

# Nonparametric Information Geometry

Giovanni Pistone

Collegio Carlo Alberto, Via Real Collegio 30, 10024 Moncalieri, Italy,  
giovanni.pistone@carloalberto.org,  
WWW home page: <http://www.giannidioresino.it/index.html>

**Abstract.** The differential-geometric structure of the set of positive densities on a given measure space has raised the interest of many mathematicians after the discovery by C.R. Rao of the geometric meaning of the Fisher information. Most of the research is focused on parametric statistical models. In series of papers by author and coworkers a particular version of the nonparametric case has been discussed. It consists of a minimalistic structure modeled according the theory of exponential families: given a reference density other densities are represented by the centered log likelihood which is an element of an Orlicz space. This mappings give a system of charts of a Banach manifold. It has been observed that, while the construction is natural, the practical applicability is limited by the technical difficulty to deal with such a class of Banach spaces. It has been suggested recently to replace the exponential function with other functions with similar behavior but polynomial growth at infinity in order to obtain more tractable Banach spaces, e.g. Hilbert spaces. We give first a review of our theory with special emphasis on the specific issues of the infinite dimensional setting. In a second part we discuss two specific topics, differential equations and the metric connection. The position of this line of research with respect to other approaches is briefly discussed.

**Keywords:** Information Geometry, Banach Manifold

## 1 Introduction

In the present paper we follow closely the presentation of Information Geometry developed by S.-I. Amari and coworkers, see e.g. in [1], [2], [3], [4], with the specification that we want to construct a Banach manifold structure in the classical sense, see e.g. [5] or [6], without any restriction to parametric models. We feel that the non parametric approach is of interest even in the case of a finite state space. We build upon our previous work in this field, namely [7], [8], [9], [10], [11], [12], [13], [14], [15], [16], [17], [18],[19], [20], [21],[22]. Other contributions are referred to in the text below. We do not discuss here the non commutative/quantum case as developed e.g., in [23], [24] and the review in [25].

The rest of this introductory section contains a review of relevant facts related with the topology of Orlicz spaces which are the model spaces in our manifold structure. The review part is based on previous joint work with M. P. Rogantin

[8] and A. Cena [11], but a number of examples and remarks are added in order to clarify potential issues and possible applications. The exponential manifold (originally introduced in the joint work with C. Sempì [7]) is critically reviewed in Sec. 2, together with applications. Differential equations are discussed in Sec. 3, with examples. Sec. 4 deals with the Hilbert bundle of the exponential manifold and the computation of the metric derivative. It builds upon previous work on non parametric connections with P. Gibilisco [9]. A variation on exponential manifolds is introduced in Sec. 5 to show it could be developed along the lines previously discussed.

### 1.1 Model spaces

In this paper we consider a fixed  $\sigma$ -finite measure space  $(\Omega, \mathcal{F}, \mu)$  and we denote by  $\mathcal{P}_>$  the set of all densities which are positive  $\mu$ -a.s. The set of densities, without any further restriction is  $\mathcal{P}_\geq$ , while  $\mathcal{P}_1$  is the set of measurable functions  $f$  with  $\int f d\mu = 1$ . In the finite state space case, i.e.  $\#\Omega < \infty$ ,  $\mathcal{P}_1$  is a plane,  $\mathcal{P}_\geq$  is the simplex,  $\mathcal{P}_>$  its topological interior. In the infinite case, the setting is much more difficult: we concentrate here mainly on strictly positive densities and we construct its geometry by taking as a guiding model the theory of exponential families, see [26], [27], [28], [29]. A non parametric approach we use was initially suggested by P. Dawid [30,31]. A geometry derived from exponential families is intrinsically based on the positivity of the densities, see [32,33].

At each  $p \in \mathcal{P}_>$  we associate a set of densities of the form  $q = e^{u-K} \cdot p$ , where  $u$  belongs to a suitable Banach space  $B_p$  and  $K$  is a constant depending on  $p$  and  $u$ . The mapping  $u \mapsto q$  will be one-to-one and its inverse  $s_p: q \mapsto u$  will be a chart of our exponential manifold  ${}^e\mathcal{P} = (\mathcal{P}_>, \{s_p\})$ . As we do not have manifold structures on the set of positive densities other than the exponential one, in the following the manifold and the set are both denoted by  $\mathcal{P}_>$ .

We refer to [5, §5-7] and [6] for the theory of manifolds modeled on Banach spaces. According to this definition, a manifold is a set  $\mathcal{P}_>$  together with a collection or atlas of charts  $s: \mathcal{U} \rightarrow B$  from a subset  $\mathcal{U} \subset \mathcal{P}_>$  to a Banach space  $B$  such that for each couple of charts the transition maps  $s' \circ s^{-1}: s(\mathcal{U} \cap \mathcal{U}') \rightarrow s'(\mathcal{U} \cap \mathcal{U}')$  are smooth functions from an open set of  $B$  into an open set of  $B'$ . In this geometric approach,  $\mathcal{P}_>$  is a set, while all structure is in model spaces  $B$ .

It should be noted that the Banach spaces are not required to be equal, as the finite dimensional case seems to suggest, but they should be isomorphic when connected by a chart. Actually this freedom is of much use in our application to statistical model, but requires a careful discussion of the isomorphism. Precisely, at each  $p \in \mathcal{P}_>$ , the model space  $B_p$  for our manifold is an Orlicz space of centered random variables, see [34], [35, Chapter II], [36], [37, Ch 8]. We review briefly our notations and recall some basic facts from these references.

If both  $\phi$  and  $\phi^{-1} = \phi_*$  are monotone, continuous functions on  $\mathbb{R}_{\geq 0}$  onto itself, we call the pair

$$\Phi(x) = \int_0^{|x|} \phi(u) du, \quad \Phi_*(y) = \int_0^{|y|} \phi^{-1}(v) dv,$$

a *Young pair*. Each Young pair satisfies the Young inequality  $|xy| \leq \Phi(x) + \Phi_*(y)$ , with equality if, and only if,  $y = \phi(x)$ . The relation in a Young pair is symmetric and either element is called a Young function.

*Example 1 (Young pairs).* We will use the following Young pairs:

	$\phi$	$\phi^{-1} = \phi_*$	$\Phi$	$\Phi_*$
(a)	$\log(1+u)$	$e^v - 1$	$(1+ x ) \log(1+ x ) -  x $	$e^{ y } - 1 -  y $
(b)	$\sinh^{-1} u$	$\sinh v$	$ x  \sinh^{-1}  x  - \sqrt{1+x^2} + 1$	$\cosh y - 1$
(c)	$\log^+ u$	$e^v$	$ x  \log^+  x  - (x-1)^+$	$e^{ y } - 1$
(2)	$u$	$v$	$\frac{1}{2}x^2$	$\frac{1}{2}y^2$

As  $\log(1+u) \leq \sinh^{-1} u = \log(u + \sqrt{1+u^2}) \leq a \log(1+u)$  if  $u \geq 0$  and  $a > 1$ , the pairs (a) and (b) are equivalent i.e.  $\Phi_a \leq \Phi_b \leq a\Phi_a$  if  $a > 1$ . Moreover, from  $\Phi_a(x) = \int_0^x (x-u)/(1+u) du$  if  $x \geq 0$ , we obtain an instance of the so called  $\Delta_2$ -condition,  $\Phi_a(ax) \leq a^2\Phi_a(x)$ . This condition is not satisfied by  $\Phi_{a*}$  as  $\Phi_{a*}(2y)/\Phi_{a*}(y)$  is unbounded as  $y \rightarrow \infty$ . The listed pairs satisfy  $\Phi_{(a)}(x) \leq \Phi_{(2)} \leq \Phi_{(a)*}$ .

In fact, around each  $p$  we consider densities of the form  $q \propto e^v$  for some random variable  $v$  and, moreover, we require the one dimensional exponential family  $q(t) \propto e^{tv}$  be defined for each  $t$  in an open interval  $I$  containing 0. In other words, we require the moment generating function  $t \mapsto \int e^{tv} p d\mu = E_p[e^{tv}]$  to be finite in a neighbourhood of 0. The set of such random variables  $v$  is a vector space and a Banach space for a properly defined norm. We discuss below those parts of the theory which are relevant for the definition of exponential manifold.

If  $\Phi(x) = \cosh x - 1$ , a real random variable  $u$  belongs to the vector space  $L^\Phi(p)$  if  $E_p[\Phi(\alpha v)] < +\infty$  for some  $\alpha > 0$ . A norm is obtained by defining the set  $\{v : E_p[\Phi(\alpha v)] \leq 1\}$  to be the closed unit ball. It follows that the open unit ball consists of those  $u$ 's such that  $\alpha u$  is in the closed unit ball for some  $\alpha > 1$ . The corresponding norm  $\|\cdot\|_{\Phi,p}$  is called Luxemburg norm and defines a Banach space, see e.g. [35, Th 7.7]. The function  $\cosh - 1$  has been chosen here because the condition  $E_p[\Phi(\alpha v)] < +\infty$  is clearly equivalent to  $E_p[e^{tv}] < +\infty$  for  $t \in [-\alpha, \alpha]$ , but other choices will define the same Banach space e.g.,  $\Phi(x) = e^{|x|} - |x| - 1$ . By abuse of notation, we will denote all these equivalent functions by  $\Phi$ .

The main technical issue in working with Orlicz spaces such as  $L^{(\cosh - 1)}(p)$  is the regularity of its unit sphere  $S = \{u : \|u\|_{(\cosh - 1),p} = 1\}$ . In fact, while  $E_p[\cosh u - 1] = 1$  implies  $u \in S$ , the latter implies  $E_p[\cosh u - 1] \leq 1$ . Subspaces of  $L^\Phi$  where this cannot happen are called *steep*. If the state space is finite, the full space is steep, see the Ex. 2 and 3 below. The relevance of steep families in exponential families is discussed in [27]. Steepness is important when related with the idea of embedding. Consider the mapping  $\Phi_+^{-1} : \mathcal{P}_> \ni p \mapsto v = \Phi_+^{-1}(p)$ ,  $\Phi_+ = \Phi|_{\mathbb{R}_>}$ . Then  $\int \Phi(v) d\mu = \int p d\mu = 1$  hence  $\|u\|_\Phi = 1$  and we have an embedding of  $\mathcal{P}_>$  into the sphere of a Banach space.

*Example 2 (Boolean state space).* In the case of a finite state space, the moment generating function is finite everywhere, but its computation can be challenging. We discuss in particular the boolean case  $\Omega = \{+1, -1\}^n$  with counting

reference measure  $\mu$  and uniform density  $p(x) = 2^{-n}$ ,  $x \in \Omega$ . In this case there is a huge literature from statistical physics, e.g., [38, Ch. VII]. A generic real function on  $\Omega$ —called in the machine learning literature pseudo-boolean [39]—has the form  $u(x) = \sum_{\alpha \in L} \hat{u}(\alpha)x^\alpha$ , with  $L = \{0, 1\}^n$ ,  $x^\alpha = \prod_{i=1}^n x_i^{\alpha_i}$ ,  $\hat{u}(\alpha) = 2^{-n} \sum_{x \in \Omega} u(x)x^\alpha$ .

As  $e^{ax} = \cosh(a) + \sinh(a)x$  if  $x^2 = 1$  i.e.  $x = \pm 1$ , we have

$$\begin{aligned} e^{tu(x)} &= \exp \left( \sum_{\alpha \in \text{Supp } \hat{u}} t\hat{u}(\alpha)x^\alpha \right) \\ &= \prod_{\alpha \in \text{Supp } \hat{u}} (\cosh(t\hat{u}(\alpha)) + \sinh(t\hat{u}(\alpha))x^\alpha) \\ &= \sum_{B \subset \text{Supp } \hat{u}} \prod_{\alpha \in B^c} \cosh(t\hat{u}(\alpha)) \prod_{\alpha \in B} \sinh(t\hat{u}(\alpha))x^{\sum_{\alpha \in B} \alpha}. \end{aligned}$$

The moment generating function of  $u$  under the uniform density  $p$  is

$$t \mapsto \sum_{B \in \mathcal{B}(\hat{u})} \prod_{\alpha \in B^c} \cosh(t\hat{u}(\alpha)) \prod_{\alpha \in B} \sinh(t\hat{u}(\alpha)),$$

where  $\mathcal{B}(\hat{u})$  are those  $B \subset \text{Supp } \hat{u}$  such that  $\sum_{\alpha \in B} \alpha = 0 \pmod{2}$ . We have

$$E_p[\Phi](tu) = \sum_{B \in \mathcal{B}_0(\hat{u})} \prod_{\alpha \in B^c} \cosh(t\hat{u}(\alpha)) \prod_{\alpha \in B} \sinh(t\hat{u}(\alpha)) - 1,$$

where  $\mathcal{B}_0(\hat{u})$  are those  $B \subset \text{Supp } \hat{u}$  such that  $\sum_{\alpha \in B} \alpha = 0 \pmod{2}$  and moreover  $\sum_{\alpha \in \text{Supp } \hat{u}} \alpha = 0$ .

If  $S$  is the  $\{1, \dots, n\} \times \text{Supp } \hat{u}$  matrix with elements  $\alpha_i$  we want to solve the system  $Sb = 0 \pmod{2}$  to find all elements of  $\mathcal{B}$ ; we want to add the equation  $\sum b = 0 \pmod{2}$  to find  $\mathcal{B}_0$ . The simplest example is the simple effect model  $u(x) = \sum_{i=1}^n c_i x_i$ .

*Example 3 (The sphere  $S$  is not smooth).* We look for the moment generating function of the density  $p(x) \propto (a+x)^{-\frac{3}{2}}e^{-x}$ ,  $x > 0$ , where  $a > 0$ . From the incomplete gamma integral

$$\Gamma\left(-\frac{1}{2}, x\right) = \int_x^\infty s^{-\frac{1}{2}-1} e^{-s} ds, \quad x > 0,$$

we have for  $\theta, a > 0$ ,

$$\frac{d}{dx} \Gamma\left(-\frac{1}{2}, \theta(a+x)\right) = -\theta^{-\frac{1}{2}} e^{-\theta a} (a+x)^{-\frac{3}{2}} e^{-\theta x}.$$

We have, for  $\theta \in \mathbb{R}$  and  $a > 0$

$$C(\theta, a) = \int_0^\infty (a+x)^{-\frac{3}{2}} e^{-\theta x} dx = \begin{cases} \sqrt{\theta} e^{\theta a} \Gamma\left(-\frac{1}{2}, \theta a\right) & \text{if } \theta > 0, \\ \frac{2}{\sqrt{a}} & \text{if } \theta = 0, \\ +\infty & \text{if } \theta < 0, \end{cases} \quad (1)$$

or, using the Gamma distribution with shape  $1/2$  and scale  $1$ ,  $\Gamma(-\frac{1}{2}, x) = 2x^{-1/2}e^{-x} - \sqrt{\pi}(1 - \Gamma(x; 1/2, 1))$ ,

$$C(\theta, a) = \begin{cases} 2a^{-\frac{1}{2}} - 2\sqrt{\pi}\theta e^{\theta a} (1 - \Gamma(\theta a; 1/2, 1)) & \text{if } \theta \geq 0, \\ +\infty & \text{if } \theta < 0. \end{cases}$$

The density  $p$  is obtained from (1) with  $\theta = 1$ ,

$$p(x) = C(1, a)^{-1}(a+x)^{-\frac{3}{2}}e^{-x} = \frac{(a+x)^{-\frac{3}{2}}e^{-x}}{e^a \Gamma(-\frac{1}{2}, a)}, \quad x > 0,$$

and, for the random variable  $u(x) = x$ , the function

$$\begin{aligned} \alpha \mapsto E_p[\Phi(\alpha u)] &= \frac{1}{e^a \Gamma(-\frac{1}{2}, a)} \int_0^\infty (a+x)^{-\frac{3}{2}} \frac{e^{-(1-\alpha)x} + e^{-(1+\alpha)x}}{2} dx - 1 \\ &= \frac{C(1-\alpha, a) + C(1+\alpha, a)}{2C(1, a)} - 1 \end{aligned} \quad (2)$$

is convex lower semi-continuous on  $\alpha \in \mathbb{R}$ , finite for  $\alpha \in [-1, 1]$ , infinite otherwise, hence not steep. Its value at  $\alpha = \pm 1$  is

$$E_p[\Phi(u)] = \frac{C(0, a) + C(2, a)}{2C(1, a)} - 1 \quad \text{e.g.,} = 0.8037381 \text{ if } a = \frac{1}{2}.$$

If the functions  $\Phi$  and  $\Phi_*$  are Young pair, for each  $u \in L^\Phi(p)$  and  $v \in L^{\Phi_*}(p)$ , such that  $\|u\|_{\Phi, p}, \|v\|_{\Phi_*, p} \leq 1$ , we have  $E_p[uv] \leq 2$ , hence

$$L^{\Phi_*}(p) \times L^\Phi(p) \ni (v, u) \mapsto E_p[uv]$$

is a duality mapping,  $|\langle u, v \rangle_p| \leq 2 \|u\|_{\Phi_*, p} \|v\|_{\Phi, p}$ .

A sequence  $u_n$ ,  $n = 1, 2, \dots$  is convergent to 0 for such a norm if and only if for all  $\epsilon > 0$  there exists a  $n(\epsilon)$  such that  $n > n(\epsilon)$  implies  $E_p[\Phi_1(\frac{u_n}{\epsilon})] \leq 1$ . Note that  $|u| \leq |v|$  implies

$$E_p\left[\Phi_1\left(\frac{u}{\|v\|_{\Phi_1, p}}\right)\right] \leq E_p\left[\Phi_1\left(\frac{v}{\|v\|_{\Phi_1, p}}\right)\right] \leq 1$$

so that  $\|u\|_{\Phi_1, p} \leq \|v\|_{\Phi_1, p}$ .

In defining our manifold, we need to show that Orlicz spaces defined at different points of statistical models are isomorphic, we will use frequently the fact that following lemma, see [11, Lemma 1].

**Lemma 1.** *Let  $p \in \mathcal{M}$  and let  $\Phi_0$  be a Young function. If the Orlicz spaces  $L^{\Phi_0}(p)$  and  $L^{\Phi_0}(q)$  are equal as sets, then their norms are equivalent.*

The condition  $u \in L^{\cosh^{-1}}(p)$  is equivalent to the existence of the moment generating function  $g(t) = E_p[e^{tu}]$  on a neighbourhoods of 0. The case when

such a moment generating function is defined on all of the real line is special and defines a notable subspace of the Orlicz space see e.g., [36]. Such spaces could be the model of an alternative definition of as in [40].

In fact, the Banach space  $L^\Phi(p)$ ,  $\phi = \cosh -1$  is not separable, unless the basic space has a finite number of atoms. In this sense it is an unusual choice from the point of view of functional analysis and manifold's theory. However,  $L^\Phi(p)$  is natural for statistics because for each  $u \in L^{\Phi_1}(p)$  the Laplace transform of  $u$  is well defined at 0, then the one-dimensional exponential model  $p(\theta) \propto e^{\theta u}$  is well defined.

However, the space  $L^{\Phi^*}(p)$  is separable and its dual space is  $L^\Phi(p)$ , the duality pairing being  $(u, v) \mapsto E_p[uv]$ . This duality extends to a continuous chain of spaces:

$$L^{\Phi_1}(p) \rightarrow L^a(p) \rightarrow L^b(p) \rightarrow L^{\Psi_1}(p), \quad 1 < b \leq 2, \quad \frac{1}{a} + \frac{1}{b} = 1$$

where  $\rightarrow$  denotes continuous injection.

From the duality pairing of conjugate Orlicz spaces and the characterization of the closed unit ball it follows a definition of dual norm on  $L^{\Phi^*}(p)$ :

$$N_p(v) = \sup \{E_p[uv] : E_p[\Phi(u)] \leq 1\}.$$

## 1.2 Moment generating functional and cumulant generating functional

In this section we review a number of key technical results. Most of the results are related with the smoothness of the superposition operator  $L^\Phi(p) : v \mapsto \exp \circ v$ . Superposition operators on Orlicz spaces are discussed e.g. in [34] and [41, Ch 4]. Banach analytic functions are discussed in [5], [42] and [43].

Let  $p \in \mathcal{P}_>$  be given. The following theorem has been proved in [10, Ch 2], see also [11].

### Proposition 1.

1. For  $a \geq 1$ ,  $n = 0, 1, \dots$  and  $u \in L^\Phi(p)$ ,

$$\lambda_{a,n}(u) : (w_1, \dots, w_n) \mapsto \frac{w_1}{a} \dots \frac{w_n}{a} e^{\frac{u}{a}}$$

is a continuous, symmetric,  $n$ -multi-linear map from  $L^\Phi(p)$  to  $L^a(p)$ .

2.  $v \mapsto \sum_{n=0}^{\infty} \frac{1}{n!} \left(\frac{v}{a}\right)^n$  is a power series from  $L^\Phi(p)$  to  $L^a(p)$  with radius of convergence  $\geq 1$ .
3. The superposition mapping  $v \mapsto e^{v/a}$  is an analytic function from the open unit ball of  $L^\Phi(p)$  to  $L^a(p)$ .

**Definition 1.** Let  $\Phi = \cosh -1$  and  $B_p = L_0^\Phi(p)$ ,  $p \in \mathcal{P}_>$ . The moment generating functional is  $M_p : L^\Phi(p) \ni u \mapsto E_p[e^u] \in \mathbb{R}_> \cup \{+\infty\}$ . The cumulant generating functional is  $K_p : B_p \ni u \mapsto \log M_p(u) \in \mathbb{R}_> \cup \{+\infty\}$ .

**Proposition 2.**

1.  $M_p(0) = 1$ ; otherwise, for each centered random variable  $u \neq 0$ ,  $M_p(u) > 1$ .
2.  $M_p$  is convex and lower semi-continuous, and its proper domain is a convex set which contains the open unit ball of  $L^\Phi(p)$ ; in particular the interior of such a domain is a non empty convex set.
3.  $M_p$  is infinitely Gâteaux-differentiable in the interior of its proper domain, the  $n$ th-derivative at  $u$  in the direction  $v \in L^\Phi(p)$  being

$$\left. \frac{d^n}{dt^n} M_p(u + tv) \right|_{t=0} = E_p [v^n e^u];$$

4.  $M_p$  is bounded, infinitely Fréchet-differentiable and analytic on the open unit ball of  $L^\Phi(p)$ , the  $n$ th-derivative at  $u$  evaluated in  $(v_1, \dots, v_n) \in L^\Phi(p) \times \dots \times L^\Phi(p)$  is

$$D^n M_p(u)(v_1, \dots, v_n) = E_p [v_1 \dots v_n e^u].$$

**Proposition 3.**

1.  $K_p(0) = 0$ ; otherwise, for each  $u \neq 0$ ,  $K_p(u) > 0$ .
2.  $K_p$  is convex and lower semi-continuous, and its proper domain is a convex set which contains the open unit ball of  $B_p$ ; in particular the interior of such a domain is a non empty convex set.
3.  $K_p$  is infinitely Gâteaux-differentiable in the interior of its proper domain.
4.  $K_p$  is bounded, infinitely Fréchet-differentiable and analytic on the open unit ball of  $\mathcal{V}_p$ .

Other properties of the key functional  $K_p$  are described below as they relate directly to the exponential manifold.

### 1.3 Families of Orlicz spaces

In statistical models, we associate to each density  $p$  a space of  $p$ -centered random variables to represent scores or estimating functions. For example, if the one-parameter statistical model  $p(t)$ ,  $t \in I$ ,  $I$  open interval, is regular enough, then  $u(t) = \frac{d}{dt} \log p(t)$  satisfies  $E_{p(t)} [u(t)] = 0$  for all  $t \in I$ . It is crucial to discuss how the relevant spaces of  $p$ -centered random variables depend on the variation of the density  $p$ , that is it is crucial to understand the variation of the spaces  $B_p = L_0^\Phi(p)$  and  ${}^*B_p = L_0^{\Phi^*}(p)$  along a one-dimensional statistical model  $p(t)$ ,  $t \in I$ . In Information Geometry, those spaces contain models for the tangent and cotangent spaces of the statistical models. On two different points of a regular model, they must be isomorphic, or, in particular, equal.

We use a peculiar notion of connection by arcs, which is different from what is usually meant with this name. Given  $p, q \in \mathcal{P}_>$ , the exponential model  $p(\theta) \propto p^{1-\theta} q^\theta$ ,  $0 \leq \theta \leq 1$  connects the two given densities as end points of a curve,  $p(\theta) \propto \exp\left(\theta \log \frac{q}{p}\right) \cdot p$ , where  $\log \frac{q}{p}$  is not in the exponential Orlicz space at  $p$  unless  $\theta$  can be extended to assume negative values.

**Definition 2.** We say that  $p, q \in \mathcal{P}_>$  are connected by an open exponential arc if there exist  $r \in \mathcal{P}_>$  and an open interval  $I$ , such that  $p(t) \propto e^{tr}$ ,  $t \in I$ , is an exponential model containing both  $p$  and  $q$  at  $t_0, t_1$  respectively. By the change of parameter  $s = t - t_0$ , we can always reduce to the case where  $r = p$  and  $u \in L^\Phi(p)$ .

The open connection of Def. 2 is an equivalence relation.

**Definition 3.** Let us denote by  $\mathcal{S}_p$  the interior of the proper domain of the cumulant generating functional  $K_p$ . For every density  $p \in \mathcal{P}_>$ , the maximal exponential model at  $p$  is defined to be the family of densities

$$\mathcal{E}(p) := \left\{ e^{u - K_p(u)} \cdot p : u \in \mathcal{S}_p \right\}.$$

**Proposition 4.** The following statements are equivalent:

1.  $q \in \mathcal{M}$  is connected to  $p$  by an open exponential arc;
2.  $q \in \mathcal{E}(p)$ ;
3.  $\mathcal{E}(p) = \mathcal{E}(q)$ ;
4.  $\log \frac{q}{p}$  belongs to both  $L^{\Phi_1}(p)$  and  $L^{\Phi_1}(q)$ .
5.  $L^{\Phi_1}(p)$  and  $L^{\Phi_1}(q)$  are equal as vector spaces and their norms are equivalent.

In the following proposition we have collected a number of properties of the maximal exponential model  $\mathcal{E}(p)$  which are relevant for its manifold structure.

**Proposition 5.** Assume  $q = e^{u - K_p(u)} \cdot p \in \mathcal{E}(p)$ .

1. The first two derivatives of  $K_p$  on  $\mathcal{S}_p$  are

$$\begin{aligned} DK_p(u)v &= E_q[v], \\ D^2K_p(u)(v_1, v_2) &= \text{Cov}_q(v_1, v_2) \end{aligned}$$

2. The random variable  $\frac{q}{p} - 1$  belongs to  ${}^*B_p$  and

$$DK_p(u)v = E_p \left[ \left( \frac{q}{p} - 1 \right) v \right].$$

In other words the gradient of  $K_p$  at  $u$  is identified with an element of  ${}^*B_p$ , denoted by  $\nabla K_p(u) = e^{u - K_p(u)} - 1 = \frac{q}{p} - 1$ .

3. The mapping  $B_p \ni u \mapsto \nabla K_p(u) \in {}^*B_p$  is monotonic, in particular one-to-one.
4. The weak derivative of the map  $\mathcal{S}_p \ni u \mapsto \nabla K_p(u) \in {}^*B_p$  at  $u$  applied to  $w \in B_p$  is given by

$$D(\nabla K_p(u))w = \frac{q}{p} (w - E_q[w]),$$

and it is one-to-one at each point.

5. The mapping  ${}^m\mathbb{U}_p^q : v \mapsto \frac{q}{p}v$  is an isomorphism of  ${}^*B_p$  onto  ${}^*B_q$ .



6.  $q/p \in L^{\Phi_*}(p)$ .
7.  $D(q\|p) = DK_p(u)u - K_p(u)$  with  $q = e^{u-K_p(u)}p$ , in particular  $-D(q\|p) < +\infty$ .
8.
$$B_q = L_0^{\Phi_1}(q) = \left\{ u \in L^{\Phi_1}(p) : E_p \left[ u \frac{q}{p} \right] = 0 \right\}.$$
9.  ${}^e\mathbb{U}_p^q : u \mapsto u - E_q[u]$  is an isomorphism of  $B_p$  onto  $B_q$ .

## 2 Exponential and mixture manifolds

### 2.1 Exponential manifold

If  $p, q \in \mathcal{M}$  are connected by an open exponential arc, then the random variable  $u \in \mathcal{S}_p$  such that  $q \propto e^u p$  is unique and it is equal to  $\log \frac{q}{p} - E_p \left[ \log \frac{q}{p} \right]$ . In fact,  $q \propto e^u p$  for some  $u \in L^{\Phi_1}(p)$  if and only if  $u - \log \frac{q}{p}$  is a constant. If  $u \in \mathcal{S}_p \subset B_p$ , then  $u - \log \frac{q}{p} = K_p(u)$  and, as  $u$  is centered, it follows that  $-E_p \left[ \log \frac{q}{p} \right] = K_p(u)$  and  $u = \log \frac{q}{p} - E_p \left[ \log \frac{q}{p} \right]$ . Indeed,  $u$  is the projection of  $\log \frac{q}{p}$  onto  $B_p$  in the split  $L^{\Phi_1}(p) = B_p \oplus \langle 1 \rangle$ .

**Definition 4.** We define two one-to-one mappings: the parameterization or patch  $e_p : \mathcal{S}_p \rightarrow \mathcal{E}(p)$ ,  $e_p(u) = e^{u-K_p(u)} \cdot p$  and the chart  $s_p : \mathcal{E}(p) \rightarrow \mathcal{S}_p$ ,  $s_p(q) = \log(\text{frac} q p) - E_p \left[ \log \left( \frac{q}{p} \right) \right]$ .

**Proposition 6.** If  $p_1, p_2 \in \mathcal{E}(p)$ , then the transition mapping  $s_{p_2} \circ e_{p_1} : \mathcal{S}_{p_1} \rightarrow \mathcal{S}_{p_2}$  is the restriction of an affine function from  $B_{p_1} \rightarrow B_{p_2}$

$$u \mapsto u + \log \left( \frac{p_1}{p_2} \right) - E_{p_2} \left[ u + \log \left( \frac{p_1}{p_2} \right) \right].$$

The derivative of the transition map  $s_{p_2} \circ e_{p_1}$  is the isomorphism of  $B_{p_1}$  onto  $B_{p_2}$

$$B_{p_1} \ni u \mapsto u - E_{p_2}[u] = {}^e\mathbb{U}_{p_1}^{p_2} \in B_{p_2}.$$

**Definition 5.** The exponential manifold is defined by the atlas of charts in Def. 4. It is an affine manifold because of Prop. 6. Each  $\mathcal{E}(p)$  is a connected component.

A metric topology called e-topology is induced by the exponential manifold on  $\mathcal{P}_{>}$ , namely a sequence  $\{p_n\}$ ,  $n \in \mathbb{N}$ , is e-convergent to  $p$  if and only if sequences  $\{p_n/p\}$  and  $\{p/p_n\}$  are convergent to 1 in each  $L^\alpha(p)$ ,  $\alpha > 1$ .

Mixture arcs are regular in each connected component  $\mathcal{E}$  of the exponential manifold.

### Proposition 7.

1. If  $q \in \mathcal{E}(p)$ , then the mixture model  $p(\lambda) = (1-\lambda)p + \lambda q \in \mathcal{E}(p)$  for  $\lambda \in [0, 1]$ .
2. An open mixture arc  $p(t) = (1-t)p + tq$ ,  $t \in ]-\alpha, 1+\beta[$ ,  $\alpha, \beta > 0$  is e-continuous.

## 2.2 Mixture manifold

We are not able to define a mixture manifold with the same support as the exponential manifold. For each  $p \in \mathcal{P}_>$  and each  $u \in \mathcal{S}_p$ ,  $q = e^{u - K_p(u)} \cdot p$ , the derivative of  $K_p$  at  $u$ , in the direction  $v \in B_p$ , is  $DK_p(u) \cdot v = E_p \left[ \left( \frac{q}{p} - 1 \right) v \right]$ , and it is identified to its gradient  $\nabla K_p(u) = q/p - 1 \in {}^*B_p$ . The mapping  $q \mapsto q/p - 1 \in {}^*B_p$  cannot be a chart because its values are bounded below by  $-1$  but it is strongly reminiscent of the mean parameterization  $\boldsymbol{\eta} = \nabla \psi(\boldsymbol{\theta})$  in parametric exponential families  $p_\theta = \exp(\boldsymbol{\theta} \cdot \mathbf{T} - \psi(\boldsymbol{\theta})) \cdot p_0$ .

We move to the larger set  $\mathcal{P}_1 = \{f : \int f d\mu = 1\} \supset \mathcal{P}_>$  and for each  $p \in \mathcal{E}(p)$  we introduce the subset  ${}^*\mathcal{U}_p$  defined by the condition  $\frac{q}{p} \in L^{\Phi^*}(p)$ . Our chart is the map

$$\eta_p : {}^*\mathcal{U}_p \ni q \mapsto \frac{q}{p} - 1 \in {}^*B_p. \quad (3)$$

As  $\eta_p(q)$ , for  $q \in \mathcal{U}_p \subset \mathcal{E}(p)$ , equals  $v \mapsto E_q[v]$ , it is the non parametric version of the so called expectation parameter. This mapping is bijective and its inverse is:

$$\eta_p^{-1} : {}^*B_p \ni u \mapsto (u + 1)p \in {}^*\mathcal{U}_p.$$

The collection of sets  $\{{}^*\mathcal{U}_p\}_{p \in \mathcal{P}_\geq}$  is a covering of  $\mathcal{P}_1$ .

There is a nice characterization of the elements of  ${}^*\mathcal{U}_p \cap \mathcal{P}_\geq$ : they are all the probability densities with finite divergence with respect to  $p$ , see [11, Prop 31]. Moreover  $\mathcal{U}_p \subset {}^*\mathcal{U}_p$  and  $p_1, p_2 \in \mathcal{E}(p)$  implies  ${}^*\mathcal{U}_{p_1} = {}^*\mathcal{U}_{p_2}$ . In conclusion, we can define the mixture manifold as follows.

For each pair  $p_1, p_2 \in \mathcal{E}(p)$  we have the affine transition map

$$\eta_{p_2} \circ \eta_{p_1}^{-1} : \begin{cases} {}^*B_{p_1} \rightarrow {}^*B_{p_2} \\ u \mapsto u \frac{p_1}{p_2} + \frac{p_1}{p_2} - 1 \end{cases}$$

and the subset of  $\mathcal{P}_1$ ,  ${}^*\mathcal{E}(p) = \left\{ q \in \mathcal{P}_1 : \frac{q}{p} \in L^{\Phi^*}(p) \right\}$ , which is equal to  ${}^*\mathcal{U}_q$  if  $q \in \mathcal{E}(p)$ .

**Proposition 8.** *Let  $p \in \mathcal{P}_>$  be given. The collection of charts*

$$\{({}^*\mathcal{U}_q, \eta_q) : q \in \mathcal{E}(p)\}$$

*is an affine  $C^\infty$ -atlas on  ${}^*\mathcal{E}(p)$ .*

The *mixture manifold* is defined by the atlas in Prop. 8. The mixture manifold is an extension of the exponential manifold.

**Proposition 9.** *For each density  $p \in \mathcal{P}_>$ , the inclusion  $\mathcal{E}(p) \rightarrow {}^*\mathcal{E}(p)$  is of class  $C^\infty$ .*

### 2.3 Examples of applications

*Example 4 (Divergence).* The divergence  $D(q\|r) = \mathbb{E}_q \left[ \log \left( \frac{q}{r} \right) \right]$  is  $C^\infty$  jointly in both variables for  $q, r \in \mathcal{E}(p)$ . In fact, in the  $p$  chart,  $u = s_p(q)$ ,  $v = s_p(r)$  gives

$$\begin{aligned} \mathcal{S}_p \times \mathcal{S}_p \ni (u, v) &\mapsto \mathbb{E}_q [u - K_p(u) - v + K_p(v)] \\ &= K_p(v) - K_p(u) - DK_p(u)(v - u). \end{aligned}$$

In the exponential chart the KL divergence is the Bregman divergence of  $K_p$ .

The partial derivative in  $u$  in the direction  $w$  is

$$-DK_p(u)w - D^2K_p(u)(v - u, w) + DK_p(u)w = \text{Cov}_q(u - v, w),$$

hence the direction of steepest increase is  $w \propto (u - v)$ . The partial derivative in  $v$  in the direction  $w$  is

$$DK_p(v)w - DK_p(u)w = \mathbb{E}_r[w] - \mathbb{E}_q[w].$$

This quantity is strictly positive for  $w = v - u \neq 0$  because of the monotonicity of  $K_p$ .

The second partial derivative in  $u$  in the direction  $w_1, w_2$  is

$$D^3K_p(u)(u - v, w_1, w_2) + D^2K_p(u)(w_1, w_2),$$

which reduces on the diagonal  $q = r$  to  $D^2K_p(u)(w_1, w_2) = \text{Cov}_q(w_1, w_2)$ .

The second partial derivative in  $v$  in the direction  $w_1, w_2$  is

$$D^2K_p(v)(w_1, w_2) = \text{Cov}_r(w_1, w_2)$$

which reduces on the diagonal  $q = r$  to  $D^2K_p(u)(w_1, w_2) = \text{Cov}_q(w_1, w_2)$ . Some approaches to Information Geometry are based on the Hessian on the diagonal of a divergence (yoke, potential) e.g., [44], [45].

This is a case of high regularity as we assume the densities  $q$  and  $r$  positive and connected by an open exponential arc. In our framework there is another option, namely to consider  $(q, r) \mapsto D(q\|r)$  as a mapping defined on  ${}^*\mathcal{E}(p) \times \mathcal{E}(p)$ , see the Ex 5 below. Without any regularity assumption one can look for the joint semicontinuity as in [46, Sec 9.4].

*Example 5 (Pythagorean theorem).*

Let  $p \in \mathcal{P}_>$  and  $s_p : \mathcal{E}(p) \rightarrow B_p$  and  $\eta_p : {}^*\mathcal{E}(p) \rightarrow {}^*B_p$  be charts respectively in the exponential and mixture manifold. We can exploit the duality between  ${}^*B_p$  and  $B_{spacep}$  as follows. Let be given densities  $q \in \mathcal{E}(p)$ ,  $u = s_p(q)$  and  $r \in {}^*\mathcal{E}(p) \cap \mathcal{P}_\geq$ . We have

$$\mathbb{E}_p[\eta_p(r) s_p(q)] = \mathbb{E}_p \left[ \left( \frac{r}{p} - 1 \right) u \right] = \mathbb{E}_r[u].$$

As

$$u = \log \left( \frac{q}{p} \right) - \mathbb{E}_p \left[ \log \left( \frac{q}{p} \right) \right] = \log \left( \frac{q}{p} \right) + D(p\|q),$$

we have

$$\mathbb{E}_p [\eta_p(r) s_p(q)] = -D(r \| q) + D(r \| p) + D(p \| q).$$

In particular, if the left side is zero,

$$D(r \| q) = D(r \| p) + D(p \| q),$$

which is the Pitagorean relation in Information Geometry e.g., [47].

*Example 6 (Stochastics).* On a Wiener space  $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geq 0}, \nu)$ , the geometric Brownian motion is  $Z_t = \exp(W_t - t/2)$  is strictly positive and  $\int Z_t d\nu = 1$ . From Ito's formula,  $\Phi = \cosh -1$ ,  $\alpha > 0$ ,

$$\int \Phi(\alpha Z_t) d\nu = \frac{\alpha^2}{2} \int_0^t dt \int \Phi(\alpha Z_s) d\nu.$$

It follows that  $\int \Phi(\alpha Z_t) d\nu = e^{\alpha^2 t/2} - 1$  if finite for all  $\alpha$ , hence  $Z_t \in \mathcal{E}(1)$  and  $s_1(Z_t) = W_t$ . The statistical model  $Z_t$ ,  $t \geq 0$ , is intrinsically non parametric because the vector space generated by the  $W_t$ ,  $t > 0$ , has  $L^2$  closure equal to the full gaussian space  $\left\{ \int f dW : \int_0^t f(s)^2 ds < \infty, t > 0 \right\}$ .

The exponential representation of  $\mathcal{P}_>$  is static, but a dynamic variant has been devised by Imparato [13] and [48].

## 2.4 Exponential families, parameterization

Both the mixture and the exponential manifold are intrinsic structures which are constructed with virtually no assumptions but the positivity of the densities. Specific applications will require special assumptions and special parameterization. We suggest to distinguish the manifold charts from other useful parameterization via definitions of this type. Note that we have defined our statistical manifolds in such a way the coordinate space of each  $p$ -chart is identifiable with the tangent space at  $p$ .

**Definition 6.** *Let  $A$  be an open subset of the exponential manifold  $\mathcal{P}_>$  and let  $\mathcal{M}$  be a manifold. An  $k$ -differentiable mapping  $F: A \rightarrow \mathcal{M}$  is a proper parameterization if the tangent linear form  $T_p F: T_p \mathcal{P}_> \rightarrow TF(p)\mathcal{M}$  is surjective.*

This approach is different from the widely used reverse approach, where a parameterization is a mapping from a parameter's manifold to set of densities. Given a proper parameterization, either the inverse tangent mapping is continuous, in which case the parameterization is actually a chart, or it is possible to pull back the image structure to  $\mathcal{P}_>$ . It is what we have done to build the mixture manifold bases on the mean parameterization  $q \mapsto q/p - 1$ .

A similar discussion applies when dealing with sub-manifolds. According to the general theory of Banach manifolds, a sub-manifold  $\mathcal{M}$  is a subset of  $\mathcal{P}_>$  with a tangent space  $T_p \mathcal{M}$  which splits in  $T_p \mathcal{P}_> = B_p$ , that is closed and has a complement space. Some closed subspaces of  $B_p$  split, e.g. finite dimensional subspaces.

In particular, we have the following definition of exponential family.

**Definition 7.** Let  $V$  be a closed subspace of  $B_p$ . The  $V$ -exponential family is the subset of the maximal exponential family  $\mathcal{E}(p)$  defined by

$$\mathcal{E}_V(p) = \left\{ e^{u - K_p(u)} \cdot p : u \in \mathcal{S}_p \cap V \right\}$$

If a splitting of  $V$  is known, then the exponential family is a sub-manifold of the exponential manifold. The proper definition of a sub-manifold in the framework of mixture/exponential manifold is an important topic to be investigated beyond the partial results in the literature we have summarized here.

### 3 Differential equations

The study of differential or evolution equations fits nicely in the theory of Banach manifolds [6, Ch IV], but would require a generalizations to tackle the technical issue of the non reflexive duality of the couple  $L^\Phi, L^{\Phi*}$ . We review here the language introduced in [17] and mention some examples.

Let  $p(t), t \in I$  open, be a curve in a given maximal exponential family  $\mathcal{E}(p)$ , with  $u(t)$  the  $p$ -coordinate of  $p(t)$ ,  $p(t) = \exp(u(t) - K_p(u(t))) \cdot p$ . If the curve  $I \ni t \mapsto u(t) \in B_p$  is of class  $C^1$  with derivative at  $t$  denoted  $\dot{u}(t)$ , the mapping  $p: I \ni t \mapsto p(t) \in \mathcal{E}(p)$  is differentiable and its derivative is the element of  $T_{p(t)}\mathcal{P}_>$  whose coordinate at  $p$  is  $\dot{u}(t)$ . Its velocity is defined to be the curve  $I \rightarrow T\mathcal{P}_>$  such that the  $p$ -coordinates of  $(p(t), \delta p(t)) \in T_{p(t)}\mathcal{P}_>$  are  $(u(t), \dot{u}(t))$ , that is  $\dot{u}(t) = \delta p(t) - E_p[\delta p(t)]$  and  $\delta p(t) = \dot{u}(t) - E_{p(t)}[\dot{u}(t)]$ . We have

$$\delta p(t) = \dot{u}(t) - E_{p(t)}[\dot{u}(t)] = \frac{d}{dt}(u(t) - K_p(u(t))) = \frac{d}{dt} \log \left( \frac{p(t)}{p} \right) = \frac{\dot{p}(t)}{p(t)}, \quad (4)$$

where the last equality is computed in  $L^{\Phi*}(p)$ .

We can do a similar construction in the mixture manifold. Let  $p(t), t \in I$  open, be a curve in the mixture manifold  $\mathcal{P}_1$ ,  $p(t) = (1 + v(t))p$ . If the curve  $I \ni t \mapsto v(t) \in {}^*B_p$  is of class  $C^1$ , the velocity of the curve is  $I \rightarrow (p, \delta p) \in T\mathcal{P}_1$ , with  $p$ -coordinates of  $\delta p(t)$  equal to  $\dot{v}(t)$ , that is  $\delta p(t) = \frac{p}{p(t)}\dot{v}(t)$ . It follows again  $\delta p(t) = \dot{p}(t)/p(t)$ . The two representations of the velocity equal in the *moving frame*. Note that other representations would be possible and are actually used in the literature e.g.,

$$\frac{\dot{p}(t)}{p(t)} = \frac{\frac{d}{dt} 2\sqrt{p(t)}}{\sqrt{p(t)}},$$

which is a representation to be discussed in Sec. 4.3 below, which is based on the embedding  $\mathcal{P}_> \ni p \mapsto \sqrt{p} \in L^2(\mu)$ .

A *vector field*  $F$  of the exponential manifold  $\mathcal{P}_>$ , is a section of the tangent bundle  $T\mathcal{P}_>$ ,  $F(p) \in T_p\mathcal{P}_>$ , with domain  $p \in \mathcal{D}(F)$ . A curve  $p(t), t \in I$ , is an *integral curve* of  $F$  if  $p(t) \in \mathcal{D}(F)$  and  $\delta p(t) = F(p(t)), t \in I$ . Same definition in the case of the mixture manifold. In the moving frame the differential equation can be written  $\dot{p}(t) = F(p(t))p(t)$ . In the exponential chart at  $p$  we write  $\dot{u}(t) =$

${}^e\mathbb{U}_{p(t)}^p F(p(t))$  together with  $p(t) = e^{u(t)-K_p(u)} \cdot p$ . In the mixture chart at  $p$  we write  $\dot{v}(t) = {}^m\mathbb{U}_{p(t)}^p F(p(t))$  with  $p(t) = (1 + v(t))p$ .

We discuss briefly the existence of solutions. In the exponential chart we have the differential equation

$$\begin{aligned}\dot{u}(t) &= {}^e\mathbb{U}_{p(t)}^p F(p(t)) = F(p(t)) - E_p[F(p(t))], \\ p(t) &= e^{u(t)-K_p(u(t))} \cdot p, \\ u(0) &= 0.\end{aligned}$$

We can use the duality on  $B_p \times {}^*B_p$  to check if the functional

$$\mathbf{F}(u) = F\left(e^{u(t)-K_p(u(t))} \cdot p\right) - E_p\left[F\left(e^{u(t)-K_p(u(t))} \cdot p\right)\right]$$

satisfies a one-sided Lipschitz condition  $\langle \mathbf{F}(u) - \mathbf{F}(v), u - v \rangle - \lambda \langle u - v, u - v \rangle \leq 0$ ,  $\lambda > 0$ . We have

$$\langle \mathbf{F}(u) - \mathbf{F}(v), u - v \rangle = \text{Cov}_p(F(e_p(u) \cdot p) - F(e_p(v) \cdot p), u - v).$$

The uniqueness of the solution follows easily from standard arguments for evolution equations. Proof of existence requires usually extra conditions in order to apply methods from functional analysis. We discuss a set of typical examples.

*Example 7 (One dimensional exponential and mixture families).* Let  $f \in L^\Phi(p)$  and define the vector field  $F$  whose value at each  $q \in \mathcal{E}(p)$  is represented in the frame at  $q$  by  $q \mapsto f - E_q[f]$ . We can assume without restriction that  $f \in B_p$  in which case  $f - E_q[f] = {}^e\mathbb{U}_p^q f$ . The differential equation in the moving frame is  $\dot{p}(t)/p(t) = f - E_{p(t)}[f]$ , with solution  $\log(p(t)) = \log(p(0)) + t(f - E_{p(t)}[f])$ . In the fixed frame at the initial condition  $p(0) = p$  the equation is  $\dot{u}(t) = f - E_p[f]$ , with solution  $u(t) = t(f - E_p[f])$ , hence  $p(t) = \exp(t(f - E_p[f]) - K_p(t(f - E_p[f]))) \cdot p$ . The equation in the mixture manifold,  $f \in {}^*B_p$  is  $\dot{p}(t)/p(t) = {}^m\mathbb{U}_p^{p(t)} f$ , with solution  $p(t) = p(1 + tf)$ . We have constructed here the geodesics of the two manifolds.

*Example 8 (Optimization).* Stochastic relaxation of optimization problems using tools from Information Geometry has been studied in [12], [19], [20], [49], [21], [22]. The expectation of a real function  $F \in L^\Phi(p)$  is an affine function in the mixture chart,  $E_q[F] = E_p\left[F\left(\frac{q}{p} - 1\right)\right] + E_p[F]$ , while in the exponential chart is a function of  $u = s_p(q)$ ,  $\tilde{F}(u) = E_q[F]$ . The equation for the derivative of the cumulant function  $K_p$  gives

$$\tilde{F}(u) = D K_p(u)(F - E_p[F]) + E_p[F],$$

and the derivative of this function in the direction  $v$  is the Hessian of  $K_p$  applied to  $(F - E_p[F]) \otimes v$ :

$$D\Phi(u)v = D^2 K_p(u)(F - E_p[F]) \otimes v = \text{Cov}_q(v, F).$$

The direction of steepest ascent of the expectation is  $F - E_q[F]$ , hence the equation of the flow is  $\delta p(t) = \mathbb{U}_p^{p(t)} F$ , whose solution in the exponential family with canonical statistics  $F$ . In practice, the flow is restricted to an exponential model  $\mathcal{E}_V(p)$ ,  $V \subset B_p$  and the direction of steepest ascent is a projection of  $F$  onto the  $V$ , if it exists.

*Example 9 (Heat equation).* The heat equation  $\frac{\partial}{\partial t} p(t, x) - \frac{\partial^2}{\partial x^2} p(t, x) = 0$ ,  $x \in \mathbb{R}$  for simplicity, is an example of evolution equation in  $T\mathcal{P}_>$  with vector field

$$F(p)(x) = \frac{\frac{\partial^2}{\partial x^2} p(x)}{p(x)}.$$

A proper discussion would require an extension of our construction to Sobolev-Orlicz spaces [37, Ch 8] and the solution would be based on the variational form of the heat equation. For each  $v$  with the proper domain of regularity  $D$

$$E_p[F(p)v] = \int p''(x)v(x) dx = - \int p'(x)v'(x) dx = -E_p\left[\frac{p'}{p}v'\right]$$

from which the weak form of the evolution equation follows,

$$E_{p(t)}[\delta p(t)v] + E_{p(t)}[F_0(p(t))v] = 0, \quad v \in D,$$

where  $F_0(p) = \nabla p/p$  is the vector field associated to the translation model  $p_\theta(x) = p(x - \theta)$ , see e.g., [50].

*Example 10 (Decision theory).* A further interesting example of evolution equation arises in decision theory [51]. For simplicity the sample space is  $\mathbb{R}$ ,  $q \in \mathcal{E}(p)$ ,  $q = e^{u - K_p(u)} \cdot p$  and  $\log q - \log p = u - K_p(u)$ . Assume  $u$  belongs to the Sobolev-Orlicz space

$$W_0^{\Phi,1} = \{u \in L_0^\Phi(p) : \nabla u \in L^\Phi(p)\},$$

where  $\nabla$  denotes the spatial derivative. The following expression is a statistical divergence

$$\begin{aligned} d(p, q) &= \frac{1}{4} E_p \left[ |\nabla \log q - \nabla \log p|^2 \right] \\ &= \frac{1}{4} E_p \left[ |\nabla u|^2 \right]. \end{aligned}$$

For  $u, v_0 \in W^{\Phi,1}$  we have a bilinear form

$$\begin{aligned} (u, v) &\mapsto E_p[\nabla u \nabla v] = \int u_x(x) v_x(x) p(x) dx, \\ &= - \int \nabla(u_x(x) p(x)) v(x) dx, \\ &= - \int (\Delta u(x) p(x) + \nabla u(x) \nabla p(x)) v(x) dx, \\ &= E_p[(-\Delta u - \nabla \log p \nabla u) v], \end{aligned}$$

where  $\Delta$  is the second derivative in space. We have

$$E_p [\nabla u \nabla v] = E_p [F_p(u)v], \quad F_p(u) = -\Delta u - \nabla \log p \nabla u,$$

with  $F_p(u)$  in a proper Sobolev-Orlicz space  ${}^*W_0^{\Phi,1}$ . This provides a classical setting for a weak form of evolution equation. The mapping  $q \mapsto \Delta \log q / \log q$  is represented for  $q = e^{u-K_p(u)} \cdot p$  by

$$u \mapsto \frac{\Delta(u - K_p(u)) + \Delta \log p}{(u - K_p(u)) + \log p}.$$

The mapping  $q \mapsto |\nabla \log q|^2$  is represented by

$$u \mapsto |\nabla u|^2.$$

*Example 11 (Boltzmann equation).* Orlicz spaces as a setting for Boltzmann equation has been recently suggested by [52]. We consider the space-homogeneous Boltzmann equation see e.g., [53]. On the sample space  $(\mathbb{R}^3, d\mathbf{v})$  let  $f_0$  be the standard normal density. For each  $f \in \mathcal{E}(f_0)$  we define the Boltzmann operator to be

$$Q(f)(\mathbf{v}) = \int_{\mathbb{R}^3} \int_{S^2} (f(\mathbf{v} - \mathbf{x}\mathbf{x}'(\mathbf{v} - \mathbf{w}))f(\mathbf{w} + \mathbf{x}\mathbf{x}'(\mathbf{v} - \mathbf{w})) - f(\mathbf{v})f(\mathbf{w})) |\mathbf{x}'(\mathbf{v} - \mathbf{w})| d\mathbf{x} d\mathbf{w},$$

where  $'$  denotes the transposed vector,  $S^2$  is the unit sphere  $\{\mathbf{x} \in \mathbb{R}^3 : \mathbf{x}'\mathbf{x} = 1\}$ ,  $d\mathbf{x}$  is the surface measure on  $S^2$ . The  $\mathbb{R}^{(3+3) \times (3+3)}$  matrix

$$A: \begin{cases} \mathbf{v}_* = \mathbf{v} - \mathbf{x}\mathbf{x}'(\mathbf{v} - \mathbf{w}) = (I - \mathbf{x}\mathbf{x}')\mathbf{v} + \mathbf{x}\mathbf{x}'\mathbf{w}, \\ \mathbf{w}_* = \mathbf{w} + \mathbf{x}\mathbf{x}'(\mathbf{v} - \mathbf{w}) = \mathbf{x}\mathbf{x}'\mathbf{v} + (I - \mathbf{x}\mathbf{x}')\mathbf{w} \end{cases}$$

is such that  $AA$  is the identity on  $\mathbb{R}^6$ , in particular  $\det A = \pm 1$ , and  $\mathbf{x}'(\mathbf{v} - \mathbf{w}) = -\mathbf{x}'(\mathbf{v}_* - \mathbf{w}_*)$ . Hence the measure  $|\mathbf{x}'(\mathbf{v} - \mathbf{w})| d\mathbf{v} d\mathbf{w}$  is invariant under  $A$ . The integral of the Boltzmann operator is zero:

$$\begin{aligned} \int_{\mathbb{R}^3} Q(f)(\mathbf{v}) d\mathbf{v} &= \int_{S^2} \int_{\mathbb{R}^3} \int_{\mathbb{R}^3} (f(\mathbf{v}_*)f(\mathbf{w}_*) - f(\mathbf{v})f(\mathbf{w})) |\mathbf{x}'(\mathbf{v} - \mathbf{w})| d\mathbf{w} d\mathbf{v} d\mathbf{x} = \\ &= \int_{S^2} \int_{\mathbb{R}^3} \int_{\mathbb{R}^3} f(\mathbf{v}_*)f(\mathbf{w}_*) |\mathbf{x}'(\mathbf{v}_* - \mathbf{w}_*)| d\mathbf{w}_* d\mathbf{v}_* d\mathbf{x} - \\ &= \int_{S^2} \int_{\mathbb{R}^3} \int_{\mathbb{R}^3} f(\mathbf{v})f(\mathbf{w}) |\mathbf{x}'(\mathbf{v} - \mathbf{w})| d\mathbf{w} d\mathbf{v} d\mathbf{x} = 0. \end{aligned}$$

Note that  $\mathbf{v}'\mathbf{v} + \mathbf{w}'\mathbf{w} = \mathbf{v}'_*\mathbf{v}_* + \mathbf{w}'_*\mathbf{w}_*$ , hence

$$f_0(\mathbf{v})f_0(\mathbf{w}) = (2\pi)^3 e^{-(1/2)(\mathbf{v}'\mathbf{v} + \mathbf{w}'\mathbf{w})} = f_0(\mathbf{v}_*)f_0(\mathbf{w}_*).$$



If we write  $f(\mathbf{v})/f_0(\mathbf{v}) = g(\mathbf{v})$ , the Boltzmann operator takes the form

$$Q(f)(\mathbf{v}) = f_0(\mathbf{v}) \int_{\mathbb{R}^3} \int_{S^2} (g(\mathbf{v}_*)g(\mathbf{w}_*) - g(\mathbf{v})g(\mathbf{w})) f_0(\mathbf{w}) |\mathbf{x}'(\mathbf{v} - \mathbf{w})| d\mathbf{x} d\mathbf{w} = F_0(f)(\mathbf{w}) f_0(\mathbf{v}),$$

and  $E_{f_0}[F(f)] = 0$  i.e. both  ${}^e\mathbb{U}_{f_0}^f F_0(f)$  and  ${}^m\mathbb{U}_{f_0}^f F_0(f)$  are candidate for a vector field in the exponential manifold.

## 4 The Hilbert bundle

To each positive density  $p \in \mathcal{P}_>$  we attach the Hilbert space of centered square-integrable random variables  $H_p = L_0^2(p)$  in order to define the a vector bundle  $H\mathcal{P}_>$  on the set  $\{(p, u) : p \in \mathcal{P}_>, u \in H_p\}$ . If the densities  $p$  and  $q$  both belong to the same maximal exponential family  $\mathcal{E}$ , then according to Prop. 4 we know the Banach spaces  $L^\Phi(p)$  and  $L^\Phi(q)$ ,  $\Phi = \cosh -1$ , to be equal as sets and have equivalent norms. The subspaces  $B_p$  and  $B_q$ , are continously embedded, respectively, into the Hilbert spaces  $L_0^2(p)$  and  $L_0^2(q)$ . Moreover,  ${}^e\mathbb{U}_p^q : B_p \ni u \mapsto u - E_q[u] \in B_q$  is an isomorphism. Under the same condition  $p, q \in \mathcal{E}$ ,  $L_0^2(p)$  and  $L_0^2(q)$  are continuously embedded, respectively, into the Banach spaces  ${}^*B_p = L_0^{\Phi*}(p)$  and  ${}^*B_q = L_0^{\Phi*}(q)$ , which admit the isomorphism  ${}^m\mathbb{U}_p^q : {}^*B_p \ni u \mapsto \frac{p}{q}u \in {}^*B_q$ . All spaces are embedded subspaces of the space of measurable random variables  $L^0(\mu)$ , see the diagram (5). The isomorphism  $\mathbb{U}_p^q : H_p \rightarrow H_q$  is to be defined in the next sections.

$$\begin{array}{ccccccc} L^\Phi(p) & \longleftarrow & B_p & \longrightarrow & H_p & \longrightarrow & {}^*B_p \longrightarrow L^0(\mu) \\ & & \downarrow {}^e\mathbb{U}_p^q & & \downarrow \mathbb{U}_p^q & & \downarrow {}^m\mathbb{U}_p^q \\ L^\Phi(q) & \longleftarrow & B_q & \longrightarrow & H_q & \longrightarrow & {}^*B_q \end{array} \quad (5)$$

*Example 12.* An example shows that  $p, q \in \mathcal{E}$  does not imply  $L^2(p) = L^2(q)$ : in the exponential model  $p_\theta = \theta e^{-\theta x}$  on  $(\mathbb{R}_>, dx)$ ,  $\theta > 0$ , the random variable  $v(x) = e^{x/4}$  belongs to  $L^2(p_\theta)$  if, and only if,  $\theta > 1/2$ . The equatity is not generally true even locally, unless we restrict ourself to cases were the steepness condition holds. If  $v$  belongs to both  $L^2(p)$  and  $L^2(q)$ , it belong to  $L^2(r)$  with  $r$  in the closed exponential arc between  $p$  and  $q$ , bat the convex function  $B_p \ni u \mapsto \int v^2 e^u p d\mu$  is finite at zero, but could take a  $+\infty$  value on any neighborhood of 0. To construct an example, consider the nonsteep distribution already used in Ex. 3. For the reference measure  $\mu(dx) = (1+x)^{-\frac{3}{2}} dx$ , we rewrite Eq. (1) as

$$\int_0^\infty e^{-\theta x} \mu(dx) = \begin{cases} \sqrt{\theta} e^\theta \Gamma(-\frac{1}{2}, \theta) & \text{if } \theta > 0, \\ 2 & \text{if } \theta = 0, \\ +\infty & \text{if } \theta < 0. \end{cases}$$

The exponential family  $p_\theta \propto e^{-\theta x}$  is defined for  $\theta > 0$ . The random variable  $u(x) = e^{x/2}$  has second moment

$$\int_0^\infty (u(x))^2 p_\theta(x) \mu(dx) \propto \int_0^\infty e^{-(\theta-1)x} \mu(dx),$$

which is finite for  $\theta \geq 1$  and infinite for  $0 < \theta < 1$ .

We are going to show that the  $H_p$ 's are actually isomorphic as Hilbert spaces and that our Hilbert bundle can be viewed as a push-back of the tangent bundle of the unit sphere of  $L^2(\mu)$ . In turn, this construction provides a derivation of the metric connection, see [6, VIII §4]. Connections on statistical manifolds are a key ingredient of Amari's theory [4], while the non parametric version has been done in [9] and [23] in the  $L^p$  case, commutative and non commutative, respectively. Cfr. also the critical discussion in [54]. The construction here is different. In order to have a clear cut distinction between the geometric Hilbert i.e.  $L^2(\mu)$  case and its application to statistical manifolds, we use a bold face notation for points and vectors in the former case.

#### 4.1 The sphere of $L^2(\mu)$

The unit sphere  $S = \{\mathbf{x} \in L^2(\mu) : \int \mathbf{x}^2 d\mu\} = 1$  is a Riemannian manifold with tangent bundle  $TS = \{(\mathbf{x}, \mathbf{u}) : \mathbf{x} \in S, \mathbf{u} \in \{\mathbf{x}\}^\perp\}$  and metric  $g_{\mathbf{x}}(\mathbf{u}, \mathbf{v}) = \int \mathbf{u} \mathbf{v} d\mu = \langle \mathbf{u}, \mathbf{v} \rangle$ . We will use the projection charts  $s_{\mathbf{x}}(\mathbf{y}) = \Pi_{\mathbf{x}} \mathbf{y} = \mathbf{y} - \langle \mathbf{x}, \mathbf{y} \rangle \mathbf{x}$  with domain  $\{\mathbf{y} \in S : \langle \mathbf{x}, \mathbf{y} \rangle > 0\}$  and codomain  $\{\mathbf{u} \in T_{\mathbf{x}}S : \langle \mathbf{u}, \mathbf{u} \rangle < 1\}$ . The patch is  $s_{\mathbf{x}}^{-1}(\mathbf{u}) = \mathbf{u} + \sqrt{1 - \langle \mathbf{u}, \mathbf{u} \rangle} \mathbf{x}$ .

**Proposition 10.** For  $\mathbf{x}, \mathbf{y} \in S$  and  $\mathbf{u} \in T_{\mathbf{x}}S$ , define

$$\mathbb{U}_{\mathbf{x}}^{\mathbf{y}} \mathbf{u} = \mathbf{u} - (1 + \langle \mathbf{x}, \mathbf{y} \rangle)^{-1} \langle \mathbf{u}, \mathbf{y} \rangle (\mathbf{x} + \mathbf{y}). \quad (6)$$

1.  $\mathbb{U}_{\mathbf{x}}^{\mathbf{y}} \mathbf{u} \in T_{\mathbf{y}}S$  and  $\mathbb{U}_{\mathbf{y}}^{\mathbf{x}} \circ \mathbb{U}_{\mathbf{x}}^{\mathbf{y}} \mathbf{u} = \mathbf{u}$ .
2. For  $\mathbf{u}, \mathbf{v} \in T_{\mathbf{x}}S$  the isometric property  $\langle \mathbb{U}_{\mathbf{x}}^{\mathbf{y}} \mathbf{u}, \mathbb{U}_{\mathbf{x}}^{\mathbf{y}} \mathbf{v} \rangle = \langle \mathbf{u}, \mathbf{v} \rangle$  holds, hence

$$g_{\mathbf{y}}(\mathbb{U}_{\mathbf{x}}^{\mathbf{y}} \mathbf{u}, \mathbb{U}_{\mathbf{x}}^{\mathbf{y}} \mathbf{v}) = g_{\mathbf{x}}(\mathbf{u}, \mathbf{v}). \quad (7)$$

*Proof.* The formula for  $\mathbb{U}_{\mathbf{x}}^{\mathbf{y}}$  is obtained by splitting  $\mathbf{u}$  into a component orthogonal to both  $\mathbf{x}$  and  $\mathbf{y}$ , which is left invariant, and rotating the other component in the plane generated by  $\mathbf{x}$  and  $\mathbf{y}$ . Note that  $\mathbb{U}_{\mathbf{z}}^{\mathbf{y}} \circ \mathbb{U}_{\mathbf{y}}^{\mathbf{x}} \neq \mathbb{U}_{\mathbf{z}}^{\mathbf{x}}$  unless  $\mathbf{z}$  belong to the plane generated by  $\mathbf{x}$  and  $\mathbf{y}$ . In a full definition, the transport should be associated to a specific path, but we do not discuss here this point.  $\square$

*Example 13.* Let  $\mu$  the the standard normal distribution and let  $H_n, n = 0, 1, \dots$  the Hermite polynomials:  $H_0(x) = 1, H_1(x) = x, H_2(x) = x^2 - 1, \dots$ , see [55, V.1]. The Hermite polynomials form an orthogonal basis of  $L^2(\mu)$ , hence  $(H_n)_{n \geq 1}$  is an orthogonal basis of  $T_1S = L_0^2(\mu)$ . If  $E(Y^2) = 1$ , the sequence

$$\mathbb{U}_1^Y H_n = H_n - (1 + E(Y))^{-1} E(Y H_n) (1 + Y), \quad n = 1, 2, \dots$$

is an orthogonal basis of  $T_Y S$ .

The isometric affine transport in (6) provides charts for the tangent bundle  $TS$ : given  $\mathbf{x} \in S$ , for each  $\mathbf{y} \in S$ ,  $\langle \mathbf{x}, \mathbf{y} \rangle > 0$ , and  $\mathbf{v} \in T_{\mathbf{y}}$ , then the coordinates of  $(\mathbf{y}, \mathbf{v}) \in T_{\mathbf{y}}S$  are

$$s_{\mathbf{x}}(\mathbf{y}, \mathbf{v}) = (\Pi_{\mathbf{x}}\mathbf{y}, \mathbb{U}_{\mathbf{y}}^{\mathbf{x}}\mathbf{v}) \in T_{\mathbf{x}}S \times T_{\mathbf{x}}S, \quad (8)$$

where  $\Pi_{\mathbf{x}}\mathbf{y} = \mathbf{y} - \langle \mathbf{x}, \mathbf{y} \rangle \mathbf{x}$  is the orthogonal projection on  $T_{\mathbf{x}}S$ . The transition map from  $\mathbf{x}_1$  to  $\mathbf{x}_2$  is

$$\begin{aligned} T_{\mathbf{x}_1}S \times T_{\mathbf{x}_1}S \ni (\mathbf{u}, \mathbf{v}) \mapsto \\ \left( \Pi_{\mathbf{x}_2} \left( \mathbf{u} + \sqrt{1 - \langle \mathbf{u}, \mathbf{u} \rangle^2} \mathbf{x}_1 \right), \mathbb{U}_{\mathbf{x}_1}^{\mathbf{x}_2} \mathbf{v} \right) \in T_{\mathbf{x}_2}S \times T_{\mathbf{x}_2}S. \end{aligned}$$

## 4.2 Covariant derivative on $S$

Let  $F$  be a vector field on the sphere  $S$  and let  $\mathbf{x}(t)$ ,  $t \in I$  be a curve on  $S$ ,  $\mathbf{x}(0) = \mathbf{x}$ . As  $(\mathbf{x}(t), F(\mathbf{x}(t))) \in T_{\mathbf{x}(t)}S$ , in the chart at  $\mathbf{x}$  we have

$$s_{\mathbf{x}}(\mathbf{x}(t), F(\mathbf{x}(t))) = \left( \Pi_{\mathbf{x}}\mathbf{x}(t), \mathbb{U}_{\mathbf{x}(t)}^{\mathbf{x}}F(\mathbf{x}(t)) \right).$$

We assume  $t \mapsto \mathbf{x}(t)$  is differentiable in  $L^2(\mu)$ , so that  $\frac{d}{dt}\Pi_{\mathbf{x}}\mathbf{x}(t)|_{t=0} = \Pi_{\mathbf{x}}\dot{\mathbf{x}}(0) = \dot{\mathbf{x}}(0)$ . The derivative with respect to  $\mathbf{x}$  of  $\Pi_{\mathbf{x}}\mathbf{y}$  in direction  $\mathbf{w}$  is

$$d_{\mathbf{w}}(\mathbf{x} \mapsto \Pi_{\mathbf{x}}\mathbf{y}) = -\langle \mathbf{y}, \mathbf{w} \rangle \mathbf{x} - \langle \mathbf{y}, \mathbf{x} \rangle \mathbf{w}.$$

The derivative with respect to  $\mathbf{x}$  of  $\mathbb{U}_{\mathbf{x}}^{\mathbf{y}}\mathbf{u}$  in direction  $\mathbf{w}$  is

$$d_{\mathbf{w}}(\mathbf{x} \mapsto \mathbb{U}_{\mathbf{x}}^{\mathbf{y}}\mathbf{u}) = (1 + \langle \mathbf{x}, \mathbf{y} \rangle)^{-2} \langle \mathbf{w}, \mathbf{y} \rangle \langle \mathbf{u}, \mathbf{y} \rangle (\mathbf{x} + \mathbf{y}) - (1 + \langle \mathbf{x}, \mathbf{y} \rangle)^{-1} \langle \mathbf{u}, \mathbf{y} \rangle \mathbf{w},$$

so that  $d_{\mathbf{w}}(\mathbf{x} \mapsto \mathbb{U}_{\mathbf{x}}^{\mathbf{y}}\mathbf{u})|_{\mathbf{y}=\mathbf{x}} = 0$  because  $\langle \mathbf{u}, \mathbf{x} \rangle = 0$ .

Let  $F$  be a vector field on the sphere  $S$ , and assume  $F$  is the restriction of a smooth  $L^2(\mu)$ -valued function, defined on a neighborhood of  $S$ , with directional derivative denoted  $d_{\mathbf{w}}F(\mathbf{x})$ . Let  $\mathbf{x}(t)$ ,  $t \in I$  be an  $L^2(\mu)$ -smooth curve on  $S$ ,  $\mathbf{x}(0) = \mathbf{x}$ ,  $\dot{\mathbf{x}}(0) = \mathbf{w} \in T_{\mathbf{x}}S$ . As we want to compute  $\frac{d}{dt}\mathbb{U}_{\mathbf{x}(t)}^{\mathbf{x}}F(\mathbf{x}(t))$ , we write  $\mathbb{U}_{\mathbf{x}(t)}^{\mathbf{x}}F(\mathbf{x}(t)) = \mathbb{U}_{\mathbf{x}(t)}^{\mathbf{x}}\Pi_{\mathbf{x}(t)}F(\mathbf{x}(t))$ , with  $\Pi_{\mathbf{z}}\mathbf{f} = \mathbf{f} - \langle \mathbf{f}, \mathbf{z}(t) \rangle \mathbf{z}(t)$  and  $d_{\mathbf{w}}(\mathbf{z} \mapsto \Pi_{\mathbf{z}}\mathbf{f}) = -\langle \mathbf{f}, \mathbf{w} \rangle \mathbf{z} - \langle \mathbf{f}, \mathbf{z} \rangle \mathbf{w} = -\langle \mathbf{f}, \mathbf{w} \rangle \mathbf{z}$  if  $\mathbf{f} \in T_{\mathbf{x}}S$ . From the previous computations,

$$\begin{aligned} \left. \frac{d}{dt} \mathbb{U}_{\mathbf{x}(t)}^{\mathbf{x}} F(\mathbf{x}(t)) \right|_{t=0} &= \left. \frac{d}{dt} \mathbb{U}_{\mathbf{x}(t)}^{\mathbf{x}} \Pi_{\mathbf{x}(t)} F(\mathbf{x}(t)) \right|_{t=0} \\ &= -\langle F(\mathbf{x}), \mathbf{w} \rangle \mathbf{x} + d_{\mathbf{w}}F(\mathbf{x}) - \langle d_{\mathbf{w}}F(\mathbf{x}), \mathbf{x} \rangle \mathbf{x} \\ &= -\langle F(\mathbf{x}), \mathbf{w} \rangle \mathbf{x} + \Pi_{\mathbf{x}}d_{\mathbf{w}}F(\mathbf{x}). \end{aligned} \quad (9)$$

Let  $F, G, W$  be smooth vector fields on the sphere  $S$ . From (9) we can compute the metric derivative  $\nabla_W F$ , i.e. the unique covariant derivative such that

$$D_W g(F, G) = g(D_W F, G) + g(F, D_W G),$$

see [6, VIII §4].

**Proposition 11.** *The value of the metric derivative  $D_W F$  at  $\mathbf{x} \in S$  is*

$$D_{W(\mathbf{x})} F(\mathbf{x}) = d_{W(\mathbf{x})} F(\mathbf{x}) - \langle d_{W(\mathbf{x})} F(\mathbf{x}), \mathbf{x} \rangle \mathbf{x} = \Pi_{\mathbf{x}} d_{W(\mathbf{x})} F(\mathbf{x}).$$

*Proof.* Let  $\mathbf{x}(t)$ ,  $t \in I$  be a smooth curve on  $S$ , such that  $\dot{\mathbf{x}}(t) = W(\mathbf{x}(t))$ ,  $\mathbf{x}(0) = \mathbf{x}$ ,  $\dot{\mathbf{x}}(0) = W(\mathbf{x}) = \mathbf{w}$ . Note that the first term in (9) is orthogonal to  $T_{\mathbf{x}} S$ .

$$\begin{aligned} \left. \frac{d}{dt} g_{\mathbf{x}(t)}(F(\mathbf{x}(t)), G(\mathbf{x}(t))) \right|_{t=0} &= \left. \frac{d}{dt} \langle \mathbb{U}_{\mathbf{x}(t)}^{\mathbf{x}} F(\mathbf{x}(t)), \mathbb{U}_{\mathbf{x}(t)}^{\mathbf{x}} G(\mathbf{x}(t)) \rangle \right|_{t=0} \\ &= \left\langle \left. \frac{d}{dt} \mathbb{U}_{\mathbf{x}(t)}^{\mathbf{x}} F(\mathbf{x}(t)) \right|_{t=0}, G(\mathbf{x}) \right\rangle + \left\langle F(\mathbf{x}(t)), \left. \frac{d}{dt} \mathbb{U}_{\mathbf{x}(t)}^{\mathbf{x}} G(\mathbf{x}(t)) \right|_{t=0} \right\rangle \\ &= g_{\mathbf{x}}(\Pi_{\mathbf{x}} d_{W(\mathbf{x})} F(\mathbf{x}), G(\mathbf{x})) + g_{\mathbf{x}}(F(\mathbf{x}), \Pi_{\mathbf{x}} d_{W(\mathbf{x})} G(\mathbf{x})). \end{aligned}$$

□

### 4.3 The Hilbert bundle of the exponential manifold

For each density  $p \in \mathcal{P}_{>}$  the linear mapping  $H_p \ni w \mapsto \mathbf{w}\sqrt{p}$  is an isometry onto  $L^2(\mu)$  that maps  $H_p$  onto  $T_{\sqrt{p}} S$ . In fact  $\int (\mathbf{w}\sqrt{p})^2 d\mu = E_p[w^2]$  and  $\langle \mathbf{w}\sqrt{p}, \sqrt{p} \rangle = E_p[w] = 0$ . Viceversa, if  $\mathbf{y} \in T_{\mathbf{x}} S$ , then  $E_p[(\mathbf{x}/\sqrt{p})^2] = \langle \mathbf{x}, \mathbf{x} \rangle$ . In this case the embedding  $\mathcal{P}_{>} \ni p \mapsto \sqrt{p}$  is an injection into the sphere  $S$  of  $L^2(\mu)$  and the sphere is smooth. It is the embedding used in [4] that we discuss here in the framework of Banach manifolds, see the diagram (10). Applications of the non parametric setting are in e.g. [56].

$$\begin{array}{ccccc} T\mathcal{P}_{>} & \longrightarrow & H\mathcal{P}_{>} & \xrightarrow{w \mapsto \mathbf{w}\sqrt{p}} & TS \\ \pi \downarrow & & \pi \downarrow & & \downarrow \pi \\ \mathcal{P}_{>} & \xlongequal{\quad} & \mathcal{P}_{>} & \xrightarrow{p \mapsto \sqrt{p}} & S \longrightarrow L^2(\mu) \end{array} \quad (10)$$

**Proposition 12.** *The mapping  $\mathcal{P}_{>} \ni p \mapsto \sqrt{p} \in S$  is  $C^\infty$  with derivative at  $p$  in the direction  $w \in T_p \mathcal{P}_{>}$  equal to  $\frac{1}{2} \mathbf{w}\sqrt{p} \in T_{\sqrt{p}} S$ .*

*Proof.* Consider the mapping  $\mathcal{P}_{>} \ni p \mapsto \sqrt{p} \in S$  in the charts at  $p$  and  $\sqrt{p}$ , respectively. We go from  $u \in \mathcal{S}_p$  to  $S$  with

$$u \mapsto q = \exp(u - K_p(u)) \cdot p \mapsto \sqrt{q} = \exp\left(\frac{1}{2}u - \frac{1}{2}K_p(u)\right) \sqrt{p}$$

and to  $T_{\sqrt{p}} S$  with

$$\begin{aligned} u \mapsto \sqrt{q} - \int \sqrt{pq} d\mu \sqrt{p} = \\ \left( \exp\left(\frac{1}{2}u - \frac{1}{2}K_p(u)\right) - E_p \left[ \exp\left(\frac{1}{2}u - \frac{1}{2}K_p(u)\right) \right] \right) \sqrt{p}. \end{aligned}$$

The mapping  $u \mapsto e^{u/2}$  is analytic from the open unit ball of  $B_p$  to  $H_p = L_0^2(p)$  according to prop. 1; multiplication by  $\sqrt{p}$  is an isometry of Hilbert spaces. The real function  $u \mapsto K_p(u)$  is infinitely Fréchet differentiable according to Prop. 3. The derivative is computed as

$$d_w(u \mapsto \exp\left(\frac{1}{2}u - \frac{1}{2}K_p(u)\right))\Big|_{u=0} = \exp\left(\frac{1}{2}u - \frac{1}{2}K_p(u)\right)\left(\frac{1}{2}w - d_w K_p(u)\right)\Big|_{u=0} = \frac{1}{2}w,$$

and finally applying the isometry  $\frac{1}{2}w \mapsto \frac{1}{2}w\sqrt{p}$ .  $\square$

For each  $p \in \mathcal{P}_>$  define  $I_p: H_p\mathcal{P}_> \ni u \mapsto \sqrt{p}u \in T_{\sqrt{p}}S$ . We can use the isometry  $I_p$  and the isometry  $\mathbb{U}_{\mathbf{x}}^{\mathbf{y}}$  of Prop. 10 to build an isometry

$$\mathbb{U}_p^q = I_q^{-1} \circ \mathbb{U}_{\sqrt{p}}^{\sqrt{q}} \circ I_p: H_p\mathcal{P}_> \rightarrow H_q\mathcal{P}_>.$$

as in the diagram (11).

$$\begin{array}{ccc} T_{\sqrt{p}}S & \xrightarrow{\mathbb{U}_{\sqrt{p}}^{\sqrt{q}}} & T_{\sqrt{q}}S \\ \uparrow u \mapsto p^{1/2}u & & \downarrow v \mapsto q^{-1/2}v \\ H_p\mathcal{P}_> & \xrightarrow{\mathbb{U}_p^q} & H_q\mathcal{P}_> \end{array} \quad \mathbb{U}_p^q u = q^{-1/2} \mathbb{U}_{\sqrt{p}}^{\sqrt{q}}(p^{1/2}u) \quad (11)$$

Substituting  $\mathbf{u} = \sqrt{p}u$ ,  $\mathbf{x} = \sqrt{p}$ ,  $\mathbf{y} = \sqrt{q}$  in (6),

$$\begin{aligned} \mathbf{u} - (1 + (\mathbf{x} \cdot \mathbf{y}))^{-1} (\mathbf{x} + \mathbf{y})(\mathbf{u} \cdot \mathbf{y}) &= \\ \sqrt{p}u - \left(1 + \int \sqrt{pq} \, d\mu\right)^{-1} (\sqrt{p} + \sqrt{q}) \left(\int \sqrt{pq}u \, d\mu\right) &= \\ \sqrt{p}u - \left(1 + \mathbb{E}_q \left[\sqrt{\frac{p}{q}}\right]\right)^{-1} (\sqrt{p} + \sqrt{q}) \mathbb{E}_q \left[\sqrt{\frac{p}{q}}u\right] \end{aligned}$$

so that

$$\mathbb{U}_p^q u = \sqrt{\frac{p}{q}}u - \left(1 + \mathbb{E}_q \left[\sqrt{\frac{p}{q}}\right]\right)^{-1} \left(1 + \sqrt{\frac{p}{q}}\right) \mathbb{E}_q \left[\sqrt{\frac{p}{q}}u\right] \quad (12)$$

**Proposition 13.**

1. The mapping  $\mathbb{U}_p^q$  of Eq. (12) is an isometry of  $H_p\mathcal{P}_>$  onto  $H_q\mathcal{P}_>$ .
2.  $\mathbb{U}_p^q \circ \mathbb{U}_p^q u = u$ ,  $u \in H_p\mathcal{P}_>$  and  $(\mathbb{U}_p^q)^t = \mathbb{U}_p^q$ .

*Proof.* We double-check the image:

$$\mathbb{E}_q [\mathbb{U}_p^q u] = \mathbb{E}_q \left[ \sqrt{\frac{p}{q}} u \right] - \left( 1 + \mathbb{E}_q \left[ \sqrt{\frac{p}{q}} \right] \right)^{-1} \mathbb{E}_q \left[ 1 + \sqrt{\frac{p}{q}} \right] \mathbb{E}_q \left[ \sqrt{\frac{p}{q}} u \right] = 0.$$

We double-check the isometry:

$$\begin{aligned} \mathbb{E}_q [(\mathbb{U}_p^q u)^2] &= \\ \mathbb{E}_q \left[ \frac{p}{q} u^2 \right] - 2 \left( 1 + \mathbb{E}_q \left[ \sqrt{\frac{p}{q}} \right] \right)^{-1} \mathbb{E}_q \left[ \sqrt{\frac{p}{q}} u \right] \mathbb{E}_q \left[ \sqrt{\frac{p}{q}} u \left( 1 + \sqrt{\frac{p}{q}} \right) \right] \\ &\quad + \left( \left( 1 + \mathbb{E}_q \left[ \sqrt{\frac{p}{q}} \right] \right)^{-1} \mathbb{E}_q \left[ \sqrt{\frac{p}{q}} u \right] \right)^2 \mathbb{E}_q \left[ \left( 1 + \sqrt{\frac{p}{q}} \right)^2 \right] \\ &= \mathbb{E}_p [u^2] - 2 \left( 1 + \mathbb{E}_q \left[ \sqrt{\frac{p}{q}} \right] \right)^{-1} \mathbb{E}_q \left[ \sqrt{\frac{p}{q}} u \right] \mathbb{E}_q \left[ \sqrt{\frac{p}{q}} u \right] \\ &\quad + \left( \left( 1 + \mathbb{E}_q \left[ \sqrt{\frac{p}{q}} \right] \right)^{-1} \mathbb{E}_q \left[ \sqrt{\frac{p}{q}} u \right] \right)^2 \left( 2 + 2\mathbb{E}_q \left[ \sqrt{\frac{p}{q}} \right] \right) \\ &= \mathbb{E}_p [u^2]. \end{aligned}$$

We can now define an atlas on the vector bundle  $H\mathcal{P}_>$  where the coordinates of  $(q, v)$  are defined for  $q \in \mathcal{E}p$  and  $v \in H_q$  as

$$s_p(q, v) = (s_p(q), \mathbb{U}_q^p v) \in \mathcal{S}_p \times H_p.$$

#### 4.4 Metric derivative in the Hilbert bundle

Let  $(p(t), F(t))$ ,  $t \in I$ , be a curve in  $T\mathcal{P}_>$ , i.e.  $p(t) \in \mathcal{P}_>$  and  $F(t) \in T_{p(t)}\mathcal{P}_>$ . Note that  $\mathbb{U}_{p(t)}^p F(t) \in T_p\mathcal{P}_>$ . We write  $p(0) = p$ ,  $\delta p(t) = \frac{d}{dt} \log(p(t))$ ,  $\delta p(0) = w$  and we compute  $\frac{d}{dt} \mathbb{U}_{p(t)}^p F(t) \Big|_{t=0}$  from

$$\begin{aligned} \mathbb{U}_{p(t)}^p F(t) &= \\ \sqrt{\frac{p(t)}{p}} F(t) - \left( 1 + \mathbb{E}_p \left[ \sqrt{\frac{p(t)}{p}} \right] \right)^{-1} \left( 1 + \sqrt{\frac{p(t)}{p}} \right) \mathbb{E}_p \left[ \sqrt{\frac{p(t)}{p}} F(t) \right]. \end{aligned} \quad (13)$$

The derivative of the first term in (13) is

$$\begin{aligned} \frac{d}{dt} \left( \sqrt{\frac{p(t)}{p}} F(t) \right) &= p^{-1/2} \frac{d}{dt} (p(t)^{1/2} F(t)) \\ &= p^{-1/2} \left( \frac{1}{2} p(t)^{-1/2} \dot{p}(t) F(t) + p(t)^{1/2} \dot{F}(t) \right) \\ &= \sqrt{\frac{p(t)}{p}} \left( \dot{F}(t) + \frac{1}{2} F(t) \delta p(t) \right), \end{aligned}$$

so that the derivative of the last factor is

$$\frac{d}{dt} \mathbb{E}_p \left[ \sqrt{\frac{p(t)}{p}} F(t) \right] = \mathbb{E}_p \left[ \sqrt{\frac{p(t)}{p}} \left( \dot{F}(t) + \frac{1}{2} F(t) \delta p(t) \right) \right].$$

Note that  $\mathbb{E}_p \left[ \sqrt{\frac{p(0)}{p}} F(0) \right] = \mathbb{E}_p [F(0)] = 0$ , while

$$\sqrt{\frac{p(0)}{p}} \left( \dot{F}(0) + \frac{1}{2} F(0) \delta p(0) \right) = \dot{F}(0) + \frac{1}{2} F(0) w.$$

In conclusion,

$$\left. \frac{d}{dt} \mathbb{U}_{p(t)}^p F(t) \right|_{t=0} = \dot{F}(0) + \frac{1}{2} F(0) w - \mathbb{E}_p \left[ \dot{F}(0) + \frac{1}{2} F(0) w \right]. \quad (14)$$

Note that in (14) the term  $F(0)w$  is the ordinary product of a random variable  $F(0) \in H_p = L_0^2(p)$  and a random variable  $w \in B_p = L^\Phi(p)$ . In order to define a covariant derivative of the Hilbert bundle  $H\mathcal{P}_>$  we want  $F(0)w \in L^2(p)$ . For example, this would be true if  $F$  were a vector field of the tangent space  $T\mathcal{P}_>$ .

**Definition 8.** Let  $G, F$  be vector fields in  $H\mathcal{P}_>$ , i.e.  $F(p), G(p) \in H_p \mathcal{P}_>$ . We define  $D_G F$  to be the vector field defined by  $D_G F(p) = \left. \frac{d}{dt} \mathbb{U}_{p(t)}^p F(t) \right|_{t=0}$ , where  $p(t)$  is a curve such that  $p(0) = p$  and  $\delta p(0) = G(p)$ .

We conclude this section by summarizing the previous discussion in a statement.

**Proposition 14.** 1.  $D_G F$  in Def. 8 is a covariant derivative.  
2. Let  $F_1, F_2, G$  be vector fields in  $H\mathcal{P}_>$  such that the ordinary products  $GF_1$  and  $GF_2$  are vector fields in  $H\mathcal{P}_>$ . As

$$D_G \mathbb{E}_p [F_1(p) F_2(p)] = \mathbb{E}_p [D_G F_1(p) F_2(p)] + \mathbb{E}_p [F_1(p) D_G F_2(p)],$$

$D_G$  is a metric derivative.

## 5 Deformed exponential manifold

The deformed exponential function is defined in [57, Ch. 10] as the inverse function of a deformed logarithm, with the aim to define a generalisation of entropy and exponential families. To improve consistency with the literature the  $\phi$ -notation in this section differs from what was used in previous sections.

Assume the function  $\phi: \mathbb{R}_> \rightarrow ]0, \phi(\infty)[$  is surjective, increasing and continuous. The  $\phi$ -logarithm is the function

$$\ln_\phi(v) = \int_1^v \frac{dx}{\phi(x)}, \quad v \in \mathbb{R}_>. \quad (15)$$

The  $\phi$ -logarithm, also called *deformed logarithm*,  $\ln_\phi$  is defined on  $\mathbb{R}_>$  and it is strictly increasing, concave and differentiable. Its values range between  $-\int_0^1 \frac{dx}{\phi(x)}$  and  $\int_1^{+\infty} \frac{dx}{\phi(x)}$ . If  $\int_1^{+\infty} \frac{dx}{\phi(x)} = +\infty$ , the range is of  $\ln_\phi$  is  $] -m, +\infty[$  with  $m = \int_0^1 \frac{dx}{\phi(x)} > 0$ . We assume the  $\phi$  function is affinely bounded, so that

$$\lim_{u \rightarrow \infty} \ln_\phi(u) \geq \int_1^{+\infty} \frac{dx}{Ax+B} = +\infty.$$

The  $\phi$ -exponential or *deformed exponential* is the inverse function of  $\ln_\phi$ ,

$$\exp_\phi = \ln_\phi^{-1}: ] -m, +\infty[ \rightarrow \mathbb{R}_>.$$

It is positive, increasing, convex, differentiable.

*Example 14 (Tsallis logarithm and exponential [58]).* The Tsallis logarithm with parameter  $q \in ]0, 1[$  is a deformed logarithm with  $\phi(v) = 1/v^q$ . We have the explicit form

$$\ln_q(v) = \int_1^v \frac{dx}{x^q} = \begin{cases} \frac{1}{1-q} (v^{1-q} - 1), & q \in ]0, 1[, \\ \ln(v), & q = 1. \end{cases}$$

The corresponding exponential is defined for  $q \neq 1$  by

$$\exp_q(u) = (1 + (1-q)u)^{\frac{1}{1-q}}, \quad u > -\frac{1}{1-q} = m.$$

*Example 15 (Kaniadakis exponential and logarithm [59,60]).* The Kaniadakis exponential with parameter  $\kappa \in [0, 1[$  is based on the function

$$\phi(x) = \frac{2x}{x^\kappa + x^{-\kappa}} = \frac{2x^{\kappa+1}}{x^{2\kappa} + 1}, \quad x > 0.$$

This function is linearly bounded,  $\phi(x) \leq x$ ; it is equivalent to  $2x$  for  $x \downarrow 0$  and to  $2x^{1-\kappa}$  for  $x \uparrow +\infty$ .

The deformed logarithm is

$$\ln_\kappa(v) = \int_1^v \frac{x^\kappa + x^{-\kappa}}{2} \frac{1}{x} dx = \begin{cases} \frac{1}{2\kappa} (v^\kappa - v^{-\kappa}) & \text{if } \kappa \neq 0, \\ \ln v & \text{if } \kappa = 0. \end{cases}$$

By checking the differential equation  $y' = \phi(y)$  one shows that the deformed exponential is

$$\exp_\kappa(u) = \exp \left( \int_0^u \frac{dy}{\sqrt{1 + \kappa^2 y^2}} \right) = \begin{cases} (\kappa u + \sqrt{1 + \kappa^2 u^2})^{\frac{1}{\kappa}} & \text{if } \kappa \neq 0, \\ \exp u & \text{if } \kappa = 0. \end{cases}$$



*Example 16 (Nigel J. Newton exponential [61]).* The function

$$\phi(x) = \frac{x}{x+1}, \quad x > 0,$$

has image  $\text{Range}(\phi) = ]0, 1[$ , is bounded by 1 and is linearly bounded by  $x$ . The  $\phi$ -logarithm is

$$\ln_\phi(u) = \int_1^u \frac{x+1}{x} dx = u - 1 + \ln u.$$

### 5.1 Model space

Here we built our model spaces according to the proposal of Vigelis and Cavalcante [62], if  $\phi(x) = x$ . Note that the  $\phi$ -exponential notation is not used in [62], where  $\phi$  denotes a class of deformed exponential function larger than the one used here.

**Definition 9.** For each  $p \in \mathcal{P}_>$ , we define the vector space

$$L^{\phi,p}(\mu) = \left\{ u : \exists \alpha > 0 \int \exp_\phi(\alpha |u| + \ln_\phi(p)) d\mu < +\infty \right\}$$

The vector space property is a consequence of the convexity of the  $\phi$ -exponential. Under our assumptions on  $\mu$  (locally finite) and on  $\phi$  (affinely bounded) such vector spaces are not empty. Bounded random variables whose support has finite  $\mu$  measure belong to each  $L^{\phi,p}(\mu)$ .

**Proposition 15.** The following statements are equivalent to  $u \in L^{\phi,p}(\mu)$ .

1. For all real  $\theta$  in a neighborhood of 0

$$\int \exp_\phi(\theta u + \ln_\phi(p)) d\mu < +\infty.$$

2. For some positive  $\alpha$

$$\frac{1}{2} \left( \int \exp_\phi(\alpha u + \ln_\phi(p)) d\mu + \int \exp_\phi(\alpha u + \ln_\phi(p)) d\mu \right) < +\infty.$$

For each  $u \in L^{\phi,p}(\mu)$  the set

$$\left\{ r > 0 : \int \exp_\phi(r^{-1} |u| + \ln_\phi(p)) d\mu \leq 2 \right\}$$

is an infinite interval of the positive real line. Its left end is the norm of  $u$ .

**Proposition 16.** The vector space  $L^{\phi,p}(\mu)$  is a Banach space for the norm

$$\|u\|_{\phi,p} = \inf \left\{ r > 0 : \int \exp_\phi(r^{-1} |u| + \ln_\phi(p)) d\mu \leq 2 \right\}.$$

The importance of escort measures in deformed exponential families have been pointed out in [57, §10.5].

**Definition 10.**

1. The measure  $\phi(p) \cdot \mu$  is equivalent to  $\mu$  and it is called escort measure of  $p$ .
2. If  $\mu$  is a finite measure, or if  $\phi$  is linearly bounded, then the escort measure is finite. In such a case, its formalized density is called the escort density of  $p$ . We write

$$E_{\phi,p}[u] = \frac{\int u\phi(p)d\mu}{\int \phi(p)d\mu}.$$

**Proposition 17.**

1. The Banach space  $L^{\phi,p}(\mu)$  is contained in the Lebesgue space  $L^1(\phi(p) \cdot \mu)$  and the injection is non-expansive,

$$\int |u| \phi(p) d\mu \leq \|u\|_{\phi,p}.$$

2.  $L^{\phi,p}(\mu)$  is a dense subspace of  $L^{\phi,p}(\mu)$ .
3. The space

$$T_{\phi,p} = \left\{ u \in L^{\phi,p}(\mu) : \int u\phi(p)d\mu = 0 \right\}$$

is a closed subspace of  $L^{\phi,p}(\mu)$  hence a Banach space for the induced norm.

*Proof.* From the convexity of  $\exp_\phi$

$$\exp_\phi(r^{-1}|v| + \ln_\phi(p)) \geq p + \phi(p)r^{-1}|v|,$$

hence, if  $r > \|v\|_{\phi,p}$ ,

$$2 \geq \int p d\mu + r^{-1} \int |v| \phi(p) d\mu.$$

It follows that  $v \in L^1(\phi(p) \cdot \mu)$  and  $\int u\phi(p)d\mu$  is well defined. Moreover,

$$\left| \int |u| \phi(p) d\mu \right| \leq r$$

for all  $r > \|v\|_{\phi,p}$ .

**Proposition 18.**

$$L^{\phi,q} \subset L^{\phi,p} \iff (\ln_\phi(q) - \ln_\phi(p)) \in L^{\phi,p}(\mu)$$

Given  $p, q \in \mathcal{P}_{>}$ , convexity implies for  $t \in [0, 1]$

$$\begin{aligned} \int \exp_{\phi}(t(\ln_{\phi}(q) - \ln_{\phi}(p)) + \ln_{\phi}(p)) d\mu = \\ \int \exp_{\phi}((1-t)\ln_{\phi}(p) + t\ln_{\phi}(q)) d\mu \leq 1 < +\infty. \end{aligned}$$

The inequality shows that any two distinct positive probability densities are connected by a closed arc of densities with total mass strictly smaller than one. The arc extends to negative values  $t \in ]a, 1] \supset [0, 1]$  with densities of finite mass if, and only if,  $\ln_{\phi}(q) - \ln_{\phi}(p) \in L^{\phi,p}(\mu)$ . This leads to the following definition.

**Definition 11.** *The densities  $p, q \in \mathcal{P}_{>}$  are  $\phi$ -connected by an open arc if  $]a, b[ \supset [0, 1]$  and*

$$\int \exp_{\phi}((1-t)\ln_{\phi}(p) + t\ln_{\phi}(q)) d\mu < +\infty, \quad t \in ]a, b[. \quad (16)$$

**Proposition 19.** *The relation in Definition 11 is an equivalence relation.*

*Proof.* We show transitivity for densities  $p, q, r \in \text{posdensities}$  with  $p, q$   $\phi$ -connected on  $]a, b[$  and  $q, r$  on  $]c, d[$ . The convex function

$$(\alpha, \beta) \mapsto \int \exp_{\phi}(\alpha \ln_{\phi}(p) + (1 - \alpha - \beta) \ln_{\phi}(q) + \beta \ln_{\phi}(r)) d\mu$$

is finite on the interior of the convex hull of  $(1-b, 0), (1-a, 0), (0, c), (0, d)$ , hence  $p, r$  are  $\phi$ -connected on  $] -bd/(b+d-1), b(1-c)/(b-c)[$ .

**Proposition 20.**  *$L^{\phi,q}(\mu) = L^{\phi,p}(\mu)$  if, and only if,  $p$  and  $q$  are  $\phi$ -connected by an open arc.*

*Proof.* Follows from the symmetric inclusion.

If  $p_0, p_1$  are  $\phi$ -connected by an open interval we can define a statistical model by

$$p(t) = \frac{\exp_{\phi}((1-t)\ln_{\phi}(p_0) + t\ln_{\phi}(p_1))}{\int \exp_{\phi}((1-t)\ln_{\phi}(p_0) + t\ln_{\phi}(p_1)) d\mu}.$$

If we define

$$u = \ln_{\phi}(p_1) - \ln_{\phi}(p_0) - \int (\ln_{\phi}(p_1) - \ln_{\phi}(p_0)) \phi(p) d\mu,$$

we can consider the expression for the density

$$\exp_{\phi}(tu - \psi(t) + \ln_{\phi}(p_0)) = p(t),$$

that is

$$tu - \psi(t) + \ln_{\phi}(p_0) = \ln_{\phi}(p(t)),$$

which gives

$$\psi(t) = \int \phi(p_0) (\ln_{\phi}(p(t)) - \ln_{\phi}(p_0)) d\mu.$$

## 5.2 Generating functionals

**Definition 12.** *The convex function*

$$L^{\phi,p}(\mu) \ni v \mapsto \int \exp_{\phi}(v + \ln_{\phi}(p)) d\mu \in [1, +\infty]$$

is the  $\phi$ -moment generating functional at  $p$ .

**Proposition 21.**

1. *The proper domain of the moment generating functional contains the open ball of  $L^{\phi,p}(\mu)$ . The interior  $\mathcal{S}_p$  of the proper domain of the moment generating functional in a nonempty convex open set.*
2. *For each  $u \in \mathcal{S}_p$  the mapping*

$$L^{\phi,p}(\mu) \ni v \mapsto \int \phi(\exp_{\phi}(u + \ln_{\phi}(p))) v d\mu$$

*is linear and continuous.*

3. *The moment generating functional is lower semicontinuous.*

**Proposition 22.**

1. *For each  $u \in B_p$  such that*

$$\int \exp_{\phi}(u + \ln_{\phi}(p)) d\mu < +\infty \tag{17}$$

*there exists a unique nonnegative constant  $K_p(u)$ , positive if  $u \neq 0$ , such that*

$$q = \exp_{\phi}(u - K_p(u) + \ln_{\phi}(p))$$

*is a probability density.*

2. *In particular, (17) holds if  $\|u\|_{\phi,p} < 1$ .*
3. *If  $q$  is a positive density such that  $\ln_{\phi}(p) - \ln_{\phi}(q) \in L^{\phi,p}(p)$ , then*

$$u = \ln_{\phi}(q) - \ln_{\phi}(p) - \int \phi(p) (\ln_{\phi}(q) - \ln_{\phi}(p)) d\mu \tag{18}$$

*satisfies (17) and*

$$K_p(u) = E_{\phi(p)} [\ln_{\phi}(p) - \ln_{\phi}(q)]. \tag{19}$$

**Definition 13.** *The  $\phi$ -cumulant generating function is the function  $K_p: T_p \rightarrow [0, +\infty]$  defined by*

$$K_p(u) = \sup \left\{ k \geq 0 : \int \exp_{\phi}(u - k + \ln_{\phi}(p)) d\mu \leq 1 \right\}.$$

**Proposition 23.**

1.  $K_p$  is null at 0, extended nonnegative and finite on the set

$$\left\{ u \in T_p : \int \exp_{\phi}(-u + \ln_{\phi}(p)) d\mu < +\infty \right\}.$$

2. If  $K_p(u_0) < +\infty$ , then for all  $u \in T_p$

$$K(u) \geq K(u_0) + E_{\Phi(p)}[u - u_0].$$

Therefore,  $K_p$  is strictly convex, proper, lower semicontinuous.

3. If  $\ln_{\phi}(q) - \ln_{\phi}(p) \in L^{\phi,p}(\mu)$  then its  $\phi(p)$ -centered random variable  $u$  is in the proper domain of  $K_p$  and viceversa.
4. The interior  $\mathcal{S}$  of the proper domain of  $K_p$  is a convex open set that contains the unit open ball  $\left\{ u \in L^{\phi,p}(\mu) : \|u\|_{\phi,p} < 1 \right\}$ . On this set  $K_p$  is differentiable and the derivative of  $K_p$  at  $u$  in the direction  $v$  is

$$DK_p(u)v = E_{\phi(q)}[u].$$

At this point, we consider that almost all elements for the construction of a deformed exponential manifold along the same lines are available.

## 6 Final remarks

In this paper we have reviewed a specific track to the development of Information Geometry, i.e. the construction of a classical Banach manifold structure. This is done by developing in the natural way the original suggestion by B. Efron to look at the larger exponential structure. A non parametric approach is justified by the importance of applications essentially non parametric and by the neat mathematics involved. Other options are present in the literature, the most classical and most successful is based on the embedding  $p \mapsto \sqrt{p}$  from the probability density simplex into the  $L^2$ . Variants of this basic Hilbert embedding were used, see e.g. [63]. S. Eguchi [64] has  $L_0^2$  representation based on the mapping  $u \mapsto \frac{1}{2} - \frac{1}{2}\sigma^2(u) + \frac{1}{2}(1-u)^2 = g$  which is defined on the unit  $L_0^2$  open ball and takes its values in the set of densities which are bounded below by a positive constant. The duality between the exponential and mixture manifold, could lead to an other intermediate option, i.e. to define a manifold where the regularity of the maps is defined in some weak sense. See also the discussion in [65]. Another option uses a non-exponential representation of positive densities through the so-called deformed exponentials.

## References

1. Amari, S.i.: Differential geometry of curved exponential families—curvatures and information loss. Ann. Statist. **10**(2) (1982) 357–385
2. Amari, S.: Differential-geometrical methods in statistics. Volume 28 of Lecture Notes in Statistics. Springer-Verlag, New York (1985)

3. Amari, S.: Differential geometrical theory of statistics. In: Differential geometry in statistical inference. Institute of Mathematical Statistics Lecture Notes—Monograph Series, 10. Institute of Mathematical Statistics, Hayward, CA (1987) 19–94
4. Amari, S., Nagaoka, H.: Methods of information geometry. American Mathematical Society, Providence, RI (2000) Translated from the 1993 Japanese original by Daishi Harada.
5. Bourbaki, N.: Variétés différentielles et analytiques. Fascicule de résultats / Paragraphes 1 à 7. Number XXXIII in *Éléments de mathématiques*. Hermann, Paris (1971)
6. Lang, S.: Differential and Riemannian manifolds. Third edn. Volume 160 of Graduate Texts in Mathematics. Springer-Verlag, New York (1995)
7. Pistone, G., Sempì, C.: An infinite-dimensional geometric structure on the space of all the probability measures equivalent to a given one. *Ann. Statist.* **23**(5) (October 1995) 1543–1561
8. Pistone, G., Rogantin, M.: The exponential statistical manifold: mean parameters, orthogonality and space transformations. *Bernoulli* **5**(4) (August 1999) 721–760
9. Gibilisco, P., Pistone, G.: Connections on non-parametric statistical manifolds by Orlicz space geometry. *IDAQP* **1**(2) (1998) 325–347
10. Cena, A.: Geometric structures on the non-parametric statistical manifold. PhD thesis, Dottorato in Matematica, Università di Milano (2002)
11. Cena, A., Pistone, G.: Exponential statistical manifold. *Ann. Inst. Statist. Math.* **59**(1) (2007) 27–56
12. Malagò, L., Matteucci, M., Dal Seno, B.: An information geometry perspective on estimation of distribution algorithms: boundary analysis. In: GECCO '08: Proceedings of the 2008 GECCO conference companion on Genetic and evolutionary computation, New York, NY, USA, ACM (2008) 2081–2088
13. Imparato, D.: Exponential models and Fisher information. Geometry and applications. PhD thesis, DIMAT Politecnico di Torino (2008)
14. Brigo, D., Pistone, G.: Projecting the Fokker-Planck equation onto a finite dimensional exponential family. *arXiv:0901.1308* (2009)
15. Malagò, L., Pistone, G.: A note on the border of an exponential family. *arXiv:1012.0637v1* (2010)
16. Pistone, G.:  $\kappa$ -exponential models from the geometrical viewpoint. *The European Physical Journal B Condensed Matter Physics* **71**(1) (July I 2009) 29–37
17. Pistone, G.: Algebraic varieties vs. differentiable manifolds in statistical models. In Gibilisco, P., Riccomagno, E., Rogantin, M., Wynn, H.P., eds.: *Algebraic and Geometric Methods in Statistics*. Cambridge University Press (2009) 339–363
18. Imparato, D., Trivellato, B.: Geometry of extended exponential models. In: *Algebraic and geometric methods in statistics*. Cambridge Univ. Press, Cambridge (2010) 307–326
19. Malagò, L., Matteucci, M., Pistone, G.: Towards the geometry of estimation of distribution algorithms based on the exponential family. In: Proceedings of the 11th workshop on Foundations of genetic algorithms. FOGA '11, New York, NY, USA, ACM (2011) 230–242
20. Malagò, L., Matteucci, M., Pistone, G.: Stochastic natural gradient descent by estimation of empirical covariances. In: *Evolutionary Computation (CEC), 2011 IEEE Congress on.* (2011) 949–956
21. Malagò, L.: On the geometry of optimization based on the exponential family relaxation. PhD thesis, Politecnico di Milano (2012)

22. Malagò, L., Matteucci, M., Pistone, G.: Natural gradient, fitness modelling and model selection: A unifying perspective. Paper #1747 IEEE Congress on Evolutionary Computation IEEE CEC 2013 June 20–23 Cancún México (2013)
23. Gibilisco, P., Isola, T.: Connections on statistical manifolds of density operators by geometry of noncommutative  $L^p$ -spaces. *Infin. Dimens. Anal. Quantum Probab. Relat. Top.* **2**(1) (1999) 169–178
24. Jenčová, A.: A construction of a nonparametric quantum information manifold. *J. Funct. Anal.* **239**(1) (2006) 1–20
25. Gibilisco, P., Riccomagno, E., Rogantin, M.P., Wynn, H.P., eds.: Algebraic and geometric methods in statistics. Cambridge University Press, Cambridge (2010)
26. Efron, B.: Defining the curvature of a statistical problem (with applications to second order efficiency). *Ann. Statist.* **3**(6) (1975) 1189–1242 With a discussion by C. R. Rao, Don A. Pierce, D. R. Cox, D. V. Lindley, Lucien LeCam, J. K. Ghosh, J. Pfanzagl, Niels Keiding, A. P. Dawid, Jim Reeds and with a reply by the author.
27. Barndorff-Nielsen, O.E.: Information and Exponential Families in Statistical Theory. John Wiley & Sons, New York (1978)
28. Brown, L.D.: Fundamentals of statistical exponential families with applications in statistical decision theory. Number 9 in IMS Lecture Notes. Monograph Series. Institute of Mathematical Statistics, Hayward, CA (1986)
29. Letac, G.: Lectures on natural exponential families and their variance functions. Volume 50 of Monografias de Matemática [Mathematical Monographs]. Instituto de Matemática Pura e Aplicada (IMPA), Rio de Janeiro (1992)
30. Dawid, A.P.: Discussion of a paper by Bradley Efron. *Ann. Statist.* **3**(6) (1975) 1231–1234
31. Dawid, A.P.: Further comments on: “Some comments on a paper by Bradley Efron” (*Ann. Statist.* **3** (1975), 1189–1242). *Ann. Statist.* **5**(6) (1977) 1249
32. Gzyl, H., Recht, L.: A geometry on the space of probabilities. I. The finite dimensional case. *Rev. Mat. Iberoam.* **22**(2) (2006) 545–558
33. Gzyl, H., Recht, L.: A geometry on the space of probabilities. II. Projective spaces and exponential families. *Rev. Mat. Iberoam.* **22**(3) (2006) 833–849
34. Krasnosel’skii, M.A., Rutickii, Y.B.: Convex Functions and Orlicz Spaces. Noordhoff, Groningen (1961) Russian original: (1958) Fizmatgiz, Moskva.
35. Musielak, J.: Orlicz spaces and modular spaces. Volume 1034 of Lecture Notes in Mathematics. Springer-Verlag, Berlin (1983)
36. Rao, M.M., Ren, Z.D.: Applications of Orlicz spaces. Volume 250 of Monographs and Textbooks in Pure and Applied Mathematics. Marcel Dekker Inc., New York (2002)
37. Adams, R.A., Fournier, J.J.F.: Sobolev spaces. Second edn. Volume 140 of Pure and Applied Mathematics (Amsterdam). Elsevier/Academic Press, Amsterdam (2003)
38. Gallavotti, G.: Statistical mechanics: A short treatise. Texts and Monographs in Physics. Springer-Verlag, Berlin (1999)
39. Boros, E., Hammer, P.L.: Pseudo-Boolean optimization. *Discrete Appl. Math.* **123**(1-3) (2002) 155–225 Workshop on Discrete Optimization, DO’99 (Piscataway, NJ).
40. Grasselli, M.R.: Dual connections in nonparametric classical information geometry. Technical Report math-ph/0104031 v1, arXiv (2001)
41. Appell, J., Zabrejko, P.P.: Nonlinear superposition operators. Volume 95 of Cambridge Tracts in Mathematics. Cambridge University Press, Cambridge (1990)

42. Upmeyer, H.: Symmetric Banach manifolds and Jordan  $C^*$ -algebras. Volume 104 of North-Holland Mathematics Studies. North-Holland Publishing Co., Amsterdam (1985) *Notas de Matemática [Mathematical Notes]*, 96.
43. Ambrosetti, A., Prodi, G.: A primer of nonlinear analysis. Volume 34 of Cambridge Studies in Advanced Mathematics. Cambridge University Press, Cambridge (1993)
44. Barndorff-Nielsen, O.E., Jupp, P.E.: Statistics, yokes and symplectic geometry. *Ann. Fac. Sci. Toulouse Math.* (6) **6**(3) (1997) 389–427
45. Shima, H.: The geometry of Hessian structures. World Scientific Publishing Co. Pte. Ltd., Hackensack, NJ (2007)
46. Ambrosio, L., Gigli, N., Savaré, G.: Gradient flows in metric spaces and in the space of probability measures. Second edn. Lectures in Mathematics ETH Zürich. Birkhäuser Verlag, Basel (2008)
47. Csiszár, I., Matúš, F.: Information projections revisited. *IEEE Trans. Inform. Theory* **49**(6) (2003) 1474–1490
48. Santacroce, M., Siri, P., Trivellato, B.: A dynamic approach to exponential statistical manifolds. In progress (2013)
49. Arnold, L., Auger, A., Hansen, N., Ollivier, Y.: Information-Geometric Optimization Algorithms: A Unifying Picture via Invariance Principles. [arXiv:1106.3708](https://arxiv.org/abs/1106.3708) (2011)
50. Otto, F.: The geometry of dissipative evolution equations: the porous medium equation. *Comm. Partial Differential Equations* **26**(1-2) (2001) 101–174
51. Parry, M., Dawid, A.P., Lauritzen, S.: Proper local scoring rules. *Ann. Statist.* **40**(1) (2012) 561–592
52. Majewski, W.A., Labuschagne, L.E.: On applications of orlicz spaces to statistical physics. [arXiv:1302.3460](https://arxiv.org/abs/1302.3460) (2013)
53. Villani, C.: A review of mathematical topics in collisional kinetic theory. In: *Handbook of mathematical fluid dynamics*, Vol. I. North-Holland, Amsterdam (2002) 71–305
54. Grasselli, M.R.: Dual connections in nonparametric classical information geometry. *Ann. Inst. Statist. Math.* **62**(5) (2010) 873–896
55. Malliavin, P.: Integration and probability. Volume 157 of Graduate Texts in Mathematics. Springer-Verlag, New York (1995) With the collaboration of Hélène Airault, Leslie Kay and Gérard Letac, Edited and translated from the French by Kay, With a foreword by Mark Pinsky.
56. Brigo, D., Hanzon, B., Le Gland, F.: Approximate nonlinear filtering by projection on exponential manifolds of densities. *Bernoulli* **5**(3) (1999) 495–534
57. Naudts, J.: Generalised Thermostatistics. Springer (2011)
58. Tsallis, C.: Possible generalization of Boltzmann-Gibbs statistics. *J. Statist. Phys.* **52**(1-2) (1988) 479–487
59. Kaniadakis, G.: Statistical mechanics in the context of special relativity. *Physical Review E* **66** (2002) 056125 1–17
60. Kaniadakis, G.: Statistical mechanics in the context of special relativity. ii. *Phys. Rev. E* **72**(3) (Sep 2005) 036108
61. Newton, N.J.: An infinite-dimensional statistical manifold modelled on Hilbert space. *J. Funct. Anal.* **263**(6) (2012) 1661–1681
62. Vigelis, R.F., Cavalcante, C.C.: On the  $\phi$ -family of probability distributions. *Journal of Theoretical Probability* (2011) Online First.
63. Burdet, G., Combe, P., Nencka, H.: On real Hilbertian info-manifolds. In: *Disordered and complex systems* (London, 2000). Volume 553 of AIP Conf. Proc. Amer. Inst. Phys., Melville, NY (2001) 153–158



- 64. Eguchi, S.: Tubular modelling approach to statistical method for observational studies. 2nd International Symposium on Information Geometry and its Applications Tokyo (Dec 12-16 2005)
- 65. Zhang, J., Hästö, P.: Statistical manifold as an affine space: a functional equation approach. *Journal of Mathematical Psychology* **50**(1) (2006) 60–65