $\Theta: \mathbb{R}^k \rightarrow \mathcal{A}$  given by  $\mathbf{t} = (t_1, \dots, t_x) \rightarrow e^{\langle \mathbf{t}, \mathbf{X} \rangle}$ , where  $\langle \mathbf{t}, \mathbf{X} \rangle = \sum_i t_i X_i$ .

*Comment.* The following lemma is a variation on the theme of the previous lemma, describes how cross-sequences are related to geodesic surfaces in  $G^+$ .

**Lemma 5.8.** Let  $\{a_n(t_1, t_2) \mid n \geq 0\}$  be the sequence of polynomials determined by the geodesic surface, then for any  $(t_1, t_2)$  and  $(s_1, s_2)$

$$a_n(t_1 + s_1, t_2 + s_2) = \sum_{k=0}^n a_k(t_1, s_1) a_{n-k}(t_2, s_2).$$

*Proof:* For the reader.  $\square$

Let us now consider a special class of sequences:

**Definition 5.3.** We shall say that a differentiable curve in  $w: \mathbb{R} \rightarrow \ell_1(\alpha)$  is a Scheffer sequence with generating sequence  $g \in \ell_1(\alpha)$ , whenever its Gelfand transform  $\Lambda_z(w(t))$  satisfies

$$(18) \quad \frac{d}{dt} \Lambda_z(w(t)) = \Lambda_z(g) \Lambda_z(w(T)).$$

**Proposition 5.2.** Let  $q \in \ell_1(\alpha)$  be the convolution sequence generated by  $g \in \ell_1(\alpha)$ . If  $w(t)$  is the convolution sequence satisfying (18), then

$$\text{a) } w_n(t + s) = \sum_{k=0}^n w_k(t) q_{n-k}(s),$$

and in particular

$$\text{b) } w_n(t) = \sum_{k=0}^n w_k(0) q_{n-k}(t).$$

The proof is for the reader. The geometric way to understand Scheffer sequences as geodesics is contained in

**Lemma 5.9.** Let  $w_0 \in G^+$  and  $X \in \ell_1^s(\alpha)$  be a tangent vector to  $G^+$ , then the geodesic  $w(t) \equiv w_0 * e^{tX}$  through  $w_0$  is a Scheffer sequence.

*Proof:* Just observe that  $\Lambda_z(w(t))$  satisfies (18).  $\square$

**5.4. Geodesics in  $\mathbb{D}$  and exponential families.**

We shall now consider the class of “densities” with respect to  $\alpha$  defined by

$$\mathbb{D} = \left\{ q \in G^+ \mid E_\alpha(q) \equiv \sum_{n=0}^{\infty} q_n \alpha_n = 1 \right\}.$$

As with  $G^+$ , here we differ radically from what we did above in Sections 2–4, since now positivity does not mean sequential positivity, and  $\mathbb{D}$  is richer than a collection of densities on  $\mathbb{N}$ . Nevertheless, it still can be thought of as a class of representatives for  $\mathbb{P}^+$ , the projective space arising from the equivalence relation on  $G^+$  defined by  $a_1 \sim a_2 \Leftrightarrow \exists r \in (0, \infty)$  such that  $a_1 = ra_2$ . One can regard  $\mathbb{P}^+$  as a homogeneous reductive structure, and then try to pass on this structure on to  $\mathbb{D}$ . But this is not necessarily possible.

Instead, one can simply project the connection on  $G^+$  directly on to  $\mathbb{D}$ . This is carried out in detail in Sections 3 and 4, where it is shown that curves

$$\rho(t) = \frac{\rho_0^{1-t} \rho_1^t}{E_m[\rho_0^{1-t} \rho_1^t]}$$

are geodesics. Actually, what changes from there to here is the nature of the product in the algebra, but in an abstract setting, things are similar. Let us briefly sketch the results. To go from  $G^+$  to  $\mathbb{D}$  we consider the projection

$$\Phi: G^+ \longrightarrow \mathbb{D}; \quad a \longrightarrow \Phi(a) = \frac{a}{E_\alpha(q)}.$$

*Comment.* Notice that this mapping is constant along rays  $\{ra \mid r > 0\}$ , thus it makes sense to think of  $\mathbb{D}$  as representatives for  $\mathbb{P}^+$ . Notice as well that the tangent mapping  $\tilde{\Phi} \equiv D\Phi: (TG^+)_a \rightarrow (T\mathbb{D})_\rho$ , with  $\rho = \Phi(a)$  is given by

$$(19) \quad X \longrightarrow \tilde{\Phi}(X) = \rho X - \rho E_\alpha[\rho X].$$

**Definition 5.4.** With the notations introduced above, if  $\rho(t)$  is a curve in  $\mathbb{D}$  with tangent  $X$  and  $Y(t)$  is a vector field along  $\rho(t)$ , tangent to  $\mathbb{D}$ , its covariant derivative along  $X$  is defined by

$$\tilde{D}_X(Y) = \tilde{\Phi}(D_X(Y)) = \rho D_X(Y) - \rho E_\alpha[D_X(Y)]$$

where  $D_X(Y) = \frac{d}{dt}(L_{g(t)^{-1}}(Y(t)))_{t=0}$ .

The following were proved in Section 3. The difference being that we cannot talk of true densities anymore.

**Theorem 5.7.** Let  $q(t) = q(0) * e^{tX}$  be a geodesic in  $G^+$  passing through  $q(0)$  at  $t = 0$  with speed  $X$ . Then  $\Phi(q(t))$  is a geodesic in  $\mathbb{D}$  (i.e., it satisfies  $\tilde{D}_{\dot{\rho}(t)}(\dot{\rho}(t)) = 0$ ).

**Corollary 5.1.** Let  $\rho_0$  and  $\rho_1$  be in  $\mathbb{D}$ , then

$$\rho(t) = \frac{\rho_0^{t*} * \rho_1^{(1-t)*}}{E_\alpha[\rho_0^{t*} * \rho_1^{(1-t)*}]}$$

is a geodesic going from  $\rho_0$  to  $\rho_1$  in a unit of time.

**Theorem 5.8.** Let  $\Psi: \mathbf{t} \in \mathbb{R}^k \rightarrow \rho(\mathbf{t}) \in \mathbb{D}$  be a continuously differentiable mapping such that  $X_i = \frac{\partial \Psi(\mathbf{t})}{\partial t_i}$  are linearly independent in  $\ell_1(\alpha)$ . The family  $\Pi \equiv \{\rho(\mathbf{t}) \mid \mathbf{t} \in \mathbb{R}^k\}$  is exponential if and only if, for any pair  $\mathbf{t}_1$  and  $\mathbf{t}_2$ , the curve  $s \rightarrow \Psi(s\mathbf{t}_1 + (1-s)\mathbf{t}_2)$  is a geodesic joining  $\rho(\mathbf{t}_1)$  to  $\rho(\mathbf{t}_2)$  in  $\mathbb{D}$ . In this case  $\Psi(\mathbf{t}) = \Phi(e^{(\mathbf{t}, \mathbf{X})})$ .

*Comment.* Notice that as stated, the result may not be applicable to probability densities. For that to be happen, we would have to add that  $\rho(\mathbf{t})$  is pointwise positive. How do the pointwise sequences sit in  $G^+$  is an open question.

**Acknowledgements.** We want to thank the referee for her/his comments, which contributed to clarify the presentation.

### References

- [A] S.-I. AMARI, “*Differential-geometrical methods in statistics*”, Lecture Notes in Statistics **28**, Springer-Verlag, New York, 1985.
- [ABKLR] S.-I. AMARI, O. E. BARNDORFF-NIELSEN, R. E. KASS, S. L. LAURITZEN AND C. R. RAO, “*Differential geometry in statistical inference*”, Institute of Mathematical Statistics Lecture Notes-Monograph Series **10**, Institute of Mathematical Statistics, Hayward, CA, 1987.
- [B] O. E. BARNDORFF-NIELSEN, “*Information and exponential families in statistical theory*”, Wiley Series in Probability and Mathematical Statistics, John Wiley & Sons, Ltd., Chichester, 1978.
- [BCR] C. BERG, J. P. R. CHRISTENSEN AND P. RESSEL, “*Harmonic analysis on semigroups*”, Theory of positive definite and related functions, Graduate Texts in Mathematics **100**, Springer-Verlag, New York, 1984.

- [Br] L. BROWN, Sufficient statistics in the case of independent random variables, *Ann. Math. Statist* **35** (1964), 1456–1474.
- [dB] A. DI BUCCHIANICO, “*Probabilistic and analytical aspects of the umbral calculus*”, CWI Tract **119**, Stichting Mathematisch Centrum, Centrum voor Wiskunde en Informatica, Amsterdam, 1997.
- [CPR] G. CORACH, H. PORTA AND L. RECHT, The geometry of the space of selfadjoint invertible elements in a  $C^*$ -algebra, *Integral Equations Operator Theory* **16(3)** (1993), 333–359.
- [Ef] B. EFRON, The geometry of exponential families, *Ann. Statist.* **6(2)** (1978), 362–376.
- [Es] F. ESSCHER, On the probability function in the collective theory of risk, *Scand. Actuar. J.* **15** (1932), 175–195.
- [GRS] I. GELFAND, D. RAIKOV AND G. SHILOV, “*Commutative normed rings*”, Translated from the Russian, with a supplementary chapter, Chelsea Publishing Co., New York, 1964.
- [GP] P. GIBILISCO AND G. PISTONE, Connections on non-parametric statistical manifolds by Orlicz space geometry, *Infin. Dimens. Anal. Quantum Probab. Relat. Top.* **1(2)** (1998), 325–347.
- [GR] H. GZYL AND L. RECHT, A geometry in the space of probabilities II: Projective spaces and exponential families, *Rev. Mat. Iberoamericana* (to appear).
- [J] E. T. JAYNES, Information theory and statistical mechanics, *Phys. Rev. (2)* **106** (1957), 620–630.
- [KN] S. KOBAYASHI AND K. NOMIZU, “*Foundations of differential geometry*”, Vol. II, Interscience Tracts in Pure and Applied Mathematics **15**, Vol. II, Interscience Publishers John Wiley & Sons, Inc., New York-London-Sydney, 1969.
- [K] S. KULLBACK, “*Information theory and statistics*”, 2nd. revised edition, Dover Publications, Inc., Mineola, NY, 1968.
- [LM] R. J. LARSEN AND M. L. MARX, “*An Introduction to Mathematical Statistics and its Applications*”, Prentice-Hall, Englewood Cliffs, N.J., 2006.
- [PRo] G. PISTONE AND M. P. ROGANTIN, The exponential statistical manifold: mean parameters, orthogonality and space transformations, *Bernoulli* **5(4)** (1999), 721–760.
- [PS] G. PISTONE AND C. SEMPI, An infinite-dimensional geometric structure on the space of all the probability measures equivalent to a given one, *Ann. Statist.* **23(5)** (1995), 1543–1561.

- [PR1] H. PORTA AND L. RECHT, Conditional expectations and operator decompositions, *Ann. Global Anal. Geom.* **12(4)** (1994), 335–339.
- [PR2] H. PORTA AND L. RECHT, Exponential sets and their geometric motions, *J. Geom. Anal.* **6(2)** (1996), 277–285.
- [S] R. F. STREATER, Classical and quantum info-manifolds, Analytical study of quantum information and related fields (Japanese) (Kyoto, 2000), *Sūrikaisekikenkyūsho Kōkyūroku* **1196** (2001), 32–51.
- [V] I. VAJDA, “*Theory of statistical inference and information*”, Theory and Decision Library B **11**, Kluwer Academic Publishers, Dordrecht, 1989.
- [W] D. WILLIAMS, “*Weighing the odds*”, A course in probability and statistics, Cambridge University Press, Cambridge, 2001.

Henryk Gzyl:  
Centro de Finanzas IESA  
San Bernardino  
Caracas 1010  
Venezuela

Departamento de Cómputo Científico y Estadística  
Universidad Simón Bolívar  
Caracas  
Venezuela  
*E-mail address:* hgzyl@reacciun.ve

Lázaro Recht:  
Departamento de Matemáticas  
Universidad Simón Bolívar  
Apartado Postal 89000  
Caracas 1080-A  
Venezuela  
*E-mail address:* recht@usb.ve

Primera versió rebuda el 31 de maig de 2006,  
darrera versió rebuda el 17 de gener de 2007.