Deformed Exponential Bundle: The Linear Growth Case

Luigi Montrucchio[1] and Giovanni Pistone[2(✉)]

[1] Collegio Carlo Alberto, Torino, Italy
luigi.montrucchio@unito.it
[2] de Castro Statistics, Collegio Carlo Alberto, Torino, Italy
giovanni.pistone@carloalberto.org
http://www.giannidiorestino.it

Abstract. Vigelis and Cavalcante extended the Naudts' deformed exponential families to a generic reference density. Here, the special case of Newton's deformed logarithm is used to construct an Hilbert statistical bundle for an infinite dimensional class of probability densities.

1 Introduction

Let $\mathcal{P}$ be a family of positive probability densities on the probability space $(\mathbb{X}, \mathcal{X}, \mu)$. At each $p \in \mathcal{P}$ we have the Hilbert space of square-integrable random variables $L^2(p \cdot \mu)$ so that we can define the *Hilbert bundle* consisting of $\mathcal{P}$ with linear fibers $L^2(p \cdot \mu)$. Such a bundle supports most of the structure of Information Geometry, cf. [1] and the non-parametric version in [6,7].

If $\mathcal{P}$ is an exponential manifold, there exists a splitting of each fiber $L(p \cdot \mu) = H_p \oplus H_p^{\perp}$, such that H_p is equal or contains as a dense subset, the tangent space of the manifold at p. Moreover, the geometry on $\mathcal{P}$ is affine and, as a consequence, there are natural transport mappings on the Hilbert bundle.

We shall study a similar set-up when the manifold is defined by charts based on mapping other than the exponential, while retaining an affine structure, see e.g. [10]. Here, we use $p = \exp_A(v)$, where $\exp_A$ is exponential-like function with linear growth at $+\infty$. In such a case, the Hilbert bundle has fibers which are all sub-spaces of the same $L^2(\mu)$ space.

The formalism of deformed exponentials by Naudts [4] is reviewed and adapted in Sect. 2. The following Sect. 3 is devoted to the adaptation of that formalism to the non-parametric case. Our construction is based on the work of Vigelis and Cavalcante [9], and we add a few more details about the infinite-dimensional case. Section 4 discusses the construction of the Hilbert statistical bundle in our case.

2 Background

We recall a special case of a nice and useful formalism introduced by Naudts [4]. Let $A\colon [0, +\infty[\to [0, 1[$ be an increasing, concave and differentiable function

F. Nielsen and F. Barbaresco (Eds.): GSI 2017, LNCS 10589, pp. 239–246, 2017.
https://doi.org/10.1007/978-3-319-68445-1_28

with $A(0) = 0$, $A(+\infty) = 1$ and $A'(0+) = 1$. We focus on the case $A(x) = 1 - 1/(1+x) = x/(1+x)$ that has been firstly discussed by Newton [5]. The deformed A-logarithm is the function $\log_A(x) = \int_1^x A(\xi)^{-1}\, d\xi = x - 1 + \log x$, $x \in]0, +\infty[$. The deformed A-exponential is $\exp_A = \log_A^{-1}$ which turns out to be the solution to the Cauchy problem $e'(y) = A(e(y)) = 1 + 1/(1+e(y))$, $e(0) = 1$.

In the spirit of [8,9] we consider the curve in the space of positive measures on $(\mathbb{X}, \mathcal{X})$ given by $t \mapsto \mu_t = \exp_A(tu + \log_A p)\cdot\mu$, where $u \in L^2(\mu)$. As $\exp_A(a+b) \leq a^+ + \exp_A(b)$, each μ_t is a finite measure, $\mu_t(\mathbb{X}) \leq \int (tu)^+ \, d\mu + 1$, with $\mu_0 = p \cdot \mu$. The curve is actually continuous and differentiable because the pointwise derivative of the density $p_t = \exp_A(tu + \log_A(p))$ is $\dot{p}_t = A(p_t)u$ so that $|\dot{p}_t| \leq |u|$. In conclusion $\mu_0 = p$ and $\dot{\mu}_0 = u$.

Notice that there are two ways to normalize the density p_t, either dividing by a normalizing constant $Z(t)$ to get the statistical model $t \mapsto \exp_A(tu - \log_A p)/Z(t)$ or, subtracting a constant $\psi(t)$ from the argument to get the model $t \mapsto \exp_A(tu - \psi(t) + \log_A(p))$. In the standard exponential case the two methods lead to the same result, which is not the case for deformed exponentials where $\exp_A(\alpha + \beta) \neq \exp_A(\alpha)\exp_A(\beta)$. We choose in the present paper the latter option.

3 Deformed Exponential Family Based on $\exp_A$

Here we use the ideas of [4,8,9] to construct deformed non-parametric exponential families. Recall that we are given: the measure space $(\mathbb{X}, \mathcal{X}, \mu)$; the set $\mathcal{P}$ of probability densities; the function $A(x) = x/(1+x)$. Throughout this section, the density $p \in \mathcal{P}$ will be fixed.

Proposition 1. *1. The mapping $L^1(\mu) \ni u \mapsto \exp_A(u + \log_A p) \in L^1(\mu)$ has full domain and is 1-Lipschitz. Consequently, the mapping*

$$u \mapsto \int g \exp_A(u + \log_A p)\, d\mu$$

is $\|g\|_\infty$-Lipschitz for each bounded function g.

2. For each $u \in L^1(\mu)$ there exists a unique constant $K(u) \in \mathbb{R}$ such that $\exp_A(u - K(u) + \log_A p)\cdot\mu$ is a probability.

3. It holds $K(u) = u$ if, and only if, u is constant. In such a case,

$$\exp_A(u - K(u) + \log_A p)\cdot\mu = p\cdot\mu \ .$$

Otherwise, $\exp_A(u - K(u) + \log_A p)\cdot\mu \neq p\cdot\mu$.

4. A density q takes the form $q = \exp_A(u - K(u) + \log_A p)$, with $u \in L^1(\mu)$ if, and only if, $\log_A q - \log_A p \in L^1(\mu)$.

5. If $u, v \in L^1(\mu)$

$$\exp_A(u - K(u) + \log_A p) = \exp_A(v - K(v) + \log_A p) \ ,$$

then $u - v$ is constant.

6. *The functional* $K\colon L^1(\mu) \to \mathbb{R}$ *is translation invariant. More specifically,* $c \in \mathbb{R}$ *implies* $K(u+c) = K(u) + cK(1)$.
7. *The functional* $K\colon L^1(\mu) \to \mathbb{R}$ *is continuous and quasi-convex, namely all its sub-levels* $L_\alpha = \left\{u \in L^1(\mu) \middle| K(u) \le \alpha\right\}$ *are convex.*
8. $K : L^1(\mu) \to \mathbb{R}$ *is convex.*

Proof. 1. As $\exp_A(u + \log_A p) \le u^+ + p$ and so $\exp_A(u+\log_A p) \in L^1(\mu)$ for all $u \in L^1(\mu)$. The estimate $|\exp_A(u+\log_A p) - \exp_A(v+\log_A p)| \le |u - v|$ leads to the desired result.

2. For all $\kappa \in \mathbb{R}$ the integral $I(\kappa) = \int \exp_A(u - \kappa + \log_A p)\, d\mu$ is bounded by $1 + \int (u-\kappa)^+\, d\mu < \infty$ and the function $\kappa \mapsto I(k)$ is continuous and strictly decreasing. Convexity of $\exp_A$ together with the equation for its derivative imply $\exp_A(u-\kappa+\log_A p) \ge \exp_A(u+\log_A p) - A(\exp_A(u+\log_A p))\kappa$, so that $\int \exp_A(u-\kappa+\log_A p)\, d\mu \ge \int \exp_A(u+\log_A p)\, d\mu - \kappa \int A(\exp_A(u + \log_A p))\, d\mu$, where the coefficient of κ is positive. Hence $\lim_{\kappa\to-\infty} \int \exp_A(u - \kappa+\log_A p)\, d\mu = +\infty$. For each $\kappa \ge 0$, we have $\exp_A(u-\kappa+\log_A p) \le \exp_A(u+\log_A p) \le p + u^+$ so that by dominated convergence we get $\lim_{\kappa\to\infty} I(\kappa) = 0$. Therefore $K(u)$ will be the unique value for which $\int \exp_A(u-\kappa+\log_A p)\, d\mu = 1$.
3. If the function u is a constant, then $\int \exp_A(u - u + \log_A p)\, d\mu = \int p\, d\mu = 1$ and so $K(u) = u$. The converse implication is trivial. The equality $\exp_A(u - K(u) + \log_A p) = p$ holds if, and only if, $u - K(u) = 0$.
4. If $\log_A q = u - K(u) + \log_A p$, then $\log_A q - \log_A p = u - K(u) \in L^1(\mu)$. Conversely, if $\log_A q - \log_A p = v \in L^1(\mu)$, then $q = \exp_A(v + \log_A p)$. As q is a density, then $K(v) = 0$.
5. If $u - K(u) + \log_A p = v - K(v) + \log_A p$, then $u - v = K(u) - K(v)$.
6. Clearly, $K(c) = c = cK(1)$ and $K(u+c) = K(u) + c$.
7. Observe that $\int \exp_A(u+\log_A p)\, d\mu \le 1$ if, and only if, $K(u) \le 0$. Hence $u_1, u_2 \in L_0$, implies $\int \exp_A(u_i + \log_A p)\, d\mu \le 1$, $i = 1, 2$. Thanks to the convexity of the function $\exp_A$, we have $\int \exp_A((1-\alpha)u_1 + \alpha u_2) + \log_A p\, d\mu \le (1-\alpha)\int \exp_A(u_1+\log_A p)\, d\mu + \alpha \int \exp_A(u_2 + \log_A p)\, d\mu \le 1$, that provides $K((1-\alpha)u_1 + \alpha u_2) \le 0$. Hence the sub-level L_0 is convex. Notice that all the other sub-levels are convex since they are obtained by translation of L_0. More precisely, $L_\alpha = L_0 + \alpha$. Clearly both the sets $\left\{\int \exp_A(u+\log_A p)\, d\mu \le 1\right\}$ and $\left\{\int \exp_A(u+\log_A p)\, d\mu \ge 1\right\}$ are closed in $L^1(\mu)$, since the functional $u \to \int \exp_A(u)\, d\mu$ is continuous. Hence $u \to K(u)$ is continuous as well.
8. A functional which is translation invariant and quasiconvex is necessarily convex. Though this property is more or less known, a proof is gathered below.

Lemma 1. *A translation invariant functional on a vector space* V*, namely* $I\colon V \to \mathbb{R}$ *such that for some* $v \in V$ *one has* $I(x + \lambda v) = I(x) + \lambda I(v)$ *for all* $x \in V$ *and* $\lambda \in \mathbb{R}$*, is convex if and only if* I *is quasiconvex, namely all level sets are convex, provided* $I(v) \neq 0$.

Proof. Let I be quasiconvex, then the sublevel $L_0(I) = \{x \in V : I(x) \le 0\}$ is nonempty and convex. Clearly, $L_\lambda(I) = L_0(I) + (\lambda/I(v))v$ holds for every $\lambda \in \mathbb{R}$.

Hence, if λ and μ are any pair of assigned real numbers and $\alpha \in (0,1)$, $\bar{\alpha} = 1-\alpha$, then

$$\begin{aligned}\alpha L_\lambda(I) + \bar{\alpha} L_\mu(I) &= \alpha L_0(I) + \bar{\alpha} L_0(I) + \frac{\alpha\lambda + \bar{\alpha}\mu}{I(v)} v \\ &= L_0(I) + \frac{\alpha\lambda + \bar{\alpha}\mu}{I(v)} v = L_{\alpha\lambda+\bar{\alpha}\mu}(I) \ .\end{aligned}$$

Therefore, if for any pair of points $x, y \in V$, we set $I(x) = \lambda$ and $I(y) = \mu$, then $x \in L_\lambda(I)$ and $y \in L_\mu(I)$. Consequently $\alpha x + \bar{\alpha} y \in \alpha L_\lambda(I) + \bar{\alpha} L_\mu(I) = L_{\alpha\lambda+\bar{\alpha}\mu}(I)$. That is, $I(\alpha x + \bar{\alpha} y) \leq \alpha\lambda + \bar{\alpha}\mu = \alpha I(x) + \bar{\alpha} I(y)$ that shows the convexity of I. Of course the converse holds in that a convex function is quasi-convex.

For each positive density q, define its *escort density* to be $\widetilde{q} = A(q)/\int A(q)\, d\mu$, see [4]. Notice that $0 < A(q) < 1$. The next proposition provides a subgradient of the convex function K.

Proposition 2. *Let $v \in L^1(\mu)$ and $q(v) = \exp_A(v - K(v) + \log_A p)$. For every $u \in L^1(\mu)$, the inequality $K(u+v) - K(v) \geq \int u\widetilde{q}(v)\, d\mu$ holds i.e., the density $\widetilde{q}(v) \in L^\infty(\mu)$ is a subgradient of K at v.*

Proof. Thanks to convexity of $\exp_A$ and the derivation formula, we have

$$\exp_A(u + v - K(u+v) + \log_A p) - q \geq A(q)(u - K(u+v) + K(v)) \ .$$

If we take μ-integral of both sides,

$$0 \geq \int uA(q)\, d\mu - (K(u+v) - K(v)) \int A(q)\, d\mu \ .$$

Isolating the increment $K(u+v) - K(v)$, the desired inequality obtains.

By Proposition 2, if the functional K were differentiable, the gradient mapping would be $v \mapsto \widetilde{q}(v)$, whose strong continuity requires additional assumptions. We would like to show that K is differentiable by means of the Implicit Function Theorem. That too, would require specific assumptions. In fact, it is in general not true that a superposition operator such as $L^1(\mu) \ni u \mapsto \exp_A(u + \log_A p) \in L^1(\mu)$ is differentiable, cf. [2, Sect. 1.2]. In this perspective, we prove the following.

Proposition 3. *1. The superposition operator $L^2(\mu) \ni v \mapsto \exp_A(v + \log_A p) \in L^1(\mu)$ is continuously Fréchet differentiable with derivative*

$$d\exp_A(v) = (h \mapsto A(\exp_A(v + \log_A p))h) \in \mathcal{L}(L^2(\mu), L^1(\mu)) \ .$$

2. The functional $K : L^2(\mu) \to \mathbb{R}$, implicitly defined by the equation

$$\int \exp_A(v - K(v) + \log_A p)\, d\mu = 1, \quad v \in L^2(\mu)$$

is continuously Fréchet differentiable with derivative

$$dK(v) = (h \mapsto \int h\widetilde{q}(v)\ d\mu), \quad q(v) = \exp_A(v - K(v))$$

where

$$\widetilde{q}(v) = \frac{A \circ q(v)}{\int A \circ q(v)\ d\mu}$$

is the escort density of p.

Proof. 1. It is easily seen that

$$\exp_A(v + h + \log_A p) - \exp_A(v + \exp_A p) - A[\exp_A(v + \log_A p)]h = R_2(h),$$

with the bound $|R_2(h)| \leq (1/2)\,|h|^2$. It follows

$$\frac{\int |R_2(h)|\, d\mu}{\left(\int |h|^2\ d\mu\right)^{\frac{1}{2}}} \leq \frac{\frac{1}{2}\int |h|^2\ d\mu}{\left(\int |h|^2\ d\mu\right)^{\frac{1}{2}}} = \frac{1}{2}\left(\int |h|^2\ d\mu\right)^{\frac{1}{2}} .$$

Therefore $\|R_2(h)\|_{L^1(\mu)} = o\left(\|h\|_{L^2(\mu)}\right)$ and so the operator $v \mapsto \exp_A(v + \log_A p)$ is Fréchet-differentiable with derivative $h \mapsto A(\exp_A(v + \log_A p))h$ at v. Let us show that the F-derivative is a continuous map $L^2(\mu) \to \mathcal{L}(L^2(\mu), L^1(\mu))$. If $\|h\|_{L^2(\mu)} \leq 1$ and $v, w \in L^2(\mu)$ we have

$$\begin{aligned}&\int |(A[\exp_A(v + \log_A p)] - A[\exp_A(w + \log_A p)])h|\ d\mu \\ &\quad \leq \|A[\exp_A(v + \log_A p) - A[\exp_A(w + \log_A p)]\|_{L^2(\mu)} \leq \|v - w\|_{L^2(\mu)}\ ,\end{aligned}$$

hence the derivative is 1-Lipschitz.

2. Frechét differentiability of K is a consequence of the Implicit Function Theorem in Banach spaces, see [3], applied to the C^1-mapping

$$L^2(\mu) \times \mathbb{R} \ni (v, \kappa) \mapsto \int \exp_A(v - \kappa + \log_A p)\ d\mu\ .$$

The derivative can be easily obtained from the computation of the subgradient.

In the expression $q(u) = \exp_A(u - K(u) + \log_A p)$, $u \in L^1(\mu)$, the random variable u is identified up to a constant. We can choose in the class a unique representative, by assuming $\int u\widetilde{p}\ d\mu = 0$, the expected value being well defined as the escort density is bounded. In this case we can solve for u and get

$$u = \log_A q - \log_A p - \mathrm{E}_{\widetilde{p}}\left[\log_A p - \log_A q\right]$$

In analogy with the exponential case, we can express the functional K as a divergence associated to the N.J. Newton logarithm:

$$K(u) = \mathrm{E}_{\widetilde{p}}\left[\log_A p - \log_A q(u)\right] = D_A(p\|q(u))\ .$$

It would be interesting to proceed with the study of the convex conjugation of K and the related properties of the divergence, but do not do that here.

4 Hilbert Bundle Based on $\exp_A$

In this section $A(x) = x/(1+x)$ and $\mathcal{P}(\mu)$ denotes the set of all μ-densities on the probability space $(\mathbb{X}, \mathcal{X}, \mu)$ of the form $q = \exp_A(u - K(u))$ with $u \in L^2(\mu)$ and $\mathrm{E}_\mu[u] = 0$, cf. [5]. Notice that $1 \in \mathcal{P}(\mu)$ because we can take $u = 0$. Equivalently, $\mathcal{P}(\mu)$ is the set of all densities q such that $\log_A q \in L^2(\mu)$ because in such a case we can take $u = \log_A q - \mathrm{E}_\mu[\log_A q]$. The condition for $q \in \mathcal{P}(\mu)$ can be expressed by saying that both q and $\log q$ are in $L^2(\mu)$. In fact, as $\exp_A$ is 1-Lipschitz, we have $\|q-1\|_\mu \leq \|u - K(u)\|_\mu$ and the other inclusion follows from $\log q = \log_A q + 1 - q$. An easy but important consequence of such a characterization is the compatibility of the class $\mathcal{P}(\mu)$ with the product of measures. If $q_i = \exp_A(u_i - K_1(u_i)) \in \mathcal{P}(\mu_i)$, $i = 1, 2$, the product is $(q_1 \cdot \mu_1) \otimes (q_2 \cdot \mu_2) = (q_2 \otimes q_2) \cdot (\mu_1 \otimes \mu_2)$, hence $q_2 \otimes q_2 \in \mathcal{P}(\mu_1 \otimes \mu_2)$ since $\|q_1 \otimes q_2\|_{\mu_1 \otimes \mu_2} = \|q_1\|_{\mu_1} \|q_2\|_{\mu_2}$. Moreover $\log(q_1 \otimes q_2) = \log q_1 + \log q_2$, hence $\|\log(q_1 \otimes q_2)\|_{\mu_1 \otimes \mu_2} \leq \|\log q_1\|_{\mu_1} + \|\log q_2\|_{\mu_2}$.

We proceed now to define an Hilbert bundle with base $\mathcal{P}(\mu)$. For each $p \in \mathcal{P}(\mu)$ consider the Hilbert spaces $H_p = \left\{u \in L^2(\mu) \middle| \mathrm{E}_{\widetilde{p}}[u] = 0\right\}$ with scalar product $\langle u, v\rangle_p = \int uv \, d\mu$ and form the Hilbert bundle

$$H\mathcal{P}(\mu) = \{(p, u) | p \in \mathcal{P}(\mu), u \in H_p\} \ .$$

For each $p, q \in \mathcal{P}(\mu)$ the mapping $\mathbb{U}_p^q u = u - \mathrm{E}_{\widetilde{q}}[u]$ is a continuous linear mapping from H_p to H_q. We have $\mathbb{U}_q^r \mathbb{U}_p^q = \mathbb{U}_p^r$. In particular, $\mathbb{U}_q^p \mathbb{U}_p^q$ is the identity on H_p, hence $\mathbb{U}_p^q$ is an isomorphism of H_p onto H_q. In the next proposition we construct an atlas of charts for which $\mathcal{P}(\mu)$ is a Riemannian manifold and $H\mathcal{P}(\mu)$ is an expression of the tangent bundle.

In the following proposition we introduce an affine atlas of charts and use it to define our Hilbert bundle which is an expression of the tangent bundle. The velocity of a curve $t \mapsto p(t) \in \mathcal{P}(\mu)$ is expressed in the Hilbert bundle by the so called A-score that, in our case, takes the form $A(p(t))^{-1}\dot{p}(t)$, with $\dot{p}(t)$ computed in $L^1(\mu)$.

Proposition 4. *1. $q \in \mathcal{P}(\mu)$ if, and only if, both q and $\log q$ are in $L^2(\mu)$.*

2. Fix $p \in \mathcal{P}(\mu)$. Then a positive density q can be written as

$$q = \exp_A(v - K_p(v) + \log_A p), \quad \text{with } v \in L^2(\mu) \text{ and } \mathrm{E}_{\widetilde{p}}[v] = 0,$$

if, and only if, $q \in \mathcal{P}(\mu)$.

3. For each $p \in \mathcal{P}(\mu)$ the mapping

$$s_p \colon \mathcal{P}(\mu) \ni q \mapsto \log_A q - \log_A p - \mathrm{E}_{\widetilde{p}}[\log_A q - \log_A p] \in H_p$$

is injective and surjective, with inverse $e_p(u) = \exp_A(u - K_p(u) + \log_A p)$.

4. The atlas $\{s_p | p \in \mathcal{P}(\mu)\}$ is affine with transitions

$$s_q \circ e_p(u) = \mathbb{U}_p^q u + s_p(q) \ .$$

5. *The expression of the velocity of the differentiable curve $t \mapsto p(t) \in \mathcal{P}(\mu)$ in the chart s_p is $ds_p(p(t))/dt \in H_p$. Conversely, given any $u \in H_p$, the curve $p\colon t \mapsto \exp_A(tu - K_p(tu) + \log_A p)$ has $p(0) = p$ and has velocity at $t = 0$ expressed in the chart s_p by u. If the velocity of a curve is expressed in the chart s_p by $t \mapsto \dot{u}(t)$, then its expression in the chart s_q is $\mathbb{U}_p^q \dot{u}(t)$.*
6. *If $t \mapsto p(t) \in \mathcal{P}(\mu)$ is differentiable with respect to the atlas then it is differentiable as a mapping in $L^1(\mu)$. It follows that the A-score is well-defined and is the expression of the velocity of the curve $t \mapsto p(t)$ in the moving chart $t \mapsto s_{p(t)}$.*

Proof. 1. Assume $q = \exp_A(u - K(u))$ with $u \in L_0^2(\mu)$. It follows $u - K(u) \in L^2(\mu)$ hence $q \in L^2(\mu)$ because $\exp_A$ is 1-Lipschitz. As moreover $q + \log q - 1 = u - K(u) \in L^2(\mu)$, then $\log q \in L^2(\mu)$. Conversely, $\log_a q = q - 1 + \log q = v \in L^2(\mu)$ and we can write $q = \exp_A v = \exp_A((v - \mathrm{E}_p[v]) + \mathrm{E}_p[v])$ and we can take $u = v - \mathrm{E}_\mu[v]$.

2. The assumption $p, q \in \mathcal{P}(\mu)$ is equivalent to $\log_A p, \log_A q \in L^2(\mu)$. Define $u = \log_A q - \log_A p - \mathrm{E}_{\widetilde{p}}[\log_A q - \log_A p]$ and $D_A(p\|q) = \mathrm{E}_{\widetilde{p}}[\log_A p - \log_A q]$. It follows $u \in L^2(\mu)$, $\mathrm{E}_{\widetilde{p}}[u] = 0$, and $\exp_A(u - D_A(p\|q) + \log_A p) = q$. Conversely, $\log_A q = u - K_p(u) + \log_A p \in L^2(\mu)$.
3. This has been already proved.
4. All simple computations.
5. If $p(t) = \exp_A(u(t) - K_p(u(t)) + \log_A p)$, with $u(t) = s_p(u(t))$ then in that chart the velocity is $\dot{u}(t) \in H_p$. When $u(t) = tu$ the expression of the velocity will be u. The proof of the second part follows from the fact that $\mathbb{U}_p^q$ is the linear part of the affine change of coordinates $s_q \circ e_p$.
6. Choose a chart s_p and express the curve as $t \mapsto s_p(p(t)) = u(t)$ so that $p(t) = \exp_A(u(t) - K_p(u(t)) + \log_A p)$. It follows that the derivative of $t \mapsto p(t)$ exists in $L^1(\mu)$ by derivation of the composite function and it is given by $\dot{p}(t) = A(p(t))\mathbb{U}_p^{p(t)} \dot{u}(t)$, hence $A(p(t))^{-1}\dot{p}(t) = \mathbb{U}_p^{p(t)} \dot{u}(t)$. If the velocity at t is expressed in the chart centered at $p(t)$, then its expression is the score.

5 Conclusions

We have constructed an Hilbert statistical bundle using an affine atlas of charts based on the A-logarithm with $A(x) = x/(1 + x)$. In particular, this entails a Riemannian manifold of densities. On the other end, our bundle structure could be useful in certain contexts. The general structure of the argument mimics the standard case of the exponential manifold. We would like to explicit some, hopefully new, features of our set-up.

The proof of the convexity and continuity of the functional K when defined on $L^1(\mu)$ relies on the property of translation invariance. Whenever K is restricted to $L^2(\mu)$, it is shown to be differentiable along with the deformed exponential and this, in turn, provides a rigorous construction of the A-score.

The gradient mapping of K is continuous and 1-to-1, but its inverse cannot be continuous as it takes values which are bounded functions. It would be interesting to analyze the analytic properties of the convex conjugate of K^*, as both K and K^* are the coordinate expression of relevant divergences.

If F is a section of the Hilbert bundle namely, $F\colon \mathcal{P}(\mu) \to L^2(\mu)$ with $\mathrm{E}_{\widetilde{p}}\left[F(p)\right] = 0$ for all p, differential equations take the form $A(p(t))\dot{p}(t) = F(p(t))$ in the atlas, which in turn implies $\dot{p}(t) = A(p(t))F(p(t))$ in $L^1(\mu)$. This is important for some applications e.g., when the section F is the gradient with respect to the Hilbert bundle of a real function. Namely, the gradient, $\operatorname{grad}\phi$, of a smooth function $\phi\colon \mathcal{P}(\mu) \to \mathbb{R}$ is a section of the Hilbert bundle such that

$$\frac{d}{dt}\phi(p(t)) = \left\langle \operatorname{grad}\phi(p(t)), A(p(t))\dot{p}(t) \right\rangle_\mu$$

for each differentiable curve $t \mapsto p(t) \in \mathcal{P}(\mu)$.

Acknowledgments. L. Montrucchio acknowledges the support of Collegio Carlo Alberto Foundation. G. Pistone is a member of GNAFA-INDAM and acknowledges the support of de Castro Statistics Foundation and Collegio Carlo Alberto Foundation.

References

1. Amari, S.: Dual connections on the Hilbert bundles of statistical models. In: Geometrization of Statistical Theory (Lancaster, 1987), pp. 123–151. ULDM Publ., Lancaster (1987)
2. Ambrosetti, A., Prodi, G.: A Primer of Nonlinear Analysis, Cambridge Studies in Advanced Mathematics, vol. 34. Cambridge University Press, Cambridge (1993)
3. Dieudonné, J.: Foundations of Modern Analysis. Academic Press, New York (1960)
4. Naudts, J.: Generalised Thermostatistics. Springer, London (2011). doi:10.1007/978-0-85729-355-8
5. Newton, N.J.: An infinite-dimensional statistical manifold modelled on Hilbert space. J. Funct. Anal. **263**(6), 1661–1681 (2012)
6. Pistone, G.: Nonparametric information geometry. In: Nielsen, F., Barbaresco, F. (eds.) GSI 2013. LNCS, vol. 8085, pp. 5–36. Springer, Heidelberg (2013). doi:10.1007/978-3-642-40020-9_3
7. Pistone, G., Sempi, C.: An infinite-dimensional geometric structure on the space of all the probability measures equivalent to a given one. Ann. Statist. **23**(5), 1543–1561 (1995)
8. Schwachhöfer, L., Ay, N., Jost, J., Lê, H.V.: Parametrized measure models. Bernoulli (online-to appear)
9. Vigelis, R.F., Cavalcante, C.C.: On ϕ-families of probability distributions. J. Theor. Probab. **26**, 870–884 (2013)
10. Zhang, J., Hästö, P.: Statistical manifold as an affine space: a functional equation approach. J. Math. Psychol. **50**(1), 60–65 (2006)