A Class of Non-parametric Deformed Exponential Statistical Models

Luigi Montrucchio and Giovanni Pistone

Abstract We study the class on non-parametric deformed statistical models where the deformed exponential has linear growth at infinity and is sub-exponential at zero. This class generalizes the class introduced by N.J. Newton. We discuss the convexity and regularity of the normalization operator, the form of the deformed statistical divergences and their convex duality, the properties of the escort densities, and the affine manifold structure of the statistical bundle

1 Introduction

In this paper we study a geometry on the set $\mathscr{P}$ of strictly positive probability densities on a probability space $(\mathbb{X}, \mathscr{X}, \mu)$. In some cases one is led to consider the set $\overline{\mathscr{P}}$ of probability densities i.e., without the restriction of strict positivity. There is a considerable literature on the Information Geometry in the sense defined in the Amari and Nagaoka monograph [2] on $\mathscr{P}$. There is also a non-parametric approach i.e., we are not considering the geometry induced on the parameter set of a given statistical model but on the full set of densities. This was done in [21, 23] by using logarithmic chart to represent densities.

A different approach, that leads to the construction of an Hilbert manifold on $\mathscr{P}$, has been proposed by N.J. Newton in [18, 19]. It is based on the use of the chart $p \mapsto p - 1 - \log p$ instead of a purely logarithmic chart. This paper presents a variation on the same theme by enlarging the class of permitted charts.

Let $\mathscr{M} \subset \mathscr{P}$. At each $p \in \mathscr{M}$, the Hilbert space of square-integrable random variables $L^2(p)$ provides a fiber that sits at $p \in \mathscr{M}$, so we can define the *Hilbert bundle* with base $\mathscr{M}$. The Hilbert bundle, or similar bundles with fibers which are

L. Montrucchio
Collegio Carlo Alberto, Piazza Vincenzo Arbarello 8, 10122 Turin, Italy
e-mail: luigi.montrucchio@unito.it

G. Pistone (✉)
de Castro Statistics, Collegio Carlo Alberto, Piazza Vincenzo Arbarello 8,
10122 Turin, Italy
e-mail: giovanni.pistone@carloalberto.org

vector spaces of random variables, provides a convenient framework for Information Geometry, cf. [1, 12, 21].

If $\mathcal{M}$ is an exponential manifold in the sense of [23], there exists a splitting of each fiber $L^2(p) = \mathcal{H}_p \oplus \mathcal{H}_p^\perp$, such that each $\mathcal{H}_p$ contains a dense vector sub-space which is an expression of the tangent space $T_p\mathcal{M}$ of the manifold. Moreover, the manifold on $\mathcal{M}$ is an affine manifold (it can be defined by an atlas whose transition mapping are affine) and it is also an Hessian manifold (the inner product on each fiber is the second derivative of a potential function, [24]).

When the sample space is finite and $\mathcal{M}$ is the full set $\mathcal{P}$ of positive probability densities, then $\mathcal{H}_p$ is the space of centered square integrable random variables $L_0^2(p)$ and moreover there is an identification of the fiber with the tangent space $\mathcal{H}_p \simeq T_p\mathcal{P}$. A similar situation occurs even when $\mathcal{M}$ is a finite-dimensional exponential family. It is difficult to devise set-ups other than those mentioned above, where the identification of the Hilbert fiber with the tangent space holds true. In fact, a necessary condition would be the topological linear isomorphism among fibers. One possible option would be to take as fibers the spaces of bounded functions $L_0^\infty(p)$, see G. Loaiza and H.R. Quiceno [14].

This difficulty is overcome in the N.J. Newton's setting. On a probability space $(X, \mathcal{X}, \mu)$, he considers the "balanced chart" $\mathcal{M} \ni p \mapsto \log p + p - 1 \in L_0^2(\mu)$. In this chart, all the tangent spaces are identified with the fixed Hilbert space $L_0^2(\mu)$ so that the statistical Hilbert bundle is trivialized.

N.J. Newton balanced chart falls in a larger class of "deformation" of the usual logarithmic representation. It is in fact an instance of the class of "deformed logarithm" as defined by J. Naudts [17]. It is defined as $\log_A(x) = \int_1^x dt/A(t)$, where A is a suitable increasing function. If A is bounded, then a special class of deformed logarithms results. It includes N.J. Newton balanced chart as well as other deformed logarithms, notably the G. Kaniadakis logarithm [10, 11, 20].

In this paper, we try a mixture of the various approaches by considering deformed logarithms with linear growth as established by N.J. Newton, but we do not look for a trivialization of the Hilbert bundle. Instead we construct an affine atlas of charts, each one centered at a $p \in \mathcal{M}$. This is obtained by adapting the construction of the exponential manifold of [21] to the deformed exponential models as defined by J. Naudts [17]. Moreover, we allow for a form of general reference measure by using an idea introduced by R.F. Vigelis and C.C. Cavalcante [26]. That is, each density has the form $q = \exp_A(u - K_p(u) + \log_A p)$, where $\exp_A = \log_A^{-1}$ is an exponential-like function which has a linear growth at $+\infty$ and is dominated by an exponential at $-\infty$.

The formalism of deformed exponentials is discussed in Sect. 2. This section is intended to be self-contained and contains material from the references discussed above without an explicit mention. The following Sect. 3 is devoted to the study of non-parametric deformed exponential families. In Sect. 4 we introduce the formulation of the divergence, in accordance with our approach. In Sect. 5 the construction of the Hilbert statistical bundle is outlined.

A first version of this piece of research has been presented at the GSI 2017 Conference [16] and we refer to that paper for some of the proofs.

2 Deformed Exponential

Let us introduce a class of the deformed exponential, according to the formalism introduced by [17]. Assume to be given a function A from $]0, +\infty[$ onto $]0, a[$, strictly increasing, continuously differentiable and such that $\|A'\|_\infty < \infty$. This implies $a = \|A\|_\infty$ and $A(x) \leq \|A'\|_\infty x$, so that $\int_0^1 d\xi/A(\xi) = +\infty$.

The A-logarithm is the function

$$\log_A(x) = \int_1^x \frac{d\xi}{A(\xi)} , \quad x \in]0, +\infty[.$$

The A-logarithm is strictly increasing from $-\infty$ to $+\infty$, its derivative $\log_A'(x) = 1/A(x)$ is positive and strictly decreasing for all $x > 0$, hence $\log_A$ is strictly concave.

By inverting the A-logarithm, one obtains the A-exponential, $\exp_A = \log_A^{-1}$. The function $\exp_A :]-\infty, +\infty[\to]0, +\infty[$ is strictly increasing, strictly convex, and is the solution to the Cauchy problem

$$\exp_A'(y) = A(\exp_A(y)), \quad \exp_A(0) = 1 . \tag{1}$$

As a consequence, we have the linear bound

$$\left|\exp_A(y_1) - \exp_A(y_2)\right| \leq \|A\|_\infty |y_1 - y_2| . \tag{2}$$

The behavior of the A-logarithm is linear for large arguments and super-logarithmic for small arguments. To derive explicit bounds, set

$$\alpha_1 = \min_{x \leq 1} \frac{A(x)}{x} , \quad \alpha_2 = \max_{x \leq 1} \frac{A(x)}{x} ,$$

namely, they are the best constants such that $\alpha_1 x \leq A(x) \leq \alpha_2 x$ for $0 < x \leq 1$. Note that $\alpha_1 \geq 0$ while $\alpha_2 > 0$. If in addition also $\alpha_1 > 0$, then

$$\frac{1}{\alpha_2} \log x \leq \log_A x \leq \frac{1}{\alpha_1} \log x , \quad 0 < x \leq 1 . \tag{3}$$

If otherwise $\alpha_1 = 0$, the left inequality is true only.

For $x \geq 1$ we have $A(1) \leq A(x) < \|A\|_\infty$, hence

$$\frac{1}{\|A\|_\infty}(x - 1) < \log_A x \leq \frac{1}{A(1)}(x - 1) , \quad x \geq 1 . \tag{4}$$

Under the assumptions made on the function A, the coefficient $\alpha_1 > 0$, if and only if $A'(0+) > 0$.

2.1 Examples

The main example of A-logarithm is the N.J. Newton A-logarithm [18], with

$$A(\xi) = 1 - \frac{1}{1+\xi} = \frac{\xi}{1+\xi} ,$$

so that

$$\log_A(x) = \log x + x - 1 .$$

There is a simple algebraic expression for the product,

$$\log_A(x_1 x_2) = \log_A(x_1) + \log_A(x_2) + (x_1 - 1)(x_2 - 1) .$$

Other similar examples are available in the literature. One is a special case of the G. Kaniadakis' exponential of [9], generated by

$$A(\xi) = \frac{2\xi^2}{1+\xi^2} .$$

It turns out

$$\log_A x = \frac{x - x^{-1}}{2} ,$$

whose inverse provides

$$\exp_A(y) = y + \sqrt{1 + y^2} .$$

A remarkable feature of the G. Kaniadakis' exponential is the relation

$$\exp_A(y) \exp_A(-y) = \left(y + \sqrt{1 + y^2} \right) \left(-y + \sqrt{1 + y^2} \right) = 1$$

Notice that the A function for N.J. Newton exponential is concave, while the A function of G. Kaniadakis exponential is not.

Another example is $A(\xi) = 1 - 2^{-\xi}$, which gives $\log_A(x) = \log_2(1 - 2^{-x})$ and $\exp_A(y) = \log_2(1 + 2^y)$.

Notable examples of deformed exponentials that do not fit into our set of assumptions are Tsallis q-logarithms, see [25]. For instance, for $q = 1/2$,

$$\log_{1/2} x = 2\left(\sqrt{x} - 1\right) = \int_1^x \frac{d\xi}{\sqrt{\xi}}.$$

In this case, $\log_{1/2}(0+) = -\int_0^1 d\xi/\sqrt{\xi} = -2$, so that the inverse is not defined for all real numbers. Tsallis logarithms provide models having heavy tails, which is not the case in our setting.

2.2 Superposition Operator

The deformed exponential will be employed to represent positive probability densities in the type $p(x) = \exp_A[u(x)]$, where u is a random variable on a probability space $(\mathbb{X}, \mathcal{X}, \mu)$. For this reason, we are interested in the properties of the *superposition operator*

$$S_A : u \mapsto \exp_A \circ u \qquad (5)$$

defined in some convenient functional setting. About superposition operators, see e.g. [3, Ch. 1] and [4, Ch. 3].

It is clear from the Lipschitz condition (2) that $\exp_A(u) \leq 1 + \|A\|_\infty |u|$, which in turn implies that the superposition operator S_A maps $L^\alpha(\mu)$ into itself for all $\alpha \in [1, +\infty]$ and the mapping is uniformly Lipschitz with constant $\|A\|_\infty$. Notice that we are assuming that μ is a finite measure.

The superposition operator $S_A : L^\alpha(\mu) \to L^\alpha(\mu)$ is 1-to-1 and its image consists of all positive random variables f such that $\log_A f \in L^\alpha(\mu)$. The following proposition intercepts a more general result [19]. We give a direct proof here for sake of completeness and because our setting includes deformed logarithms other than the case treated there.

Proposition 1 *1. For all $\alpha \in [1, \infty]$, the superposition operator S_A of Eq. (5) is Gateaux-differentiable with derivative*

$$dS_A(u)[h] = A(\exp_A(u))h . \qquad (6)$$

2. *S_A is Fréchet-differentiable from $L^\alpha(\mu)$ to $L^\beta(\mu)$, for all $\alpha > \beta \geq 1$.*

Proof 1. Equation (1) implies that for each couple of random variables $u, h \in L^\alpha(\mu)$

$$\lim_{t \to 0} t^{-1} \left(\exp_A(u + th) - \exp_A(u) \right) - A(\exp_A(u))h = 0$$

holds point-wise. Moreover, if each $\alpha \in [1, \infty[$, by Jensen inequality we infer that if $t > 0$ then

$$\left| t^{-1} \left(\exp_A(u + th) - \exp_A(u) \right) - A(\exp_A(u))h \right|^\alpha \leq$$
$$t^{-1} |h|^\alpha \int_0^t \left| A(\exp_A(u + rh)) - A(\exp_A(u)) \right|^\alpha \, dr \leq (2 \|A\|_\infty)^\alpha |h|^\alpha .$$

Now, dominated convergence forces the limit to hold in $L^\alpha(\mu)$. If $t < 0$, it suffices to replace h with $-h$.

Whenever $\alpha = \infty$, we can use the second-order bound

$$\left| t^{-1} \left(\exp_A(u + th) - \exp_A(u) \right) - A(\exp_A(u))h \right| =$$

$$|t|^{-1} h^2 \left| \int_0^t (t - r) \frac{d}{dr} A(\exp_A(u + rh)) \, dr \right| \le \frac{t}{2} \| h \|_\infty^2 \| A' \|_\infty \| A \|_\infty .$$

As $\| A' \cdot A \|_\infty < \infty$, the RHS goes to 0 as $t \to 0$ uniformly for each $h \in L^\infty(\mu)$.

2. Given $u, h \in L^\alpha(\mu)$, thanks again to Taylor formula,

$$\int \left| \exp_A(u + h) - \exp_A(u) - A(\exp_A(u))h \right|^\beta \, d\mu \le$$

$$\int |h|^\beta \int_0^1 \left| A(\exp_A(u + rh)) - A(\exp_A(u)) \right|^\beta \, dr \, d\mu .$$

By means of Hölder inequality, with conjugate exponents α/β and $\alpha/(\alpha - \beta)$, the RHS is bounded by

$$\left(\int |h|^\alpha \, d\mu \right)^{\frac{\beta}{\alpha}} \left(\iint \left| A(\exp_A(u + rh)) - A(\exp_A(u)) \right|^{\frac{\alpha\beta}{\alpha - \beta}} \, dr \, d\mu \right)^{\frac{\alpha - \beta}{\alpha}} .$$

Consequently,

$$\| h \|_{L^\alpha(\mu)}^{-1} \left\| \exp_A(u + h) - \exp_A(u) - A(\exp_A(u))h \right\|_{L^\beta(\mu)} \le$$

$$\left(\iint \left| A(\exp_A(u + rh)) - A(\exp_A(u)) \right|^{\frac{\alpha\beta}{\alpha - \beta}} \, dr \, d\mu \right)^{\frac{\alpha - \beta}{\alpha\beta}} .$$

In order to show that the RHS vanishes as $\| h \|_{L^\alpha(\mu)} \to 0$, observe that for all $\delta > 0$ we have

$$\left| A(\exp_A(u + rh)) - A(\exp_A(u)) \right| \le \begin{cases} 2 \| A \|_\infty & \text{always,} \\ \| A' \|_\infty \| A \|_\infty \delta & \text{if } |h| \le \delta, \end{cases}$$

so that, decomposing the double integral as $\iint = \iint_{|h| \le \delta} + \iint_{|h| > \delta}$, we obtain

$$\iint \left| A(\exp_A(u + rh)) - A(\exp_A(u)) \right|^\gamma \, dr \, d\mu \le$$

$$(2 \| A \|_\infty)^\gamma \mu \{ |h| > \delta \} + \left(\| A' \|_\infty \| A \|_\infty \delta \right)^\gamma \le$$

$$(2 \| A \|_\infty)^\gamma \delta^{-\alpha} \int |h|^\alpha \, d\mu + \left(\| A' \|_\infty \| A \|_\infty \delta \right)^\gamma ,$$

where $\gamma = \alpha\beta/(\alpha - \beta)$ and we have used Čebičev inequality. Now it is clear that the last bound implies the conclusion for each $\alpha < \infty$. The case $\alpha = \infty$ follows a fortiori. $\qquad\square$

Remark 1 It is not generally true that the superposition operator S_A be Fréchet differentiable for $\alpha = \beta$, cf. [3, §1.2]. We repeat here the well known counter-example.

Assume μ is a non-atomic probability measure. For each $\lambda \in \mathbb{R}$ and $\delta > 0$ define the simple function

$$h_{\lambda,\delta}(x) = \begin{cases} \lambda & \text{if } |x| \leq \delta, \\ 0 & \text{otherwise.} \end{cases}$$

For each $\alpha \in [1, +\infty[$ we have

$$\lim_{\delta \to 0} \left\| h_{\lambda,\delta} \right\|_{L^\alpha(\mu)} = \lim_{\delta \to 0} |\lambda| \, \mu \{|x| \leq \delta\}^{1/\alpha} = 0 \, .$$

Differentiability at 0 in $L^\alpha(\mu)$ would imply for all λ

$$0 = \lim_{\delta \to 0} \frac{\left\| \exp_A(h_{\lambda,\delta}) - 1 - A(1)h_{\lambda,\delta} \right\|_{L^\alpha(\mu)}}{\left\| h_{\lambda,\delta} \right\|_{L^\alpha(\mu)}} =$$

$$\lim_{\delta \to 0} \frac{\left| \exp_A(\lambda) - 1 - A(1)\lambda \right| \mu \{x \mid |x| \leq \delta\}^{1/\alpha}}{|\lambda| \, \mu \{x \mid |x| \leq \delta\}^{1/\alpha}} = \left| \frac{\exp_A(\lambda) - 1}{\lambda} - A(1) \right| \, ,$$

which is a contradiction.

Remark 2 Theorems about the differentiability of the deformed exponential are important because of computations like $\frac{d}{d\theta} \exp_A(v(\theta)) = \exp'_A(v(\theta))\dot{v}(\theta)$ are essential for the geometrical theory of statistical models. Several variations in the choice of the combination domain space - image space are possible. Also, one could look at a weaker differentiability property than Frechét differentiability. Our choice is motivated by the results of the following sections. A large class of cases is discussed in [19].

Remark 3 It would also be worth to study the action of the superposition operator on spaces of differentiable functions, for example Gauss-Sobolev spaces of P. Malliavin [15]. If μ is the standard Gaussian measure on $\mathbb{R}^n$, and u is a differentiable function such that $u, \frac{\partial}{\partial x_i} u \in L^2(\mu)$, $i = 1, \ldots, n$, then it follows that $\exp_A(u) \in L^2(\mu)$ as well as $\frac{\partial}{\partial x_i} \exp_A(u) \in L^2(\mu)$, since

$$\frac{\partial}{\partial x_i} \exp_A(u(x)) = A(\exp_A(u(x)) \frac{\partial}{\partial x_i} u(x) \, .$$

We do not pursue this line of investigation here.

3   Deformed Exponential Family Based on $\exp_A$

According to [5, 26], let us define the deformed exponential curve in the space of positive measures on $(\mathbb{X}, \mathscr{X})$ as follows

$$t \mapsto \mu_t = \exp_A(tu + \log_A p) \cdot \mu , \quad u \in L^1(\mu) .$$

We have the following inequality:

$$\exp_A(x + y) \leq \|A\|_\infty x^+ + \exp_A(y).$$

Actually, it is true for $x \leq 0$, as being $\exp_A$ increasing. For $x = x^+ > 0$ the inequality follows from Eq. (2). As a consequence, each μ_t is a finite measure, $\mu_t(\mathbb{X}) \leq t \|A\|_\infty \int u^+ \, d\mu + 1$, with $\mu_0 = p \cdot \mu$. The curve is actually continuous and differentiable in $L^1(\mu)$ because the point-wise derivative of the density $p_t = \exp_A(tu + \log_A(p))$ is $\dot{p}_t = A(p_t)u$ so that $|\dot{p}_t| \leq \|A\|_\infty |u|$. In conclusion $\mu_0 = p \cdot \mu$ and $\dot{\mu}_0 = A(p)u \cdot \mu$.

There are two ways to normalize the density p_t to total mass 1, 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 $t \mapsto \exp_A(tu - \psi(t) + \log_A(p))$. Unlike the standard exponential case, where these two methods lead to the same result, this 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.

Here we use the ideas of [5, 17, 26] to construct deformed non-parametric exponential families. Recall that we are given: the probability space $(\mathbb{X}, \mathscr{X}, \mu)$; the set $\mathscr{P}$ of the positive probability densities and the function A satisfying the conditions set out in Sect. 2. Throughout this section, the density $p \in \mathscr{P}$ will be fixed.

The following proposition is taken from [16] where a detailed proof is given.

Proposition 2   *1. The mapping $L^1(\mu) \ni u \mapsto \exp_A(u + \log_A p) \in L^1(\mu)$ has full domain and is $\|A\|_\infty$ -Lipschitz. Consequently, the mapping*

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

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

2. For each $u \in L^1(\mu)$ there exists a unique constant $K_p(u) \in \mathbb{R}$ such that $\exp_A(u - K_p(u) + \log_A p) \cdot \mu$ is a probability.
3. $K_p(u) = u$ if, and only if, u is constant. In such a case,

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

Otherwise, $\exp_A(u - K_p(u) + \log_A p) \cdot \mu \neq p \cdot \mu$.
4. A density q is of the form $q = \exp_A(u - K_p(u) + \log_A p)$, with $u \in L^1(\mu)$ if, and only if, $\log_A q - \log_A p \in L^1(\mu)$.

5. If
$$\exp_A(u - K_p(u) + \log_A p) = \exp_A(v - K_p(v) + \log_A p) \,,$$

with $u, v \in L^1(\mu)$, then $u - v$ is constant.
6. The functional $K_p \colon L^1(\mu) \to \mathbb{R}$ is translation invariant. More specifically,

$$K_p(u + c) = K_p(u) + c K_p(1)$$

holds for all $c \in \mathbb{R}$.
7. $K_p \colon L^1(\mu) \to \mathbb{R}$ is continuous and convex.

3.1 Escort Density

For each positive density $q \in \overline{\mathcal{P}}$, its *escort density* is defined as

$$\text{escort}\,(q) = \frac{A(q)}{\int A(q)\,d\mu} \,,$$

see [17]. Notice that $0 \le A(q) \le \|A\|_\infty$. In particular, $\widetilde{q} = \text{escort}\,(q)$ is a bounded positive density. Hence, $\text{escort}\left(\overline{\mathcal{P}}\right) \subseteq \overline{\mathcal{P}} \cap L^\infty(\mu)$. Clearly, the inclusion $\text{escort}\,(\mathcal{P}) \subseteq \mathcal{P} \cap L^\infty(\mu)$ is true as well.

Proposition 3 *1. The mapping* $\text{escort} \colon \overline{\mathcal{P}} \to \overline{\mathcal{P}} \cap L^\infty(\mu)$ *is a.s. injective.*
2. *A bounded positive density* $\widetilde{q}$ *is an escort density, i.e.,* $\widetilde{q} \in \text{escort}\left(\overline{\mathcal{P}}\right)$ *if, and only if,*

$$\lim_{\alpha \uparrow \|A\|_\infty} \int A^{-1}\left(\alpha \frac{\widetilde{q}}{\|\widetilde{q}\|_\infty}\right) d\mu \ge 1 \,. \tag{7}$$

3. *Condition* (7) *is fulfilled if* $\mu\left\{\widetilde{q} = \|\widetilde{q}\|_\infty\right\} > 0$. *In particular, every density taking a finite number of different values, i.e., a simple density, is an escort density.*
4. *If* $\widetilde{q}_1 = \text{escort}\,(q_1)$ *is an escort density, and* q_2 *is a bounded positive density such that*

$$\mu\left\{\widetilde{q}_1 > t\,\|\widetilde{q}_1\|_\infty\right\} \le \mu\left\{q_2 > t\,\|q_2\|_\infty\right\}, \quad t > 0 \,,$$

then q_2 *is an escort density as well.*

Proof 1. Let $\text{escort}\,(q_1) = \text{escort}\,(q_2)$ for μ-almost all x. Say, $\int A \circ q_1\,d\mu \ge \int A \circ q_2\,d\mu$. Then $A(q_2(x)) \le A(q_1(x))$, for μ-almost all x. Since A is strictly increasing, it follows $q_2(x) \le q_1(x)$ for μ-almost all x, which, in turn, implies $q_1 = q_2$ μ-a.s. because both μ-integrals are equal to 1. Thus the escort mapping is a.s. injective.
2. Fix a $\widetilde{q} \in \overline{\mathcal{P}} \cap L^\infty(\mu)$, and define the function

$$f(\alpha) = \int A^{-1} \left(\alpha \frac{\widetilde{q}}{\|\widetilde{q}\|_\infty} \right) d\mu, \quad \alpha \in [0, \|A\|_\infty [\, .$$

It is finite, increasing, continuous and $f(0) = 0$. It is clear that the range condition (7) is necessary because $\widetilde{q} = \text{escort}(q)$ implies $q = A^{-1} \left(\left(\int A(q) \, d\mu \right) \widetilde{q} \right)$ and, in turn, $1 = \int A^{-1} \left(\left(\int A(q) \, d\mu \right) \widetilde{q} \right) d\mu$, given that q is a probability density. If we take $\alpha = \int A(q) \, d\mu \, \|\widetilde{q}\|_\infty \le \|A\|_\infty$, the range condition is satisfied. Conversely, if the range condition holds, there exists $\alpha \le \|A\|_\infty$ such that $q = A^{-1} \left(\alpha \frac{\widetilde{q}}{\|\widetilde{q}\|_\infty} \right)$ is a positive probability density whose escort is $\widetilde{q}$.

3. This is a special case of Item 2, in that

$$f(\alpha) = \int A^{-1} \left(\alpha \frac{\widetilde{q}}{\|\widetilde{q}\|_\infty} \right) d\mu \ge A^{-1}(\alpha) \mu \left\{ \widetilde{q} = \|\widetilde{q}\|_\infty \right\} \, .$$

Therefore, $f(\alpha) \uparrow +\infty$, as $\alpha \uparrow \|A\|_\infty$.

4. For each bounded positive density q we have

$$\int A^{-1} \left(\frac{q}{\|q\|_\infty} \right) d\mu = \int_0^{+\infty} \mu \left\{ \frac{q}{\|q\|_\infty} > A(t) \right\} dt =$$

$$\int_0^{\|A\|_\infty} \mu \left\{ \frac{q}{\|q\|_\infty} > s \right\} \frac{1}{A' \left(A^{-1}(s) \right)} ds \, .$$

Now the necessary condition of Item 3. Follows from Item 1. and our assumptions. □

The previous proposition shows that the range of the escort mapping is uniformly dense as it contains all simple densities. Moreover, in the partial order induced by the rearrangement of the normalized density (that is for each q the mapping $t \mapsto \mu \left\{ \frac{q}{\|q\|_\infty} > t \right\}$), it contains the full right interval of each element. But the range of the escort mapping is not the full set of bounded positive densities, unless the σ-algebra $\mathscr{X}$ is generated by a finite partition. To provide an example, consider on the Lebesgue unit interval the densities $q_\delta(x) \propto (1 - x^{1/\delta})$, $\delta > 0$, and $A(x) = x/(1 + x)$. The density q_δ turns out to be an escort if, and only if, $\delta \le 1$.

3.2 Gradient of the Normalization Operator K_p

Proposition 2 shows that the functional K_p is a global solution of an equation. We now study its local properties by the implicit function theorem as well as the related subgradients of the convex function K_p. We refer to [7, Part I] for the general theory of convex functions in infinite dimension.

For every $u \in L^1(\mu)$, let us write

$$q(u) = \exp_A(u - K_p(u) + \log_A p) \tag{8}$$

while $\widetilde{q}(u) = \text{escort}(q(u))$ denotes its escort density.

Proposition 4 *1. The functional $K_p \colon L^1(\mu) \to \mathbb{R}$ is Gateaux-differentiable with derivative*

$$\frac{d}{dt}K_p(u + tv)\bigg|_{t=0} = \int v\widetilde{q}(u)\,d\mu\,.$$

It follows that $K_p \colon L^1(\mu) \to \mathbb{R}$ is monotone and globally Lipschitz.

2. For every $u, v \in L^1(\mu)$, the inequality

$$K_p(u + v) - K_p(u) \geq \int v\widetilde{q}(u)\,d\mu$$

holds, i.e., the density $\widetilde{q}(u) \in L^\infty(\mu)$ is the unique subgradient of K_p at u.

Proof 1. Consider the equation

$$F(t, \kappa) = \int \exp_A(u + tv - \kappa + \log_A p)\,d\mu - 1 = 0, \quad t, \kappa \in \mathbb{R}\,,$$

so that $\kappa = K_p(u + tv)$. Derivations under the integral hold by virtue of the bounds

$$\left|\frac{\partial}{\partial t}\exp_A(u + tv - \kappa + \log_A p)\right| =$$
$$\left|A(\exp_A(u + tv - \kappa + \log_A p))v\right| \leq \|A\|_\infty |v|$$

and

$$\left|\frac{\partial}{\partial \kappa}\exp_A(u + tv - \kappa + \log_A p)\right| = \left|A(\exp_A(u + tv - \kappa + \log_A p))\right| \leq \|A\|_\infty\,.$$

Furthermore, the partial derivative with respect to κ is never zero. Thanks to the implicit function theorem, there exists the derivative $(d\kappa/dt)_{t=0}$ which is the desired Gateaux derivative. Since $\widetilde{q}(u)$ is positive and bounded, K_p is monotone and globally Lipschitz.

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

$$\exp_A(u + v - K_p(u + v) + \log_A p) \geq q + A(q)(v - (K_p(u + v) - K_p(v)))\,,$$

where $q = \exp_A(u - K_p(u) + \log_A p)$. If we take μ-integral of both sides,

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

Isolating the increment $K_p(u + v) - K_p(v)$, the desired inequality obtains. Therefore, $\widetilde{q}(u)$ is a subgradient of K_p at u. From Item 1. we deduce that $\widetilde{q}(u)$ is the unique subgradient and further $\widetilde{q}(u)$ is the Gateaux differential of K_p at u. $\square$

We can also establish Fréchet-differentiability of the functional, under more stringent assumptions.

Proposition 5 *Let $\alpha \geq 2$.*

1. *The superposition operator*

$$L^\alpha(\mu) \ni v \mapsto \exp_A(v + \log_A p) \in L^1(\mu)$$

   is continuously Fréchet differentiable with derivative

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

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

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

   is continuously Fréchet differentiable with derivative

$$dK_p(v) = \left(h \mapsto \int h\widetilde{q}(v) \, d\mu\right),$$

   where $\widetilde{q}(u) = \text{escort}(q(u))$.

Proof 1. Setting $\beta = 1$ in Proposition 1, we get easily the assertion. It remains just to check that the Fréchet derivative is continuous, i.e., that the Fréchet derivative is a continuous map $L^\alpha(\mu) \to \mathscr{L}(L^\alpha(\mu), L^1(\mu))$. If $\|h\|_{L^\alpha(\mu)} \leq 1$ and $v, w \in L^\alpha(\mu)$ we have

$$\int \left|(A[\exp_A(v + \log_A p)] - A[\exp_A(w + \log_A p)])h\right| \, d\mu$$
$$\leq \|A[\exp_A(v + \log_A p) - A[\exp_A(w + \log_A p)]\|_{L^\sigma(\mu)} \,,$$

where $\sigma = \alpha/(\alpha - 1)$ is the conjugate exponent of α. On the other hand,

$$\|A[\exp_A(v + \log_A p) - A[\exp_A(w + \log_A p)]\|_{L^\sigma(\mu)}$$
$$\leq \|A'\|_\infty \|A\|_\infty \|v - w\|_{L^\sigma(\mu)}$$

and so the map $L^\alpha(\mu) \to \mathscr{L}(L^\alpha(\mu), L^1(\mu))$ is continuous whenever $\alpha \geq \sigma$, i.e., $\alpha \geq 2$.

2. Fréchet differentiability of K_p is a consequence of the Implicit Function Theorem in Banach spaces, see [6], applied to the C^1-mapping

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

The value of the derivative is given by Proposition 4.   □

4 Deformed Divergence

In analogy with the standard exponential case, define the A-divergence between probability densities as

$$D_A(q \| p) = \int \left(\log_A q - \log_A p \right) \text{escort}(q) \, d\mu, \quad \text{for } q, p \in \mathscr{P} \ .$$

Since $\log_A$ is strictly concave with derivative $1/A$, we have

$$\log_A(x) \le \log_A(y) + \frac{1}{A(y)} (x - y)$$

for all $x, y > 0$ and with equality if, and only if, $x = y$. Hence

$$A(y) \left(\log_A(y) - \log_A(x) \right) \ge y - x \ . \tag{9}$$

It follows in particular that $D_A(\cdot \| \cdot)$ is a well defined, possibly extended valued, function.

Observe further that by Proposition 2, $\log_A q - \log_A p \in L^1(\mu)$, and so $D_A(q \| p) < \infty$, whenever $q = q(u)$.

The binary relation D_A is a faithful divergence in that it satisfies the following Gibbs' inequality.

Proposition 6 *It holds $D_A(q \| p) \ge 0$ and $D_A(q \| p) = 0$ if and only if $p = q$.*

Proof From inequality (9) it follows

$$D_A(q \| p) = \frac{1}{\int A(q) \, d\mu} \int \left(\log_A q - \log_A p \right) A(q) \, d\mu$$

$$\ge \frac{1}{\int A(q) \, d\mu} \int (q - p) \, d\mu = 0.$$

Moreover, equality holds if and only if $p = q$ μ-a.e.   □

There are other alternative definitions that may fully candidate to be a divergence measure. For instance:

$$I_A(q \| p) = - \int \log_A(p/q)q \; d\mu.$$

or also

$$\widetilde{D}_A(q \| p) = \int A(q/p) \log_A(p/q)p \; d\mu.$$

By means of the concavity of $\log_A$, it is not difficult to check that both satisfy Gibbs' condition of Proposition 6, as well as they equal the Kullback–Leibner functional in the non-deformed case. Observe further that the functional $I_A(q \| p)$ is closely related to Tallis' divergence (see [25] and also [14]). In fact, if one replaces $\log_A$ with the q-logarithm, one gets just Tallis' q-divergence.

However our formulation for the divergence is motivated by the structure of the deformed exponential representation. As it will be now seen, our definition of divergence is more adapted to the present setting and it turns out be closely related to the normalizing operator.

In the equation

$$q = \exp_A(u - K_p(u) + \log_A p), \quad u \in L^1(\mu) \,, \; q \in \mathscr{P} \,, \tag{10}$$

the random variable u is identified up to an additive constant for any fixed density q. There are at least two options for selecting an interesting representative member in the equivalence class.

One option is to impose the further condition $\int u\widetilde{p} \, d\mu = 0$, where $\widetilde{p} = \text{escort}(p)$, the integral being well defined, given that the escort density is bounded. This restriction provides a unique element u_q. On the other hand, if we solve Eq. (10) with respect to $u - K(u)$, we get the desired relation:

$$K_p(u_q) = E_{\widetilde{p}}\big[\log_A p - \log_A q\big] = D_A(p \| q), \tag{11}$$

where $u = u_q$ is uniquely characterized by the two equations: $E_{\widetilde{p}}[u] = 0$ and $q = \exp_A(u - K_p(u) + \log_A p)$.

Observe further that Eq. (11) entails the relation

$$K_p(u) = D_A(p \| q(u)) \quad \forall u \in L^1(\mu) \,.$$

The previous choice is that followed in the construction of the non-parametric exponential manifold, see [22, 23].

With regard to the non-deformed case, Eq. (11) yields the Kulback-Leibler divergence with p and q exchanged, with respect to what is considered more natural in Statistical Physics, see for example the comments [13].

For this purpose, we undertake another choice for the random variable in the equivalence class. More specifically, in Eq. (10) the random variable u will be now centered with respect to $\widetilde{q} = \text{escort}(q)$, i.e., $E_{\widetilde{q}}[u] = 0$.

To avoid confusion let us rewrite Eq. (10) as follows and where for convenience the function K_p is replaced with $H_p = -K_p$:

$$q = \exp_A(v + H_p(v) + \log_A p), \quad v \in L^1(\mu), \quad E_{\widetilde{q}}[v] = 0, \tag{12}$$

so that

$$H_p(v_q) = E_{\widetilde{q}}\left[\log_A q - \log_A p\right] = D_A(q\|p),$$

where $v = v_q$ is the solution to the two equations $E_{\widetilde{q}}[v] = 0$ and $q = \exp_A(v + H_p(v) + \log_A p)$. There are hence two notable representations of the same probability density q:

$$q = \exp_A(u - K_p(u) + \log_A p) = \exp_A(v + H_p(v) + \log_A p)$$

which implies $u_q - v_q = K_p(u_q) + H_p(v_q)$. This, in turn, leads to

$$-E_{\widetilde{p}}\left[v_q\right] = E_{\widetilde{q}}\left[u_q\right] = K_p(u_q) + H_p(v_q) = K_p(u_q) - K_p(v_q).$$

This provides the following remarkable relation

$$H_p(v_q) = E_{\widetilde{q}}\left[u_q\right] - K_p(u_q). \tag{13}$$

4.1 Variational Formula

We now present a variational formula in the spirit of the classical one by Donsker-Varadhan. Next proposition provides the convex conjugate of K_p, in the duality $L^\infty(\mu) \times L^1(\mu)$.

In what follows, the operator $\eta \mapsto \hat{\eta}$ denotes the inverse of the escort operator, i.e., $\eta = \text{escort}(\hat{\eta})$. In the light of the results established in Sect. 3.1, this operator maps a dense subset of $\overline{\mathscr{P}} \cap L^\infty(\mu)$ onto $\overline{\mathscr{P}}$.

Proposition 7 *1. The convex conjugate function of K_p:*

$$K_p^*(w) = \sup_{u \in L^1(\mu)} \left(\int wu \, d\mu - K_p(u)\right), \quad w \in L^\infty(\mu) \tag{14}$$

has domain contained into $\overline{\mathscr{P}} \cap L^\infty(\mu)$. More precisely,

$$\text{escort}(\mathscr{P}) \subseteq \text{dom} K_p^* \subseteq \overline{\mathscr{P}} \cap L^\infty(\mu).$$

2. *$K_p^*(w) \geq 0$ for all $w \in L^\infty(\mu)$. For any $\eta \in \text{escort}(\mathscr{P})$, the conjugate $K_p^*(\eta)$ is given by the* Legendre transform:

$$K_p^*(\eta) = \int \eta \, u_{\hat{\eta}} \, d\mu - K_p(u_{\hat{\eta}}) .$$

So that $\quad K_p^*(\eta) = H_p(v_{\hat{\eta}}) = D_A(\hat{\eta} \| p);$ *equivalently:*

$$K_p^*(\text{escort}\,(q)) = D_A(q \| p) \quad \forall p, q \in L^1(\mu).$$

3. *It holds the inversion formula*

$$K_p(u) = \max_{\eta \in \text{escort}(\mathscr{P})} \left(\int \eta u \, d\mu - D_A(\hat{\eta} \| p) \right)$$

$$= \max_{q \in \mathscr{P}} \left(\int \text{escort}\,(q) \, u \, d\mu - D_A(q \| p) \right), \quad \forall u \in L^1(\mu).$$

Proof   1. It follows from the fact that K_p is monotone and translation invariant. Let us first suppose $w \notin L_+^\infty(\mu)$. That means that

$$\int w \chi_C \, d\mu < 0$$

is true for some indicator function χ_C. If we consider the cone generated by the function $-\chi_C$, we can write

$$K_p^*(w) \geq \sup_{u \in \, cone(-\chi_C)} \left(\int wu \, d\mu - K_p(u) \right) \geq \sup_{u \in \, cone(-\chi_C)} \int wu \, d\mu = +\infty,$$

since $K_p(u) \leq 0$ when $u \in cone(-\chi_C)$. Now consider the case in which $w \geq 0$. If we set $u = \lambda \in \mathbb{R}$, we have $K_p(\lambda) = \lambda$ and consequently

$$K_p^*(w) \geq \sup_{\lambda \in \mathbb{R}} \left(\lambda \int w \, d\mu - \lambda \right) . \tag{15}$$

This sup is $+\infty$, unless $\int w \, d\mu = 1$. Hence, $K_p^*(w) < \infty$ implies $w \in \overline{\mathscr{P}}$. Summarizing, the domain of K_p^* is contained into $\overline{\mathscr{P}} \cap L^\infty(\mu)$, and this proves one of the two claimed inclusions. The other one will be a direct consequence of the next point.

2. Equation (15) implies $K_p^* \geq 0$. By Proposition 4 the concave and Gateaux differentiable function $u \mapsto \int \eta u \, d\mu - K_p(u)$ has derivative at u given by $\eta - dK_p(u) = \eta - \text{escort}\,(q(u))$, where $q(u) = \exp_A(u - K_p(u) + \log_A p)$. Under our assumptions, the derivative vanishes at $u = u_{\hat{\eta}}$ and the sup in the definition of K_p^* is attained at that point. The maximum value is $K_p^*(\eta) = \int \eta u \, d\mu - K_p(u)$, by setting $u = u_{\hat{\eta}}$.
   The last formula follows straightforward from Eq. (13).

3. For a well-known property of Fenchel–Moreau duality theory, we have:

$$K_p(u) \geq \int wu \, d\mu - K_p^*(w) \quad \forall u \in L^1(\mu), \quad \forall w \in L^\infty(\mu)$$

$$K_p(u) = \int wu \, d\mu - K_p^*(w) \iff w \in \partial K_p(u).$$

Clearly in our case $\partial K_p(u)$ is a singleton and the image of ∂K_p is the set escort $(\mathscr{P})$. Therefore

$$K_p(u) = \max_{w \in \text{escort}(\mathscr{P})} \left(\int wu \, d\mu - K_p^*(w) \right).$$

By Item 2 the desired inversion formula obtains. $\square$

5   Hilbert Bundle Based on $\exp_A$

We shall introduce the Hilbert manifold of probability densities as defined in [18, 19]. A slightly more general set-up than the one used in that reference will be introduced. By means of a general A function, we provide an atlas of charts, and define a linear bundle as an expression of the tangent space.

Let $\mathscr{P}(\mu)$ denote the set of all μ-densities on the probability space $(\mathbb{X}, \mathscr{X}, \mu)$ of the kind

$$q = \exp_A(u - K_1(u)), \quad u \in L^2(\mu), \quad E_\mu[u] = 0. \tag{16}$$

Notice that $1 \in \mathscr{P}(\mu)$ because we can take $u = 0$.

Proposition 8   *1. $\mathscr{P}(\mu)$ is the set of all densities q such that $\log_A q \in L^2(\mu)$, in which case $u = \log_A q - E_\mu\left[\log_A q\right]$.*
  2. If in addition $A'(0+) > 0$, then $\mathscr{P}(\mu)$ is the set of all densities q such that both q and $\log q$ are in $L^2(\mu)$.
  3. Let $A'(0+) > 0$. On a product space with reference probability measures μ_1 and μ_2, and densities respectively q_1 and q_2, we have $q_1 \in \mathscr{P}(\mu_1)$ and $q_2 \in \mathscr{P}(\mu_2)$ if, and only if, $q_1 \otimes q_2 \in \mathscr{P}(\mu_1 \otimes \mu_2)$.

Proof   1. From Eq. (16), it follows $\log_A q = u - K_1(u) \in L^2(\mu)$, provided $u \in L^2(\mu)$. Conversely, let $\log_A q \in L^2(\mu)$. Equation (16) yields

$$u = \log_A q - K_1(u) \quad and \quad K_1(u) = -\log_A q.$$

Therefore $u = \log_A q - E_\mu\left[\log_A q\right]$ and $u \in L^2(\mu)$.
  2. Write
$$\left|\log_A q\right|^2 = \left|\log_A q\right|^2 (q < 1) + \left|\log_A q\right|^2 (q \geq 1)$$

and use the bounds of Eqs. (3) and (4) to get

$$E_\mu \left[\left| \log_A q \right|^2 \right] \leq \frac{1}{\alpha_2^2} E_\mu \left[|\log q|^2 \, (q < 1) \right] + \frac{1}{A(1)^2} E_\mu \left[|q - 1|^2 \, (q \geq 1) \right] \leq$$

$$\frac{1}{\alpha_2^2} E_\mu \left[|\log q|^2 \right] + \frac{1}{A(1)^2} E_\mu \left[q^2 \right] .$$

We deduce that the two conditions q and $\log q$ in $L^2(\mu)$ imply $\log_A q \in L^2(\mu)$. Conversely, let $\log_A q \in L^2(\mu)$. By means of the other two bounds (recall that $\alpha_1 > 0$) we have too

$$E_\mu \left[\left| \log_A q \right|^2 \right] \geq \frac{1}{\alpha_1^2} E_\mu \left[|\log q|^2 \, (q < 1) \right] + \frac{1}{\|A\|_\infty^2} E_\mu \left[(q - 1)^2 (q \geq 1) \right] .$$

Consequently, $E_\mu \left[(q - 1)^2 (q \geq 1) \right] < +\infty$. This in turn gives $E_\mu \left[(q - 1)^2 \right] < +\infty$, and so $q \in L^2(\mu)$.

Once again, the previous inequality provides the condition $E_\mu \left[|\log q|^2 \, (q < 1) \right] < +\infty$. On the other hand, $E_\mu \left[|\log q|^2 \, (q \geq 1) \right] < +\infty$ since $|\log q|^2 \, (q \geq 1) \leq (q - 1)^2 (q \geq 1)$. Therefore, $\log q \in L^2(\mu)$.

3. We deduce by the previous item that: $q_1 \otimes q_2 \in \mathscr{P}(\mu_1 \otimes \mu_2)$ if and only if both $q_1 \otimes q_2$ and $\log(q_1 \otimes q_2)$ are in $L^2(\mu_1 \otimes \mu_2)$.

The first condition is equivalent to both $q_1 \in L^2(\mu_1)$ and $q_2 \in L^2(\mu_2)$. The second one is equivalent to $\log q_1 + \log q_2 \in L^2(\mu_1 \otimes \mu_2)$. On the other hand, we have

$$E_{\mu_1 \otimes \mu_2} \left[(\log q_1 + \log q_2)^2 \right] =$$
$$E_{\mu_1} \left[\log^2 q_1 \right] + E_{\mu_2} \left[\log^2 q_2 \right] + 2 \, E_{\mu_1} \left[\log q_1 \right] E_{\mu_2} \left[\log q_2 \right]. \tag{17}$$

By Eq. (17), $q_1 \in \mathscr{P}(\mu_1)$ and $q_2 \in \mathscr{P}(\mu_2)$ imply $q_1 \otimes q_2 \in \mathscr{P}(\mu_1 \otimes \mu_2)$. Conversely, assume $q_1 \otimes q_2 \in \mathscr{P}(\mu_1 \otimes \mu_2)$. This implies that it holds, $E_{\mu_1 \otimes \mu_2} \left[(\log q_1 + \log q_2)^2 \right] < +\infty$. Since $E_{\mu_i} \left[\log q_i \right] \leq E_{\mu_1} \left[q_i - 1 \right] = 0$. We have $E_{\mu_1} \left[\log q_1 \right] E_{\mu_2} \left[\log q_2 \right] \geq 0$. In view of Eq. (17), we can infer that $q_1 \in \mathscr{P}(\mu_1)$ and $q_2 \in \mathscr{P}(\mu_2)$ ☐

We proceed now to define an Hilbert bundle with base $\mathscr{P}(\mu)$. The notion of Hilbert bundle has been introduced in Information Geometry by [1]. We are here using an adaptation to the A-exponential of arguments elaborated by [8, 21]. Notice that the construction depends in a essential way on the specific conditions we are assuming for the present class of deformed exponential.

At each $q \in \mathscr{P}(\mu)$ the escort density $\widetilde{q}$ is bounded, so that we can define the fiber given by the Hilbert spaces

$$\mathscr{H}_q = \left\{ u \in L^2(\mu) | E_{\widetilde{q}} [u] = 0 \right\}$$

with scalar product $\langle u, v \rangle_q = \int uv \, d\mu$. The Hilbert bundle is

$$H\mathscr{P}(\mu) = \left\{ (q, u) | q \in \mathscr{P}(\mu), u \in \mathscr{H}_q \right\} .$$

For each $p, q \in \mathscr{P}(\mu)$ the mapping $\mathbb{U}_p^q u = u - E_{\widetilde{q}}[u]$ is a continuous linear mapping from $\mathscr{H}_p$ to $\mathscr{H}_q$. Moreover, $\mathbb{U}_q^r \mathbb{U}_p^q = \mathbb{U}_p^r$. In particular, $\mathbb{U}_q^p \mathbb{U}_p^q$ is the identity on $\mathscr{H}_p$ and so $\mathbb{U}_p^q$ is an isomorphism of $\mathscr{H}_p$ onto $\mathscr{H}_q$.

In the next proposition an affine atlas of charts is constructed in order to define our Hilbert bundle which is an expression of the tangent bundle. The velocity of a curve $t \mapsto p(t) \in \mathscr{P}(\mu)$ is given in the Hilbert bundle by the so called A-score that, in our case, takes the form $A(p(t))^{-1}\dot{p}(t)$, where $\dot{p}(t)$ is computed in $L^1(\mu)$.

The following proposition is taken from [16] where a detailed proof is presented.

Proposition 9 *1. Fix $p \in \mathscr{P}(\mu)$. A positive density $q \in \mathscr{P}(\mu)$ if and only if*

$$q = \exp_A(u - K_p(u) + \log_A p), \ \text{with } u \in L^2(\mu) \text{ and } E_{\widetilde{p}}[u] = 0.$$

2. For any fixed $p \in \mathscr{P}(\mu)$ the mapping $s_p \colon \mathscr{P}(\mu) \to \mathscr{H}_p$ defined by

$$q \mapsto \log_A q - \log_A p + D_A(p\|q)$$

is injective and surjective, with inverse $e_p(u) = \exp_A(u - K_p(u) + \log_A p)$.
3. The atlas $\{s_p | p \in \mathscr{P}(\mu)\}$ is affine with transitions

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

4. The velocity of the differentiable curve $t \mapsto p(t) \in \mathscr{P}(\mu)$ in the chart s_p is $ds_p(p(t))/dt \in \mathscr{H}_p$. Conversely, given any $u \in \mathscr{H}_p$, the curve

$$p \colon t \mapsto \exp_A(tu - K_p(tu) + \log_A p)$$

satisfies $p(0) = p$ and has velocity u at $t = 0$, expressed in the chart s_p. If the velocity of a curve is $t \mapsto \dot{u}(t)$, in a chart s_p, then $\mathbb{U}_p^q \dot{u}(t)$ is its velocity in the chart s_q.
5. If $t \mapsto p(t) \in \mathscr{P}(\mu)$ 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 $t \mapsto s_{p(t)}$.

We end here our discussion of the geometry of the Hilbert bundle, because our aim is limited to show the applicability of the analytic results obtained in the previous section. A detailed discussion of the relevant geometric objects e.g., the affine covariant derivative, is not attempted here.

6   Final Remarks

A non-parametric Hilbert manifold based on a deformed exponential representation of positive densities has been firstly introduced by N.J. Newton [18, 19]. We have derived regularity properties of the normalizing functional K_p and discussed the relevant Fenchel conjugation. In particular, we have discussed some properties of the escort mapping and a form of the divergence that appears to be especially adapted to our set-up. We have taken a path different from that of N.J. Newton original presentation. We allow for a manifold defined by an atlas containing charts centered at each density in the model. In conclusion, we have discussed explicitly a version of the Hilbert bundle as a family of codimension 1 sub-vector spaces of the basic Hilbert space.

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. (1987)
2. Amari, S., Nagaoka, H.: Methods of Information Geometry. American Mathematical Society, Providence (2000). Translated from the 1993 Japanese original by Daishi Harada
3. Ambrosetti, A., Prodi, G.: A Primer of Nonlinear Analysis. Cambridge Studies in Advanced Mathematics, vol. 34. Cambridge University Press, Cambridge (1993)
4. Appell, J., Zabrejko, P.P.: Nonlinear Superposition Operators. Cambridge Tracts in Mathematics, vol. 95. Cambridge University Press, Cambridge (1990). https://doi.org/10.1017/CBO9780511897450
5. Ay, N., Jost, J., Lê, H.V., Schwachhöfer, L.: Information Geometry. Springer, Berlin (2017)
6. Dieudonné, J.: Foundations of Modern Analysis. Academic Press, New York (1960)
7. Ekeland, I., Témam, R.: Convex Analysis and Variational Problems. Classics in Applied Mathematics, vol. 28, English edn. Society for Industrial and Applied Mathematics (SIAM) (1999). https://doi.org/10.1137/1.9781611971088. Translated from the French
8. Gibilisco, P., Pistone, G.: Connections on non-parametric statistical manifolds by Orlicz space geometry. IDAQP **1**(2), 325–347 (1998)
9. Kaniadakis, G.: Non-linear kinetics underlying generalized statistics. Phys. A **296**(3–4), 405–425 (2001)
10. Kaniadakis, G.: Statistical mechanics in the context of special relativity. Phys. Rev. E **66**, 056, 125 1–17 (2002)
11. Kaniadakis, G.: Statistical mechanics in the context of special relativity. ii. Phys. Rev. E **72**(3), 036,108 (2005). https://doi.org/10.1103/PhysRevE.72.036108
12. Kass, R.E., Vos, P.W.: Geometrical Foundations of Asymptotic Inference. Wiley Series in Probability and Statistics: Probability and Statistics. Wiley, New York (1997). https://doi.org/10.1002/9781118165980. A Wiley-Interscience Publication
13. Landau, L.D., Lifshits, E.M.: Course of Theoretical Physics. Statistical Physics, vol. V, 3rd edn. Butterworth-Heinemann, Oxford (1980)

14. Loaiza, G., Quiceno, H.R.: A q-exponential statistical Banach manifold. J. Math. Anal. Appl. **398**(2), 466–476 (2013). https://doi.org/10.1016/j.jmaa.2012.08.046

15. Malliavin, P.: Integration and Probability. Graduate Texts in Mathematics, vol. 157. Springer-Verlag (1995). With the collaboration of Hlne Airault, Leslie Kay and Grard Letac. Edited and translated from the French by Kay, With a foreword by Mark Pinsky

16. Montrucchio, L., Pistone, G.: Deformed exponential bundle: the linear growth case. In: Nielsen, F., Barbaresco, F. (eds.) Geometric Science of Information. LNCS, vol. 10589, pp. 239–246. Springer (2017). Proceedings of the Third International Conference, GSI 2017, Paris, France, November 7–9

17. Naudts, J.: Generalised Thermostatistics. Springer-Verlag London Ltd. (2011). https://doi.org/10.1007/978-0-85729-355-8

18. Newton, N.J.: An infinite-dimensional statistical manifold modelled on Hilbert space. J. Funct. Anal. **263**(6), 1661–1681 (2012). https://doi.org/10.1016/j.jfa.2012.06.007

19. Newton, N.J.: Infinite-dimensional statistical manifolds based on a balanced chart. Bernoulli **22**(2), 711–731 (2016). https://doi.org/10.3150/14-BEJ673

20. Pistone, G.: κ-exponential models from the geometrical viewpoint. Eur. Phys. J. B Condens. Matter Phys. **71**(1), 29–37 (2009). https://doi.org/10.1140/epjb/e2009-00154-y

21. Pistone, G.: Nonparametric information geometry. In: Nielsen, F., Barbaresco, F. (eds.) Geometric Science of Information. Lecture Notes in Computer Science, vol. 8085, pp. 5–36. Springer, Heidelberg (2013). Proceedings First International Conference, GSI 2013 Paris, France, August 28–30

22. Pistone, G., Rogantin, M.: The exponential statistical manifold: mean parameters, orthogonality and space transformations. Bernoulli **5**(4), 721–760 (1999)

23. Pistone, G., Sempi, C.: An infinite-dimensional geometric structure on the space of all the probability measures equivalent to a given one. Ann. Stat. **23**(5), 1543–1561 (1995)

24. Shima, H.: The Geometry of Hessian Structures. World Scientific Publishing Co. Pte. Ltd., Hackensack, NJ (2007). https://doi.org/10.1142/9789812707536

25. Tsallis, C.: Possible generalization of Boltzmann-Gibbs statistics. J. Stat. Phys. **52**(1–2), 479–487 (1988)

26. Vigelis, R.F., Cavalcante, C.C.: On ϕ-families of probability distributions. J. Theor. Probab. **26**, 870–884 (2013)