



# A two-parameter family of non-parametric, deformed exponential manifolds

Nigel J. Newton<sup>1</sup> 

Received: 1 September 2022 / Revised: 1 September 2022 / Accepted: 13 November 2022 /  
Published online: 6 December 2022  
© The Author(s) 2022

## Abstract

We construct a new family of non-parametric statistical manifolds by means of a two-parameter class of deformed exponential functions, that includes functions with power-law, linear and sublinear rates of growth. The manifolds are modelled on weighted, mixed-norm Sobolev spaces that are especially suited to this purpose, in the sense that an important class of nonlinear superposition operators (those used in the construction of divergences and tensors) act continuously on them. We analyse variants of these operators, that map into “subordinate” Sobolev spaces, and evaluate the associated gain in regularity. With appropriate choice of parameter values, the manifolds support a large variety of the statistical divergences and entropies appearing in the literature, as well as their associated tensors, eg. the Fisher-Rao metric. Manifolds of finite measures and probability measures are constructed; the latter are shown to be smoothly embedded submanifolds of the former.

**Keywords** Banach manifold · Fisher-Rao metric · Information Theory · Log-Sobolev inequality · Non-parametric statistics · Sobolev spaces

**Mathematics Subject Classification** 46N30 · 60D05 · 60H15 · 62B10 · 93E11

## 1 Introduction

Great progress has been made during the last five decades on the theory of information geometry, and its application in many scientific fields. The fundamental parametric theory is well developed, and is treated pedagogically in a number of texts (See, for example, [1, 3, 5, 9, 12]). The non-parametric theory, on the other hand, is largely

---

Communicated by Jürgen Jost.

---

✉ Nigel J. Newton  
nigeljnewton@gmail.com

<sup>1</sup> University of Essex, Colchester CO4 3SQ, UK

to be found in a series of research papers. A notable exception is the text [2], which treats parametric and non-parametric theories in a unified way. The step from the parametric to the non-parametric setting is not an easy one, since it introduces the infinite-dimensional spaces of Functional Analysis.

The parametric exponential model is arguably the nucleus of the subject. Its extension to the non-parametric setting was accomplished by G. Pistone and his co-workers in the fundamental series of papers [4, 7, 18, 19]. The manifolds there constructed are “maximally inclusive” in a precise sense, and various statistical divergences, including the Amari  $\alpha$ -divergences, for  $\alpha$  in the interval  $[-1, 1]$ , are smooth on them. As with parametric exponential manifolds, the log of the density is used as a chart. This requires a model space with a particularly strong topology: the exponential Orlicz space, which has a number of disadvantages. Since its publication, several variations of the exponential Orlicz manifold have been developed. In [11], the exponential function was replaced by the Tsallis  $q$ -deformed exponential, which has an important interpretation in statistical mechanics. (See [20], and Chapter 7 in [13].) The model space used is  $L^\infty$ , which significantly restricts membership of the manifold constructed. A large class of deformed exponential functions (the “ $\varphi$ -functions”) was used in [21, 22] to construct inclusive manifolds of probability measures, in which the model spaces are Musielak-Orlicz spaces.

The constructions in these references begin with the tangent space at a generic point,  $P$ , of a set of measures. A representation of tangent vectors, derived from the (deformed) logarithm, is then used to construct a *local* chart, which naturally maps to a model space defined in terms of  $P$ . However, the model spaces required in this approach can be difficult to use in practice. A different approach was taken in [14, 15], where a *global* chart was used with the specific deformed logarithm  $\log_d = y - 1 + \log y$  to construct an inclusive manifold modelled on Lebesgue  $L^p$  spaces (including the Hilbert space  $L^2$ ). The corresponding deformed exponential has bounded derivatives of all orders; a property that has a number of advantages. Both the probability density and its (non-deformed) log (objects of central importance to information geometry) belong to the model space and, considered as superposition operators mapping into this space, are continuous.

The sample space in all these manifolds is an abstract probability space: a set of “outcomes”, a class of measurable subsets of these outcomes, and a probability measure attaching a number between 0 and 1 to each subset. This has the advantage of generality: the sample space could be  $\mathbb{R}^d$ , or the path space of a stochastic process, ... However, topologies, metrics and linear structures on the sample space play important roles in most applications, including the theory of partial differential equations. A natural direction for research in the non-parametric theory is to specialise the manifolds outlined above to such problems by incorporating the topology of the sample space in the manifolds. One way of achieving this is to use model spaces of Sobolev type. This was carried out in the context of the exponential Orlicz manifold in [10], where it was applied to the spatially homogeneous Boltzmann equation. It was carried out in the context of the  $L^p$  manifolds in [17]. The resulting fusion of information and sample space topologies is mutually beneficial. For example, it was shown in [17] that log-Sobolev embedding strengthens the topology of the raw  $L^p$  manifolds in a useful way.

This paper takes the approach of [14, 15, 17] further, by constructing a new class of non-parametric manifolds based on a two-parameter deformed exponential, dubbed the  $\eta$ -exponential. The paper has two primary aims: (i) to provide manifolds on which a wider class of divergences and entropy functions can be accommodated; (ii) to refine the Sobolev space methods of [17] in order to increase the degree of smoothness they confer on these quantities. Regarding the first aim, there is a vast literature on the importance of different divergences and entropies to particular branches of science, a full review of which is beyond the scope of this article. Let us mention, however, the special volume (13) of *Entropy* on applications of the Tsallis  $q$ -divergences and entropies. The review article by Tsallis [20], in particular, cites many applications in which the  $q$  parameter should be strictly greater than or strictly less than the value  $q = 1$  of the Boltzmann-Gibbs theory. In the context of Amari's  $\alpha$ -divergences, this translates to values of  $\alpha$  both greater than and less than 1. The author was motivated, in particular, by the study of multi-objective measures of error in *nonlinear filtering* that lead naturally to divergences with  $\alpha < -1$ , [16].

The two-parameter  $\eta$ -exponential we use corresponds to a deformed exponential introduced in [8], but is reparametrised into the Amari setting. The manifolds accommodate  $\alpha$ -divergences and entropies over a range of  $\alpha$  values, but are especially suited to those for which  $\alpha \in [\eta_-, \eta_+]$ , for the chosen parameters  $-\infty < \eta_- < 1 \leq \eta_+ < \infty$ . The parameter values  $\pm 1$  yield the linear-growth deformed exponential of [14, 15, 17]; values of  $\eta_-$  other than  $-1$  yield deformed exponentials with power law or sublinear growth. For a more detailed account of deformed exponentials, and their use in Statistical Mechanics, the reader is referred to [13].

The paper is structured as follows. Section 2 introduces the model Sobolev spaces, expanding considerably on the material in [17]. It also introduces a new class of "subordinate" Sobolev spaces, which are later used in the analysis of superposition operators derived from Amari's  $\alpha$ -embedding maps. (The latter can be used in the analysis of divergences and entropies.) Section 3 introduces the two-parameter deformed exponential and uses it to construct manifolds of finite measures. Section 4 then shows that the subsets of *probability* measures are smoothly embedded submanifolds of those of Sect. 3.

## 2 The Sobolev spaces

The manifolds are modelled on mixed-norm, weighted Sobolev spaces generalising those defined in [17]. The spaces are based on a *reference* probability measure  $\mu$  on the sample space  $\mathbb{R}^d$ . This takes the form

$$\mu(dx) = r(x)dx = \exp(l_r(x))dx, \quad (1)$$

where  $l_r : \mathbb{R}^d \rightarrow \mathbb{R}$  is a continuous function such that  $\mu(\mathbb{R}^d) = 1$ . Stronger results can be obtained with the following additional hypothesis on  $l_r$ .

- (E) The log-density  $l_r$  is constructed as follows. Let  $\theta : [0, \infty) \rightarrow [0, \infty)$  be a strictly increasing, convex function that is twice continuously differentiable on  $(0, \infty)$ , is

such that  $\lim_{z \downarrow 0} \theta'(z) < \infty$ ,  $-\sqrt{\theta}$  is convex and, for some  $t \in (1, 2]$ ,

$$\theta(z) = \begin{cases} 0 & \text{if } z = 0 \\ c + z^t & \text{if } z \geq z_0 \end{cases}, \quad \text{where } z_0 \geq 0, \text{ and } c \in \mathbb{R}. \tag{2}$$

$l_r$  then takes the special form

$$l_r(x) := \sum_i (C - \theta(|x_i|)), \tag{3}$$

where  $C \in \mathbb{R}$  is such that  $\mu(\mathbb{R}^d) = 1$ . (Some examples are given in [17] including the Gaussian case, in which  $c = z_0 = 0$  and  $t = 2$ .)

The model spaces used in the construction of the manifolds comprise measurable functions defined on  $\mathbb{R}^d$  having weak derivatives of various orders, that belong to the Lebesgue spaces  $L^\lambda = L^\lambda(\mu)$  for various exponents  $\lambda$ . ( $f \in L^\lambda$  if and only if  $E_\mu |f|^\lambda := \int |f(x)|^\lambda \mu(dx) < \infty$ .) Under (E),  $\mu$  is a product measure and the model spaces admit a log-Sobolev embedding result.

Let  $C^\infty(\mathbb{R}^d; \mathbb{R})$  be the space of continuous functions with continuous partial derivatives of all orders, and let  $C_0^\infty(\mathbb{R}^d; \mathbb{R})$  be the subspace of those functions having compact support. For any  $\lambda \in [2, \infty)$ , and any  $0 \leq k \leq \lambda$ , the space  $W^{k,\lambda}$  is the mixed-norm Sobolev space comprising measurable functions  $a \in L^\lambda$  that have weak partial derivatives up to order  $k$ , those of order  $i$  belonging to the Lebesgue space  $L^{\lambda/i}$ . We shall also use the ‘‘subordinate’’ spaces  $W^{k,\lambda;l}$ , for certain integer values of  $l$ . Let  $\lambda^\circ$  be the following Lebesgue exponent: if (E) holds and  $k \geq 1$  then  $\lambda^\circ = \lambda$ , otherwise  $\lambda^\circ = \lambda - \epsilon$  for some  $0 < \epsilon \ll 1$ . For  $1 \leq l \leq \lfloor \lambda^\circ \rfloor$ , the space  $W^{k,\lambda;l}$  comprises measurable functions  $a \in L^{\lambda^\circ/l}$  that have weak partial derivatives up to order  $k_l := \min\{k, \lfloor \lambda^\circ \rfloor - l\}$ , those of order  $i$  belonging to the Lebesgue space  $L^{\lambda^\circ/(l+i)}$ . (For convenience  $W^{k,\lambda;0} := W^{k,\lambda}$ .)

Model spaces with more general derivative structures were developed in [17], including fixed-norm spaces; however, the Lebesgue exponents in  $W^{k,\lambda}$  and its subordinates are especially suited to the deformed logarithms used here. Weak derivatives are defined in the usual way: for any  $\varphi \in C_0^\infty(\mathbb{R}^d; \mathbb{R})$ ,

$$\int (\partial_i a) \varphi \, dx = - \int a (\partial_i \varphi) \, dx \quad \text{where } \partial_i a \text{ is shorthand for } \frac{\partial a}{\partial x_i}. \tag{4}$$

In order to express higher-order weak derivatives in an efficient way, we use the following standard ‘‘multi-index’’ notation. Let  $S := \{0, \dots, k\}^d$  be the set of  $d$ -tuples of integers in the range  $0 \leq s_i \leq k$ . For  $s \in S$ , we define  $|s| = \sum_i s_i$ , and denote by  $S_i := \{s \in S : 1 \leq |s| \leq i\}$  the set of  $d$ -tuples of weight at most  $i$ . For appropriate  $a$ , we define the following

$$D^s a = \partial_1^{s_1} \dots \partial_d^{s_d} a$$

$$\|a\|_{W^{k,\lambda}}^\lambda = \|a\|_{L^\lambda}^\lambda + \sum_{s \in S_k} \|D^s a\|_{L^{\lambda/(|s|)}}^\lambda$$

$$\|a\|_{W^{k,\lambda;l}}^\lambda = \|a\|_{L^{\lambda^\circ/l}}^\lambda + \sum_{s \in S_{k_l}} \|D^s a\|_{L^{\lambda^\circ/(l+|s|)}}^\lambda, \quad 1 \leq l \leq \lfloor \lambda^\circ \rfloor. \tag{5}$$

**Theorem 1** (i) For any  $1 \leq l \leq \lfloor \lambda^\circ \rfloor$ ,  $W^{k,\lambda;l}$  and  $W^{k,\lambda}$  are Banach spaces with respect to the norms in (5);

(ii) For any  $1 \leq l \leq \lfloor \lambda^\circ \rfloor$ ,  $C_0^\infty(\mathbb{R}^d; \mathbb{R})$  is dense in  $W^{k,\lambda;l}$  and  $W^{k,\lambda}$ .

**Proof** Both parts are proved in Theorem 1 and Lemmas 1 and 2 in [17]. (The only property required of the log density,  $l_r$ , is its continuity.) Part (ii) is a consequence of the non-increasing nature of the Lebesgue exponents in  $W^{k,\lambda}$  and  $W^{k,\lambda;l}$ .  $\square$

The spaces admit the following continuous embeddings:

$$W^{k,\lambda;0} := W^{k,\lambda} \prec W^{k,\lambda;l} \prec W^{k,\lambda;\tilde{l}}, \quad \text{where } 1 \leq l < \tilde{l} \leq \lfloor \lambda^\circ \rfloor. \tag{6}$$

The spaces  $W^{k,\lambda}$  will be used as model spaces for manifolds of *finite measures* in Sect. 3, and centred versions of them, as model spaces for manifolds of *probability measures* in Sect. 4. The following theorem derives some properties of particular types of map acting on them. It will be used in the sequel.

**Theorem 2** (i) For any  $\psi \in C^\infty(\mathbb{R}; \mathbb{R})$  having bounded derivatives of all orders, the nonlinear superposition operator  $\Psi : W^{k,\lambda} \rightarrow W^{k,\lambda}$ , defined by  $\Psi(a)(x) = \psi(a(x))$ , is continuous. Its spatial derivatives are given by the Faà di Bruno formula

$$D^s \Psi(a) = F_s(a) := \sum_{\pi \in \Pi(s)} \psi^{(|\pi|)}(a) \prod_{\sigma \in \pi} D^\sigma a, \tag{7}$$

where  $\pi = \{\sigma_1, \dots, \sigma_{|\pi|} \in S_{|s|}; 1 \leq |\sigma_j| \leq |s|, \sum_j \sigma_j = s\}$  is a partition of  $s$ ,  $|\pi|$  is the cardinal of  $\pi$ , and  $\Pi(s)$  is the set of all such partitions.

(ii) For appropriate Banach spaces of functions on  $\mathbb{R}^d$ ,  $A$ ,  $B$  and  $C$ , let  $\Pi_{A,B} : A \times B \rightarrow C$  be defined by  $\Pi_{A,B}(a, b)(x) = a(x)b(x)$ .  $\Pi_{A,B}$  is a well defined, continuous, bilinear map in the following instances:

$$\begin{aligned} A = B = C &= L^\infty \cap W^{k,\lambda} \quad \text{normed by } \|\cdot\|_{L^\infty} + \|\cdot\|_{W^{k,\lambda}}; \\ A &= L^\infty \cap W^{k,\lambda}, \quad B = W^{k,\lambda}, \quad C = W^{k,\lambda;1}; \\ A &= L^\infty \cap W^{k,\lambda}, \quad B = C = W^{k,\lambda;l} \quad 1 \leq l \leq \lfloor \lambda^\circ \rfloor; \\ A &= W^{k,\lambda;l}, \quad B = W^{k,\lambda;\tilde{l}}, \quad C = W^{k,\lambda;l+\tilde{l}} \quad 1 \leq l, \tilde{l}; \quad l + \tilde{l} \leq \lfloor \lambda^\circ \rfloor. \end{aligned} \tag{8}$$

The spatial derivatives of  $\Pi_{A,B}$  are given by the Leibniz formula

$$D^s \Pi_{A,B}(a, b) = H_s(a, b) := \sum_{\sigma \leq s} \frac{s!}{\sigma!(s-\sigma)!} D^\sigma a D^{s-\sigma} b, \tag{9}$$

where  $s! := s_1! \cdots s_d!$ , and  $\sigma \leq s$  if and only if  $\sigma_i \leq s_i$  for  $1 \leq i \leq d$ .

(iii) For  $\psi$  as in part (i), and any  $0 \leq l \leq \lfloor \lambda^\circ \rfloor$ , the superposition operator  $\Psi_l : W^{k,\lambda} \rightarrow W^{k,\lambda;l}$ , defined by  $\Psi_l(a)(x) = \psi(a(x))$ , is of class  $C^l$ . Its derivatives are as follows:

$$\Psi_l^{(i)}(u_1, \dots, u_i)(x) = \psi^{(i)}(a(x))u_1(x) \cdots u_i(x), \quad \text{for } 1 \leq i \leq l. \quad (10)$$

The proof makes use of the following Lemma.

**Lemma 1** Let  $a \in W^{k,\lambda}$ , let  $(a_n \neq a)$  and  $(b_n)$  be sequences converging to  $a$  in the sense of  $W^{k,\lambda}$ , and let  $B$  be the unit ball of  $W^{k,\lambda}$ . For any continuous, bounded function  $f : \mathbb{R} \rightarrow \mathbb{R}$ ,

$$\|a_n - a\|_{W^{k,\lambda}}^{-1} \|(f(b_n) - f(a))(a_n - a)\|_{L^{\lambda^\circ}} \rightarrow 0 \quad (11)$$

and  $\sup_{u \in B} \|(f(a_n) - f(a))u\|_{L^{\lambda^\circ}} \rightarrow 0. \quad (12)$

**Proof** We use the generalised Hölder inequality,

$$\|(f(b_n) - f(a))(a_n - a)\|_{L^{\lambda^\circ}} \leq \|f(b_n) - f(a)\|_A \|a_n - a\|_E^{1/\lambda}, \quad (13)$$

where  $A$  and  $E$  are the following Banach spaces. If (E) does not hold or  $k = 0$ , then  $A = L^{\lambda\lambda^\circ/\epsilon}$  and  $E = L^1$ ; (13) is then the classical Hölder inequality on dual Lebesgue spaces. If (E) holds and  $k \geq 1$  then  $A = \exp L^{1/\beta}(\mu)$  and  $E = L^1 \log^\beta L(\mu)$ , where  $\beta = (t - 1)t$ ; these are Orlicz spaces based on the complementary Young functions:

$$G_\beta(z) = \int_0^z (\exp(y^{1/\beta}) - 1) dy \quad \text{and} \quad F_\beta(z) = \int_0^z \log^\beta(y + 1) dy. \quad (14)$$

It follows from a log-Sobolev embedding theorem (see, for example, Theorem 7.12 in [6]), and the following representation for first-order weak derivatives

$$\partial_i |a_n - a|^\lambda = \lambda |a_n - a|^{\lambda-1} \text{sgn}(a_n - a) \partial_i (a_n - a) \in L^1, \quad 1 \leq i \leq d,$$

that  $\| |a_n - a|^\lambda \|_E \leq K \|a_n - a\|_{W^{k,\lambda}}^\lambda$ , for some  $K < \infty$ . (See the proof of Lemma 4 in [17] for fuller details.) Now  $f(b_n) - f(a)$  is bounded and converges to zero in probability, and so it converges to zero in the sense of  $A$  (with either definition of  $A$ ), and (11) follows. A similar argument establishes (12).  $\square$

**Proof of Theorem 2** A proof of part(i) is given in Proposition 2 of [17]. It involves a sequence  $f_n \in C^\infty(\mathbb{R}^d; \mathbb{R})$  converging to  $a$  in the sense of  $W^{k,\lambda}$ .  $D^s \Psi(f_n)$  is defined in the classical sense, and is equal to  $F_s(f_n)$ . It is then shown that  $F_s(f_n)$  converges to  $F_s(a)$  in the sense of  $L^{\lambda/|s|}$ .  $F_s(a)$  is then shown to be the weak derivative of  $\Psi(a)$  by the standard procedure of integrating  $D^s F_s(f_n)\varphi$  by parts for a  $\varphi \in C_0^\infty(\mathbb{R}^d; \mathbb{R})$ . Finally, continuity is established by repeating the argument for a sequence  $W^{k,\lambda} \ni a_n \rightarrow a$ .

The proof of part (ii) is similar. Let  $f_n, g_n \in C^\infty(\mathbb{R}^d; \mathbb{R})$  be sequences converging to  $a$  (respectively  $b$ ) in the sense of  $W^{k,\lambda;l}$  (respectively  $W^{k,\lambda;\bar{l}}$ ). Clearly  $f_n b \rightarrow ab$

in the sense of  $L^{\lambda^\circ/(l+\tilde{l})}$ . Furthermore, for any  $s \in S_{k_{l+\tilde{l}}}$ ,

$$H_s(f_n, g_n) - H_s(a, g_n) = \sum_{\sigma \leq s} \frac{s!}{\sigma!(s-\sigma)!} (D^\sigma f_n - D^\sigma a) D^{s-\sigma} g_n,$$

and so it follows from Hölder’s inequality that

$$\|H_s(f_n, g_n) - H_s(a, g_n)\|_{L^{\lambda^\circ/(|s|+l+\tilde{l})}} \rightarrow 0.$$

A similar argument can be applied to  $H_s(a, g_n) - H_s(a, b)$ , and so  $D^s(f_n g_n) \rightarrow H_s(a, b)$  in  $L^{\lambda^\circ/(|s|+l+\tilde{l})}$ . Once again, an argument involving integration by parts establishes that  $H_s(a, b) = D^s \Pi(a, b)$ , and an argument involving sequences  $W^{k,\lambda;l} \ni a_n \rightarrow a$ , and  $W^{k,\lambda;\tilde{l}} \ni b_n \rightarrow b$  establishes continuity. Similar arguments can be used with the other cases.

The case  $l = 0$  of part (iii) is established in part (i). Let  $(a_n \neq a)$  be a sequence converging to  $a$  in the sense of  $W^{k,\lambda}$ , and let

$$\Delta_n := \psi(a_n) - \psi(a) - \psi^{(1)}(a)(a_n - a). \tag{15}$$

Then, according to the mean-value theorem,  $\Delta_n = \delta_n(a_n - a)$ , where

$$\delta_n = \psi^{(1)}((1 - \beta_n)a + \beta_n a_n) - \psi^{(1)}(a) \quad \text{for some } 0 \leq \beta_n(x) \leq 1.$$

Lemma 1 shows that  $\|a_n - a\|_{W^{k,\lambda}}^{-1} \|\Delta_n\|_{L^{\lambda^\circ}} \rightarrow 0$ , and so  $\Psi_1 : W^{k,\lambda} \rightarrow L^{\lambda^\circ}$  is differentiable, with derivative as in (10) with  $i = 1$ . That this derivative is continuous follows from (12).

For  $0 \leq i \leq l - 1$ , let  $F_{s,i}(a)$  be as in (7), but with  $\psi$  replaced by  $\psi^{(i)}$ ; then

$$\begin{aligned} F_s(a_n) - F_s(a) - F_{s,1}(a)(a_n - a) &= \sum_{\pi \in \Pi(s)} \psi^{(|\pi|)}(a) \sum_j \Gamma_{\pi,j} D^{\sigma_j} (a_n - a) \\ &= \sum_{\pi \in \Pi(s)} \left( \Gamma_{\pi,0} \delta_\pi(a_n - a) + \sum_j (\Gamma_{\pi,j} \zeta_\pi + \Theta_{\pi,j}) D^{\sigma_j} (a_n - a) \right), \end{aligned}$$

where  $\{\sigma_1, \dots, \sigma_j, \dots, \sigma_{|\pi|}\}$  is an enumeration of  $\pi$ ,

$$\begin{aligned} \Gamma_{\pi,0} &= \prod_{\sigma \in \pi} D^\sigma a, & \Gamma_{\pi,j} &= \prod_{m \neq j} D^{\sigma_m} a, \\ \delta_\pi &= \psi^{(|\pi|+1)}(\beta_n a_n + (1 - \beta_n)a) - \psi^{(|\pi|+1)}(a), \\ \zeta_\pi &= \psi^{(|\pi|)}(a_n) - \psi^{(|\pi|)}(a), \\ \Theta_{\pi,j} &= \psi^{(|\pi|)}(a_n) \left( \prod_{m < j} D^{\sigma_m} a_n - \prod_{m < j} D^{\sigma_m} a \right) \prod_{m > j} D^{\sigma_m} a. \end{aligned}$$

Hölder’s inequality now shows that

$$\begin{aligned}
 R_1 &:= \|\Gamma_{\pi,0}\delta_\pi(a_n - a)\|_{L^{\gamma_0}} \leq \|\Gamma_{\pi,0}\|_{L^{\lambda^\circ/|s|}} \|\delta_\pi(a_n - a)\|_{L^{\lambda^\circ}} \\
 R_2 &:= \|\Gamma_{\pi,j}\zeta_\pi D^{\sigma_j}(a_n - a)\|_{L^{\gamma_0}} \leq \|\Gamma_{\pi,j}\zeta_\pi\|_{L^{\gamma_j}} \|D^{\sigma_j}(a_n - a)\|_{L^{\lambda^\circ/|\sigma_j|}} \\
 R_3 &:= \|\Theta_{\pi,j} D^{\sigma_j}(a_n - a)\|_{L^{\gamma_0}} \leq \|\Theta_{\pi,j}\|_{L^{\gamma_j}} \|D^{\sigma_j}(a_n - a)\|_{L^{\lambda^\circ/|\sigma_j|}}
 \end{aligned}$$

where  $\gamma_0 = \lambda^\circ/(|s| + 1)$  and  $\gamma_j = \lambda^\circ/(|s| - |\sigma_j| + 1)$ . We now claim that

$$\|a_n - a\|_{W^{k,\lambda}}^{-1} R_i \rightarrow 0 \quad \text{for } i = 1, 2, 3.$$

That this is true of  $R_1$  follows from Lemma 1. Regarding  $R_2$ ,  $\zeta_\pi$  is bounded and converges to zero in probability and so, according to the dominated convergence theorem,  $\Gamma_{\pi,j}\zeta_\pi \rightarrow 0$  in the sense of  $L^{\gamma_j}$ . Finally, the bracketed term in  $\Theta_{\pi,j}$  can be expanded as a telescopic sum of products each containing one of the differences  $D^{\sigma_m}(a_n - a)$ . Hölder’s inequality then shows that  $\Theta_{\pi,j} \rightarrow 0$  in the sense of  $L^{\gamma_j}$ .

We have thus shown that  $D^s \Psi_1 : W^{k,\lambda} \rightarrow L^{\lambda^\circ/(|s|+1)}$  is differentiable, and

$$(D^s \Psi_1)^{(1)}u = F_{s,1}(a)u + \sum_{\pi \in \Pi(s)} \psi^{(l\pi)}(a) \sum_j \Gamma_{\pi,j} D^{\sigma_j} u.$$

We can now apply the Leibniz and Faà di Bruno formulae to  $D^s(\psi^{(1)}(a)u)$  to show that it is equal to  $(D^s \Psi_1)^{(1)}u$ , and is continuous in  $(a, u)$ . This proves (10) for the case  $l = 1$ .

We now proceed by induction on  $l$ . Suppose that (10) is correct for  $l$ ; then, since  $W^{k,\lambda;l} \subset W^{k,\lambda;l+1}$ ,  $\Psi_{l+1}$  is of class  $C^l$ , with derivatives as in (10). Setting  $\Delta_{l,n} = \psi^{(l)}(a_n) - \psi^{(l)}(a) - \psi^{(l+1)}(a)(a_n - a)$ , we can apply the arguments used above on  $\Delta_n$  of (15), and the fact that

$$\sup_{u_i \in B} \|\Pi(\Delta_{l,n}, u_1 \cdots u_l)\|_{W^{k,\lambda;l+1}} \leq K \|\Delta_{l,n}\|_{W^{k,\lambda;1}}, \quad \text{for some } K < \infty,$$

where  $B$  is the unit ball of  $W^{k,\lambda}$ , to show that  $\Psi_{l+1}^{(l)}$  is of class  $C^1$ . □

### 3 The manifolds of finite measures

The charts of the statistical manifolds developed here are based on a two-parameter family of  $\eta$ -deformed logarithms. These are defined in terms of Amari’s  $\alpha$ -logarithms,  $\Lambda_\alpha : (0, \infty) \rightarrow \mathbb{R}$ :

$$\Lambda_\alpha(y) = \begin{cases} \frac{2}{1-\alpha} (y^{(1-\alpha)/2} - 1) & \text{if } \alpha \neq 1 \\ \log y & \text{if } \alpha = 1 \end{cases} \tag{16}$$

The  $\eta$ -logarithm is defined for  $\eta = (\eta_-, \eta_+)$ ,  $(-\infty < \eta_- < 1 \leq \eta_+ < \infty)$  as:

$$\log_\eta(y) = \Lambda_{\eta_-}(y) + \Lambda_{\eta_+}(y) \tag{17}$$



The deformed logarithm  $\log_{(-1,+1)}$  is that used to construct a family of highly inclusive statistical manifolds in [14, 15, 17]. Setting  $\kappa = (\eta_+ - \eta_-)/4$  and  $r = (2 - \eta_- - \eta_+)/4$ ,  $\log_\eta$  is essentially the two-parameter  $(\kappa, r)$ -logarithm defined in [8]. The different weightings for the two components of the deformed logarithm used here have no effect on the membership or properties of the manifolds, but are more convenient in the context of information geometry.

Now  $\inf_y \log_\eta y = -\infty$ ,  $\sup_y \log_\eta y = +\infty$ , and  $\log_\eta \in C^\infty((0, \infty); \mathbb{R})$  with strictly positive first derivative  $y^{-(1+\eta_-)/2} + y^{-(1+\eta_+)/2}$  and so, according to the inverse function theorem,  $\log_\eta$  is a diffeomorphism from  $(0, \infty)$  onto  $\mathbb{R}$ . Let  $\exp_\eta$  be its inverse. This can be thought of as a deformed exponential function. Using  $f^{(n)}$  to denote the  $n$ -th derivative of a function  $f$ , we have

$$\exp_\eta^{(1)} = \frac{1}{1 + \exp_\eta^{\delta/2}} \exp_\eta^{[1+\eta_+]/2} = \frac{\exp_\eta^{\delta/2}}{1 + \exp_\eta^{\delta/2}} \exp_\eta^{[1+\eta_-]/2}, \tag{18}$$

where  $\delta := \eta_+ - \eta_-$ . So  $\exp_\eta$  satisfies the differential inequality

$$\exp_\eta^{(1)} < \exp_\eta^{[1+\eta_-]/2} \tag{19}$$

and, since  $\exp_\eta(0) = 1$ , there exists a  $K_\eta < \infty$  such that

$$\exp_\eta(z) \leq K_\eta(1 + z^{2/(1-\eta_-)}) \text{ for all } z \geq 0. \tag{20}$$

If  $\eta_- = -1$  (as is the case in [14, 15, 17]) then the exponent is 1, and  $\exp_\eta$  has linear growth; otherwise it has sublinear or power law growth. The exponent itself grows without limit as  $\eta_-$  approaches 1 from below.

For any  $\alpha \in \mathbb{R}$ , the Amari embedding map  $\xi_\alpha : \mathbb{R} \rightarrow \mathbb{R}$  is as follows:

$$\xi_\alpha(z) = \Lambda_\alpha \circ \exp_\eta(z). \tag{21}$$

These maps can be used in the analysis of a large class of divergences, and their associated tensors. The maps  $\xi_{-1}$  and  $\xi_{+1}$  are especially important since they will represent the density of a measure and its log, respectively. The following lemma establishes some of their properties in the context of the  $\eta$ -log.

**Lemma 2** (i) For any  $\alpha \in \mathbb{R}$ ,  $\xi_\alpha \in C^\infty(\mathbb{R}; \mathbb{R})$ ; its derivatives are

$$\xi_\alpha^{(i)} = \frac{f_{\alpha,i}(\exp_\eta)}{(1 + \exp_\eta^{\delta/2})^{(2i-1)}} \text{ for } 1 \leq i < \infty, \tag{22}$$

where  $f_{\alpha,1}(y) = y^{(\eta_+ - \alpha)/2}$  and

$$\begin{aligned} f_{\alpha,i+1}(y) &= (y^{\delta/2} + y^\delta)y^{(\eta_- + 1)/2} f_{\alpha,i}^{(1)}(y) - (i - 1/2)\delta y^\delta y^{(\eta_- - 1)/2} f_{\alpha,i}(y) \\ &= (1 + y^{\delta/2})y^{(\eta_+ + 1)/2} f_{\alpha,i}^{(1)}(y) - (i - 1/2)\delta y^{(2\eta_+ - \eta_- - 1)/2} f_{\alpha,i}(y). \end{aligned} \tag{23}$$

(ii) For any  $1 \leq i < \infty$ , and any  $\alpha \in \mathbb{R}$ ,

$$\limsup_{z \rightarrow \infty} z^{-\beta_i} |\xi_\alpha^{(i)}(z)| < \infty, \quad \text{where } \beta_i := \frac{1 - \alpha}{1 - \eta_-} - i. \tag{24}$$

(iii) For any  $1 \leq i < \infty$ , and any  $\alpha \leq \eta_+$ ,

$$\limsup_{z \rightarrow -\infty} |\xi_\alpha^{(i)}(z)| < \infty. \tag{25}$$

(iv) If  $\alpha \in [\eta_-, \eta_+]$ , then  $\xi_\alpha^{(i)}$  is bounded for all  $1 \leq i < \infty$ .

**Proof** Part (i) is straightforward.

The power of  $y$  in  $f_{\alpha,1}(y)$  is  $(\eta_+ - \alpha)/2 = \delta/2 + (\eta_- - \alpha)/2$ , and so  $\xi_\alpha^{(1)}(z)$  grows as  $\exp_\eta(z)^{(\eta_- - \alpha)/2}$  for large  $z$ . It now follows from (20) that (24) is correct for  $i = 1$ . That it is also correct for  $i \geq 2$  follows from an induction argument based on the first representation of  $f_{\alpha,i+1}$  in (23).

If  $\alpha \leq \eta_+$  then the power of  $y$  in  $f_{\alpha,1}(y)$  is greater than or equal to 0, and so (25) is correct for  $i = 1$ . That it is also correct for  $i \geq 2$  follows from an induction argument based on the *second* representation of  $f_{\alpha,i+1}$  in (23). Part (iv) is an immediate consequence of parts (ii) and (iii).  $\square$

Let  $\theta_0 := (1 - \eta_-)\lambda/2$ . We assume that  $\theta_0 > 1$ ; it then follows from (20) and (24) that, for any  $0 \leq i < \lambda/\theta_0$  and any  $a \in L^\lambda$ ,

$$\exp_\eta^{(i)}(a) \in L^{\theta_0\lambda/(\lambda-i\theta_0)}. \tag{26}$$

(If  $i \geq \lambda/\theta_0$  then  $\exp_\eta^{(i)}$  is bounded.) We can now construct the manifold  $M (= M_\eta^{k,\lambda})$ . This is the set of finite measures on  $\mathbb{R}^d$  satisfying the following:

- (M1)  $P$  is mutually absolutely continuous with respect to Lebesgue measure;
- (M2)  $\log_\eta p \in G (= G^{k,\lambda} := W^{k,\lambda})$ .

Here,  $p$  denotes the density of  $P$  with respect to the reference probability measure  $\mu$ . Its density with respect to Lebesgue measure is  $pr$ , where  $r$  is as in (1). The chart  $\phi : M \rightarrow G$  is defined by:

$$\phi(P) = \log_\eta p. \tag{27}$$

**Proposition 1**  $\phi$  is a bijection onto  $G$ . Its inverse is

$$\phi^{-1}(a) = P(dx) = \exp_\eta(a(x))\mu(dx). \tag{28}$$

**Proof** It follows from (M2) that, for any  $P \in M$ ,  $\phi(P) \in G$ . Suppose, conversely, that  $a \in G$ ; since  $\exp_\eta(a) \in L^1$ , we can define the finite measure  $P(dx) = \exp_\eta(a(x))\mu(dx)$ . Since  $\exp_\eta$  is strictly positive,  $P$  satisfies (M1). That it also satisfies (M2) follows from the fact that  $\log_\eta \exp_\eta(a) = a \in G$ . We have thus shown that  $P \in M$ , and clearly  $\phi(P) = a$ .  $\square$

**Remark 1** This proposition shows that  $M$  is, in one sense, nothing more than a (whole) Banach space. Manifold theory enters the picture with the introduction of base-point dependent tensors such as the Fisher-Rao metric and Amari-Chentsov tensor on the tangent bundle.

The tangent space at basepoint  $P$ ,  $T_P M$ , is the linear space of signed measures,  $U$ , that are absolutely continuous with respect to Lebesgue measure and take the form

$$U(dx) = \exp_{\eta}^{(1)}(a(x))u(x)\mu(dx), \quad \text{for some } u \in G, \tag{29}$$

where  $a = \phi(P)$ .  $U$  is well defined because of (26). The representation in (29) is obtained from the tangent map of the chart;  $u$  is then the natural representation of the tangent vector,  $U$ , in the model space. The tangent bundle is the disjoint union  $TM := \cup_{P \in M}(P, T_P M)$ , and is globally trivialised by the chart  $\Phi : TM \rightarrow G \times G$ , where

$$\Phi(P, U) = (a, u), \quad \text{and } a \text{ and } u \text{ are as in (29)}. \tag{30}$$

We now investigate some of the smoothness properties of Amari’s embedding maps (21) in the context of the manifold  $M$ . These, in turn, can be used to analyse the smoothness properties of divergences and tensors.

**Proposition 2** (i) For any  $\alpha \in [\eta_-, \eta_+]$ , any  $0 \leq l \leq \lfloor \lambda^\circ \rfloor$  and any  $a \in G$ ,  $\xi_\alpha(a) \in W^{k, \lambda; l}$ . The superposition operator  $\Xi_{\alpha, l} : G \rightarrow W^{k, \lambda; l}$ , defined by  $\Xi_{\alpha, l}(a)(x) = \xi_\alpha(a(x))$ , is of class  $C^l$ , with derivatives:

$$\Xi_{\alpha, l}^{(i)}(a)(u_1, \dots, u_i)(x) = \xi_\alpha^{(i)}(a(x))u_1(x) \cdots u_i(x). \tag{31}$$

(ii) For any  $1 \leq \theta < \theta_0$  (as defined before (26)), the superposition operator  $\text{Exp}_{\eta, \theta} : G \rightarrow L^\theta$ , defined by  $\text{Exp}_{\eta, \theta}(a)(x) = \exp_{\eta}(a(x))$  is of class  $C^{\lfloor \lambda \rfloor - 1}$ , with derivatives:

$$\text{Exp}_{\eta, \theta}^{(i)}(u_1, \dots, u_i)(x) = \exp_{\eta}^{(i)}(a(x))u_1(x) \cdots u_i(x). \tag{32}$$

**Proof** Part (i) is a special case of Theorem 2(iii). Part (ii) can be proved in a similar way; the essential differences are that the derivatives of  $\exp_{\eta}$  are not necessarily bounded, and the range space of the superposition operator has a weaker topology. Let  $(a_n \in G \setminus \{a\})$  be a sequence converging to  $a$  in the sense of  $G$ . For any  $1 \leq i \leq \lfloor \lambda \rfloor - 1$  let

$$\begin{aligned} \Delta_n &:= \exp_{\eta}^{(i-1)}(a_n) - \exp_{\eta}^{(i-1)}(a) - \exp_{\eta}^{(i)}(a)(a_n - a) \\ \Gamma_n &:= \exp_{\eta}^{(i)}(a_n) - \exp_{\eta}^{(i)}(a). \end{aligned} \tag{33}$$

According to the mean-value theorem  $\Delta_n = \delta_n(a_n - a)$ , where

$$\delta_n = \exp_{\eta}^{(i)}(\beta_n a_n + (1 - \beta_n)a) - \exp_{\eta}^{(i)}(a) \quad \text{for some } 0 \leq \beta_n(x) \leq 1.$$

It follows from (26) and Hölder’s inequality that, for any  $u_1, \dots, u_i$  in the unit ball of  $G$ ,

$$\|\Delta_n u_1 \cdots u_{i-1}\|_{L^\theta} \leq \|\Delta_n\|_{L^\gamma} \quad \text{and} \quad \|\Gamma_n u_1 \cdots u_i\|_{L^\theta} \leq \|\Gamma_n u_i\|_{L^\gamma},$$

where  $\gamma := \lambda\theta/(\lambda - i\theta)$ . In order to prove part (ii), it thus suffices to show that

$$\|a_n - a\|_G^{-1} \|\Delta_n\|_{L^\gamma} \rightarrow 0 \quad \text{and} \quad \sup_{\|u\|_G=1} \|\Gamma_n u\|_{L^\gamma} \rightarrow 0. \tag{34}$$

According to (26) and the de la Vallée-Poussin theorem,  $\exp_\eta^{(i)}(a_n)^\gamma$  is uniformly integrable. Now  $\delta_n$  and  $\Gamma_n$  both converge to zero in measure, and so (34) follows from the Lebesgue-Vitaly theorem.  $\square$

**Remark 2** (i) There is a vast choice of range spaces for superposition operators of this type, each of which results in operators with different properties. Proposition 2 is not intended to be exhaustive, but to cover some of the more interesting and useful cases.

(ii) The case  $l = 0$  is worth special mention since the domain and range spaces of the superposition operators are then both  $G$ . If, for example,  $\eta_- \leq -1$  then the density  $p (= \xi_{-1}(a) + 1)$  belongs to the model space and varies continuously on the manifold, as does the log of the density.

The superposition operators  $\Xi_{\alpha,l}$  can be used in the analysis of divergences and entropies. This analysis was carried out for the  $\alpha$ -divergences,  $\alpha \in [-1, 1]$ , in [15], where it was shown that they are of class  $C^l(M \times M)$ , for values of  $l$  dependent on  $\lambda$ . Although we do not pursue these issues here, it is clear that similar methods can be used with the manifolds of this paper for any  $\alpha \in [\eta_-, \eta_+]$ . We would also expect the  $(\kappa, r)$  divergences of [8] to exhibit an equivalent degree of smoothness. Divergences can be used to define various tensor fields on  $M$ , which depend naturally on the superposition operators,  $\Xi_{\alpha,l}$ . In particular, the Fisher-Rao metric on  $M$  can be expressed in terms of the maps  $(\xi_\alpha, \alpha \in [\eta_-, \eta_+])$  in two different ways, according to the value of  $\eta_-$ :

$$\langle U, V \rangle_P = \begin{cases} E_\mu \xi_0^{(1)}(a) \xi_0^{(1)}(a) uv & \text{if } \eta_- \leq 0 \\ E_\mu \exp_\eta(a)^{\eta_-} \xi_{\eta_-}^{(1)}(a) \xi_{\eta_-}^{(1)}(a) uv & \text{if } \eta_- > 0, \end{cases} \tag{35}$$

where  $(a, u) = \Phi(P, U)$  and  $(a, v) = \Phi(P, V)$ .

**Corollary 1** *The Fisher-Rao metric is of class  $C^l$  on  $M$ , where*

$$l = \begin{cases} \lfloor \lambda^\circ \rfloor - 2 & \text{if } \eta_- \leq 0 \\ \lfloor \lambda^\circ - 2\eta_-/(1 - \eta_-) \rfloor - 2 & \text{if } \eta_- > 0. \end{cases} \tag{36}$$

**Proof** The case  $\eta_- \leq 1$  follows from a repeated application of Lemma 1, starting with  $\psi = (\xi_0^{(1)})^2$ . A similar technique can be applied if  $\eta_- > 0$ ; the essential difference is

that, at each stage, we must use Hölder’s inequality and Proposition 2(ii) with  $\theta = \eta_-$  to remove the term in  $\exp_\eta$ . □

The Fisher-Rao metric is positive definite and dominated by the chart-induced norm on  $T_P M$ . However the norms are not equivalent, and so the metric is a *weak Riemannian metric*. The Fisher-Rao metric and higher-order tensor fields, such as the Amari-Chentsov tensor, become smoother with increasing values of  $\lambda$ . As Corollary 1 shows, log-Sobolev embedding plays a role in these results for certain integer values of  $\lambda$ .

Of course, the use of Sobolev model spaces enables the analysis of quantities that depend on the weak derivatives of probability densities, such as the Hyvärinen divergence, and this is one of the motivations for extending the results of [14, 15]. The manifolds may also be useful in the theory of partial differential equations, as discussed in the final section of [17]. These aspects will be pursued elsewhere.

### 4 The manifolds of probability measures

Let  $M_0 \subset M$  be the subset of the manifold of Sect. 3, whose members are *probability* measures, and let  $L_0^\lambda$  (respectively  $G_0$ ) be the co-dimension 1 subspaces of  $L^\lambda$  (respectively  $G$ ) whose members have zero  $\mu$ -mean. Let  $\phi_0 : M_0 \rightarrow G_0$  be defined by

$$\phi_0(P) = \phi(P) - E_\mu \phi(P) = \log_\eta p - E_\mu \log_\eta p. \tag{37}$$

**Proposition 3** (i)  $\phi_0$  is a bijection onto  $G_0$ .

- (ii)  $(M_0, G_0)$  is a  $C^{[\lambda]-1}$ -embedded submanifold of  $(M, G)$ . The inclusion map  $\rho : G_0 \rightarrow G$  takes the form  $\rho(a) = a + Z(a)$ , where  $Z : G_0 \rightarrow \mathbb{R}$  is an (implicitly defined), additive normalisation constant.
- (iii)  $\rho$  and all its derivatives are bounded on bounded sets. The first (and if  $\lambda > 2$ , second) derivatives of  $\rho$  are as follows:

$$\begin{aligned} \rho_a^{(1)} u &= u - E_{P_a} u \\ \rho_a^{(2)}(u, v) &= -\frac{E_\mu \exp_\eta^{(2)}(\rho(a))(u - E_{P_a} u)(v - E_{P_a} v)}{E_\mu \exp_\eta^{(1)}(\rho(a))}, \end{aligned} \tag{38}$$

where  $P_a(dx) := \exp_\eta^{(1)}(\rho(a(x)))\mu(dx)/E_\mu \exp_\eta^{(1)}(\rho(a))$ , is the escort probability [13].

**Proof** Let  $\Upsilon : G_0 \times \mathbb{R} \rightarrow (0, \infty)$  be defined by

$$\Upsilon(a, z) = E_\mu \exp_\eta(a + z) = E_\mu \text{Exp}_\eta(a + z), \tag{39}$$

where  $\text{Exp}_\eta$  is as defined in Proposition 2 with  $\theta = 1$ . It follows from Proposition 2, that  $\Upsilon$  is of class  $C^{\lceil\lambda\rceil-1}$  and that, for any  $u \in G_0$ ,

$$\Upsilon_{a,z}^{(1,0)}u = E_\mu \exp_\eta^{(1)}(a+z)u \quad \text{and} \quad \Upsilon_{a,z}^{(0,1)} = E_\mu \exp_\eta^{(1)}(a+z) > 0. \tag{40}$$

Since  $\exp_\eta$  is monotone increasing,

$$\Upsilon(a, z) \geq E_\mu \mathbf{1}_{[-1,\infty)}(a) \exp_\eta(a+z) \geq \mu(a \geq -1) \exp_\eta(z-1),$$

and so  $\lim_{z \uparrow \infty} \Upsilon(a, z) = \infty$ . Furthermore, the monotone convergence theorem shows that

$$\lim_{z \downarrow -\infty} \Upsilon(a, z) = E_\mu \lim_{z \downarrow -\infty} \psi(a+z) = 0.$$

So  $\Upsilon(a, \cdot)$  is a bijection with strictly positive derivative, and the inverse function theorem shows that it is a  $C^{\lceil\lambda\rceil-1}$ -isomorphism. The implicit mapping theorem shows that  $Z : G_0 \rightarrow \mathbb{R}$ , defined by  $Z(a) = \Upsilon(a, \cdot)^{-1}(1)$ , is of class  $C^{\lceil\lambda\rceil-1}$ . For some  $a \in G_0$ , let  $P$  be the probability measure with density  $p = \exp_\eta(a + Z(a))$ ; then  $\phi_0(P) = a$  and  $P \in M_0$ , which proves part (i).

The argument above shows that the inclusion map,  $\rho$ , is of class  $C^{\lceil\lambda\rceil-1}$ . Let  $c : G \rightarrow G_0$  be the (linear) superposition operator defined by  $c(a)(x) = a(x) - E_\mu a$ ; then  $c$  is continuous, and has derivative  $c_a^{(1)}u = u - E_\mu u$ . Now  $c \circ \rho$  is the identity map of  $G_0$ , which shows that  $\rho$  is homeomorphic onto its image,  $\rho(G_0)$ , endowed with the relative topology. Furthermore, for any  $u \in G_0$ ,

$$u = (c \circ \rho)_a^{(1)}u = c_{\rho(a)}^{(1)}\rho_a^{(1)}u,$$

and so  $\rho_a^{(1)}$  is a toplinear isomorphism, and its image,  $\rho_a^{(1)}G_0$ , is a closed linear subspace of  $G$ . Let  $E_a$  be the one dimensional subspace of  $G$  defined by  $E_a = \{y \exp_\eta^{(1)}(\rho(a)) : y \in \mathbb{R}\}$ . If  $u \in E_a$  and  $v \in \rho_a^{(1)}G_0$  then there exist  $y \in \mathbb{R}$  and  $w \in G_0$  such that

$$E_\mu uv = y E_\mu \exp_\eta^{(1)}(\rho(a))(w - E_{P_a}w) = 0.$$

So  $E_a \cap \rho_a^{(1)}G_0 = \{0\}$ , and  $\rho_a^{(1)}$  splits  $G$  into the direct sum  $E_a \oplus \rho_a^{(1)}G_0$ . We have thus shown that  $\rho$  is a  $C^{\lceil\lambda\rceil-1}$ -immersion, and this completes the proof of part (ii).

Jensen’s inequality shows that there exists a  $K_\eta < \infty$  such that

$$-\log E_\mu \exp_\eta^{(1)}(\rho(a)) \leq -E_\mu \log \exp_\eta^{(1)}(\rho(a)) \leq K_\eta E_\mu |\log p| \leq K_\eta \|a\|.$$

So, for bounded  $B$ ,  $\inf_{P \in B} E_\mu \exp_\eta^{(1)}(\rho(a)) > 0$ . □

For any  $P \in M_0$ , the tangent space  $T_P M_0$  is a subspace of  $T_P M$  of co-dimension 1; in fact, as shown in the proof of Proposition 3(ii),

$$T_P M = T_P M_0 \oplus \{y\hat{U}, y \in \mathbb{R}\}, \quad \text{where } \hat{U}\phi = \psi^{(1)}(\phi(P)). \quad (41)$$

Let  $\Phi_0 : TM_0 \rightarrow G_0 \times G_0$  be defined as follows:

$$\Phi_0(P, U) = \Phi(P, U) - E_\mu \Phi(P, U). \quad (42)$$

Then  $\Phi \circ \Phi_0^{-1}(a, u) = (\rho(a), \rho_a^{(1)}u)$ . For any  $(P, U) \in TM_0$ ,  $U\phi = \rho_a^{(1)}u = u - E_{P_a}u$ , and so tangent vectors in  $T_P M_0$  are distinguished from those merely in  $T_P M$  by the fact that their total mass is zero.

Any regularity possessed by divergences, entropies and tensors on  $M$  involving fewer than  $\lceil \lambda \rceil$  derivatives is also enjoyed by their restrictions to  $M_0$ .

## 5 Concluding remarks

This paper has developed a family of non-parametric statistical manifolds that use the two-parameter deformed logarithm of (17), and a variety of model spaces of Sobolev type. It has shown that the mixed-norm space  $W^{k,\lambda}$  is especially suited to this application. The Amari embedding maps,  $\xi_\alpha$ , which are central to the analysis of divergences, entropies and associated tensors, “lift” to continuous nonlinear superposition operators acting on the Sobolev model spaces. (A rare property in the theory of such operators.) Variants of the superposition operators having Sobolev range spaces with weaker topologies enjoy greater regularity; they were shown to admit multiple derivatives on the manifolds, according to the values of the parameters  $k$ ,  $\lambda$  and  $\eta$ . Of course, this paper takes only the first step in a fuller analysis of the information geometry of the manifolds constructed. However, for reasons of space, we shall go no further here.

**Data Availability** Data sharing is not applicable to this article as no datasets were generated or analysed during the current study.

## Declarations

**Conflict of interest** The author states that there is no conflict of interest.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

1. Amari, S.-I., Nagaoka, H.: *Methods of Information Geometry*, Translations of Mathematical Monographs, 191. American Mathematical Society, Providence (2000)
2. Ay, N., Jost, J., Van Lê, H., Schwachhöfer, L.: *Information Geometry*, *Ergebnisse der Mathematik und ihrer Grenzgebiete. 3. Folge / A Series of Modern Surveys in Mathematics*, vol 64. Springer, Cham (2017). <https://doi.org/10.1007/978-3-319-56478-4>
3. Barndorff-Nielsen, O.E.: *Information and Exponential Families in Statistical Theory*, Wiley (1978)
4. Cena, A., Pistone, G.: Exponential statistical manifold. *Ann. Inst. Stat. Math.* **59**, 27–56 (2007)
5. Chentsov, N.N.: *Statistical Decision Rules and Optimal Inference*. Translations of Mathematical Monographs, vol. 53. American Mathematical Society, Providence (1982)
6. Cianchi, A., Pick, L., Slavíková, L.: Higher-order Sobolev embeddings and isoperimetric inequalities. *Adv. Math.* **273**, 568–650 (2015)
7. Gibilisco, P., Pistone, G.: Connections on non-parametric statistical manifolds by Orlicz space geometry. *Infin.-Dimens. Anal. Quantum Probab. Relat. Top* **1**, 325–347 (1998)
8. Kaniadakis, G., Lissia, M., Scarfone, A.M.: Two-parameter deformations of logarithm, exponential, and entropy: a consistent framework for generalized statistical mechanics. *Phys. Rev. E* **71**, 046128 (2005)
9. Lauritzen, S.L.: *Statistical Manifolds*, IMS Lecture Notes Series, 10, Institute of Mathematical Statistics (1987)
10. Lods, B., Pistone, G.: Information geometry formalism for the spatially homogeneous Boltzmann equation. *Entropy* **17**, 4323–4363 (2015)
11. Loaiza, G., Quiceno, H.R.: A  $q$ -exponential statistical Banach manifold. *J. Math. Anal. Appl.* **398**, 466–476 (2013)
12. Murray, M.K., Rice, J.W.: *Differential Geometry and Statistics*, Monographs in Statistics and Applied Probability, 48, Chapman Hall (1993)
13. Naudts, J.: *Generalised Thermostatistics*, Springer (2011)
14. Newton, N.J.: An infinite-dimensional statistical manifold modelled on Hilbert space. *J. Funct. Anal.* **263**, 1661–1681 (2012)
15. Newton, N.J.: Infinite-dimensional statistical manifolds based on a balanced chart. *Bernoulli* **22**, 711–731 (2016). <https://doi.org/10.3150/14-BEJ673>
16. Newton, N.J.: Nonlinear filtering and information geometry: a Hilbert manifold approach, in: Ay, N., Gibilisco, P., Matúš, F. (eds.) *Information Geometry and its Applications*, *Proceedings in Mathematics and Statistics*, 252, Springer, Cham, 189–208 (2018). [https://doi.org/10.1007/978-3-319-97798-0\\_7](https://doi.org/10.1007/978-3-319-97798-0_7)
17. Newton, N.J.: A class of non-parametric statistical manifolds modelled on Sobolev space. *Inf Geometry Springer* **2**, 283–312 (2019). <https://doi.org/10.1007/s41884-019-00024-z>
18. Pistone, G., Rogantin, M.P.: The exponential statistical manifold: mean parameters, orthogonality and space transformations. *Bernoulli* **5**, 721–760 (1999)
19. Pistone, G., Sempì, C.: An infinite-dimensional geometric structure on the space of all the probability measures equivalent to a given one. *Ann. Stat.* **23**, 1543–1561 (1995)
20. Tsallis, C.: The nonadditive entropy  $S_q$  and its applications in Physics and elsewhere, *Entropy*, 13, 1765–1804 (2011). <https://doi.org/10.3390/e13101765>
21. Vigelis, R.F., Cavalcante, C.C.: On  $\varphi$ -families of probability distributions. *J. Theor. Probab.* **26**, 870–884 (2013)
22. Vieira, F.L.J., de Andrade, L.H.F., Vigelis, F.R., Cavalcante, C.C.: A Deformed Exponential Statistical Manifold, *Entropy*, 21 (2019). <https://doi.org/10.3390/e21050496>

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.