# Statistical manifold as an affine space: A functional equation approach

Jun Zhang [a,*] and Peter Hästö [b,1]

[a] *Department of Psychology, University of Michigan, Ann Arbor, MI 48109–1109, USA*

[b] *Department of Mathematical Sciences, Norwegian University of Science and Technology, 7491 Trondheim, NORWAY*

**Abstract**

A statistical manifold $M_\mu$ consists of positive functions $f$ such that $f\,d\mu$ defines a probability measure. In order to define an atlas on the manifold, it is viewed as an affine space associated with a subspace of the Orlicz space $L^\Phi$. This leads to a functional equation whose solution, after imposing the linearity constrain in line with the vector space assumption, leads to a general form of mappings between the affine probability manifold and the vector (Orlicz) space. These results generalize the exponential statistical manifold and clarify some foundational issues in non-parametric information geometry.

## 1 Introduction

Information geometry investigates the differential geometric structure of a manifold of probability density functions. It has found many applications in theoretical statistics, information theory, stochastic process, neural computation, machine learning, Bayesian statistics, and other related fields (Amari, 1985; Amari & Nagaoka, 2000). Recently, an interest in the geometry associated with non-parametric probability densities has arisen (Pistone & Sempi, 1995; Giblisco & Pistone, 1998; Pistone & Rogantin, 1999; Pistone, 2001; Grasselli, 2001). Non-parametric statistical models are important in a wider range of areas including psychological measurement and perception, e.g., Townsend et al. (2001). Unlike in the parametric case where the manifold of probability density functions inherits a Euclidean topology from the space of its natural parameters, a major challenge for the non-parametric case is to define a suitable topology and develop a corresponding notion of convergence. Fortunately, this obstacle was overcome by the introduction of an exponential statistical manifold by Pistone & Sempi (1995). These authors, among other things, gave an explicit formula for a chart of the manifold formed by all density functions absolutely continuous with respect to a given one.

In this note we address some foundational issues of non-parametric information geometry, with the goal of extending the Pistone-Sempi formula that maps the manifold of probability density functions to a subset in an Orlicz space (a special Banach space, see Appendix A). The Banach-space valued map is the infinite-dimension extension of the coordinate functions which take values in $\mathbb{R}^n$. Recall that $\mathbb{R}^n$ has both a topological structure, in terms of its canonical topology, an algebraic structure as a vector space, and a geometric or

* Corresponding author.
*Email addresses:* `junz@umich.edu` (Jun Zhang), `peter.hasto@helsinki.fi` (Peter Hästö).
*URLs:* `http://www-personal.umich.edu/~junz/` (Jun Zhang), `http://www.helsinki.fi/~hasto/` (Peter Hästö).

1 Supported in part by a Gehring-Finland Post-doctoral Fellowship, by the Finnish Academy of Science and Letters and by the Research Council of Norway, Project 160192/V30

affine structure as a set of points. Our approach is to exploit this affine structure and view the statistical manifold as an affine space associated with the vector space $L^\Phi$.

An affine space is a homogeneous set of points such that no point stands out in particular. Affine spaces differ from vector spaces in that no origin has been selected. So affine space is fundamentally a geometric structure – an example being the plane. The structure of an affine space is given by an operation $\oplus\colon A\times U\to A$ which associates to a point $a$ in the affine space $A$ and a vector $u$ in the vector space $U$ another point $a\oplus u$ in $A$. We think of this as a translation of a point $a$ in its space $A$ by a vector $u$. Notice that it makes no sense to add two points of $A$ in this setting.

One advantage of modelling a statistical manifold as a generalized affine space is to address the issue of representation of probability measures, i.e., from the probability simplex ($A$) to an extended vector space $U$. If the operation $\oplus$ exists and is continuous and differentiable, a global mapping is established between $U$ and $A$. This allows the setting up of a global atlas at any given point of $A$ (recall that the usual differential manifold only assumes the existence of an atlas locally).

The structure of the rest of this note is as follows: In the next section we review the Pistone–Sempi framework, and consider an obvious generalization of their original formulation. In Section 3 we introduce the affine structure on the probability manifold and derive a corresponding functional equation. We then solve this equation and derive a general expression for the $\oplus$ operation. This generalizes the exponential model in a natural way. The reader who needs some refreshing of their Orlicz space theory can start by consulting Appendix A.

## 2 The Pistone–Sempi Framework Revisited

Pistone & Sempi (1995) introduced a non-parametric exponential statistical manifold consisting of all densities that are absolutely continuous with respect to a given one. In this section we review the parts of that article most relevant to our considerations.

Let $X$ be a set and $\mu$ a $\sigma$-finite probability measure on $X$, in other words $\mu(X)=1$. We will consider the set

$$\mathcal{M}_\mu := \{p\in L^1(X,\mu)\colon p>0\ a.e., \|p\|_{L^1(X,\mu)}=1\}.$$

This set will be endowed with a structure of a differentiable manifold. Since $X$ and $\mu$ will be fixed, we abbreviate $L^1=L^1(X,\mu)$ and $L^1(p)=L^1(X,p\,\mu)$, and similarly for other function spaces. The expectation operator $\mathrm{E}_p(f)$ over any function $f$ on $X$ is defined as

$$\mathrm{E}_p(f)=\int_X fp\,d\mu=\|f\|_{L^1(p)}.$$

Recall that a differential manifold $\mathcal{M}$ of a set of functions is defined as follows: there exists a system of charts $\{\Omega_i,\phi_i\}_{i\in\mathbb{N}}$, collectively called an atlas, such that

(i) each $\Omega_i$ is an open set on $\mathcal{M}$ and the union $\cup_{i\in\mathbb{N}}\Omega_i$ covers $\mathcal{M}$;

(ii) the associated mappings $\phi_i\colon\Omega_i\to B$ are all homeomorphisms (here $B$ is some Banach space) and have the property that whenever $\Omega_i$ and $\Omega_j$ have non-empty intersection then the mapping $\phi_j^{-1}\circ\phi_i$ is differentiable in this common domain of definition $\Omega_i\cap\Omega_j$.

The functions $\phi_i$ themselves are called coordinate functions. In the finite-dimensional case, the coordinate functions are valued in (subsets of) $\mathbb{R}^n$. However, in the construction of Pistone and Sempi, the coordinate functions $\phi_i$ are valued on subsets of an "exponential Orlicz space" (which is a Banach space) defined by the Young function $\Phi(t)=\exp(|t|)-1$; we denote this Orlicz space by $\exp L(p)=\exp L(X,p\,\mu)$.

A major achievement of Pistone & Sempi (1995) is to provide an atlas $\phi_p:\mathcal{M}_\mu\to\exp L_0(p)$ for $\mathcal{M}_\mu$ centered at any of its point $p$:

$$\phi_p(q)=\log\tfrac{q}{p}-\mathrm{E}_p\big(\log\tfrac{q}{p}\big). \tag{1}$$

Here $\exp L_0(p)$ is a sub-space of $\exp L(p)$ (normed according to the Young's function $\exp(|\cdot|)-1$) consisting of the functions with zero mean:

$$\exp L_0(p)=\{u\in\exp L(p):\mathrm{E}_p(u)=0\}.$$

In other words, the space $\exp L_0(p)$ contains all random variables with zero expectation value. Pistone & Sempi (1995) showed that in an open neighborhood surrounding $p$, the $\mathrm{E}_p\{\cdot\}$ term in (1) is always finite and hence $\phi_p$ is a well-defined chart for any reference point $p\in\mathcal{M}_\mu$, with $\phi_p(p)=0$. (Note that the chart is not global because, in the infinite-dimension setting, the second term in (1) may be unbounded for certain $q\in\mathcal{M}_\mu$.) The inverse of the coordinate mapping $\phi_p^{-1}:\exp L_0(p)\to\mathcal{M}_\mu$ gives the exponential family of density functions

$$\phi_p^{-1}(u)=\frac{pe^u}{\mathrm{E}_p(e^u)}. \tag{2}$$

Again this formula is well-defined only when $\mathrm{E}_p(e^u)$ is finite. For this reason, Pistone and Sempi require that $u$

lies in the closed unit ball $B^{\exp}(p)$ of $\exp L(p)$. Under this condition, the coordinate transformation $\phi_{p_2} \circ \phi_{p_1}^{-1}$ between two atlases centered at different points $p_1, p_2$ of $\mathcal{M}_\mu$ is simply

$$u \mapsto u + \log \tfrac{p_1}{p_2} - \mathrm{E}_{p_2}\big(u + \log \tfrac{p_1}{p_2}\big).$$

The construction of the mapping (1) from $\mathcal{M}_\mu$ to the space of zero-mean random variables can be understood in the following way. Suppose we start from the entire Orlicz space $\exp L(p)$, as opposed to $\exp L_0(p)$, and restrict the coordinate function $u \in \exp L(p)$ to lie in the closed unit ball $B^{\exp}$, i.e. $\|u\|_{\exp L(p)} \leq 1$. By the definition of the norm and the monotonicity of the Young function, this implies that

$$\int_X \exp|u|\, p\, d\mu \leq \int_X \exp\left(|u|/\|u\|_{\exp L(p)}\right) p\, d\mu = 1,$$

i.e. $\exp|u|$ is in $L^1(p)$. Thus $\exp u$ is also in $L^1(p)$ since

$$\int_X \exp u\, p\, d\mu \leq \int_X \exp|u|\, p\, d\mu.$$

So there is an obvious way to get a function of unit integral (and hence an element of $\mathcal{M}_\mu$) using $\exp u$, namely

$$u \mapsto \frac{p \exp u}{\mathrm{E}_p(\exp u)} \tag{3}$$

which is just (2). It is clear that in this case $\phi_p^{-1}\colon B^{\exp} \to M_\mu$ as defined above is many-to-one. Specifically, $\phi_p^{-1}(u) = \phi_p^{-1}(v)$ if and only if $\exp u = c \exp v$, or expressed differently, $u = v + \log c$, for some positive constant $c$. Because of this extra degree of freedom, we can define a foliation of $\exp L(p) = \cup_c \exp L_c(p)$ where $\exp L_c(p) = \{u \in \exp L(p) : \mathrm{E}_p(u) = \log c\}$. We can then require that $\phi_p^{-1}(q)$ be defined on a particular leaf of this foliation of the Banach space. When $c = 1$ this leads to the Pistone–Sempi's formula of $\phi_p$, i.e., (1).

We now investigate whether this construction can be extended to an arbitrary Orlicz space $L^\Phi$, using an arbitrary class of Young function $\Phi$, instead of from $\exp L$, using the particular Young function $\exp(|\cdot|) - 1$. Introduce a strictly increasing function $\Phi : \mathbb{R} \to (0, \infty)$ that is convex on $[0, \infty)$ and satisfies $\Phi(-t) = (\Phi(t))^{-1}$ (and $\Phi(0) = 1$). Now $\Psi(t) = \Phi(|t|) - 1$ is a Young function. With an abuse of notation, we still denote the corresponding Orlicz space as $L^\Phi$ rather than $L^\Psi$. The arguments of the above paragraph, which was made to $\Phi(t) = e^t$, can be now made to any such $\Phi$ in general. Everything works fine up until (3). When we require, in general, that $\phi_p^{-1}(u) = \phi_p^{-1}(v)$ if and only if $\Phi(u) = c\Phi(v)$, we run into a problem. Let us define $u_c = \Phi^{-1}(c\Phi(u))$, a function on $X$. The problem is that we do not know if for a fixed $u \in B^\Phi$ (the unit ball in $L^\Phi$) there exists a positive constant $c$ such that

$$\mathrm{E}_p(u_c) = 0.$$

Hence we must look for another foliation procedure. Since our goal is to turn the multiplicative constant into an additive one, it is natural to look at the logarithm. We easily see that

$$\mathrm{E}_p(\log(\Phi(u_c)) = \log(c) + \mathrm{E}_p(\log \Phi(u)). \tag{4}$$

To make sure that this formula makes sense, we need to prove that the integral involved in the right hand side is absolutely convergent. It follows from our assumption $\log \Phi(-t) = -\log \Phi(t)$ that

$$\mathrm{E}_p(|\log \Phi(u)|) = \mathrm{E}_p(\log \Phi(|u|)).$$

An application of Jensen's inequality yields

$$\mathrm{E}_p(\log \Phi(|u|)) \leq \log \mathrm{E}_p(\Phi(|u|)) < \infty$$

where we used also $u \in L^\Phi$. That $\mathrm{E}_p(|\log \Phi(u))$ (and hence $\mathrm{E}_p(\log \Phi(u)))$ is finite allows us to impose, with reference to (4), the foliation

$$\mathrm{E}_p(\log(\Phi(u_c))) = 0.$$

Therefore the Pistone-Sempi approach will work if we use any positive function $\Phi(t)$ that is strictly increasing and convex on $[0, \infty)$, and satisfies $\Phi(t)\Phi(-t) = 1$. The corresponding map is

$$u \mapsto \frac{p\Phi(u)}{\mathrm{E}_p(\Phi(u))}.$$

We call this the pseudo-exponential map. But this generalization of Pistone-Sempi is somewhat trivial because it still uses the same kind of normalization for mapping $B^\Phi$ to $\mathcal{M}_\mu$ and hence does not generate new insights into the representation of probability density functions using Orlicz space functions. We turn to another approach in the next section.

## 3 The Affine Space Model of Statistical Manifolds

An affine space is a set of points in which each point can be "translated" to any other point through an associated vector space. More precisely, the set $A$ is an affine space associated to the vector space $U$ if there exists an operation $\oplus\colon A \times U \to A$, called "right translation" or "translation" for short, through which $U$ acts on $A$ transitively. Writing out the axioms, this means that

(i) $(p \oplus u) \oplus u' = p \oplus (u + u')$ for all $p \in A$ and $u, u' \in U$.

(ii) $(p \oplus 0) = p$ for all $p \in A$.

(iii) The restricted mapping $f_u(p) = p \oplus u$ is surjective for every fixed $u \in U$.

It can be easily verified that the exponential model satisfies the above axioms with

$$p \oplus u = \frac{pe^u}{\mathrm{E}_p(e^u)}.$$

The Pistone–Sempi formula (2) can be viewed as defining the right translation for the exponential family. Of course, the affine structure of the exponential map, with log-likelihoods as score functions, is well known, e.g., (Amari, 1985; Murray & Rice, 1993). So one way to generalize the exponential model is to see what other representations $\oplus$ could have.

Let us denote $p \oplus u$ as $F(p, u)$. Then the first axiom above can be written as $F(F(p, u), u') = F(p, u + u')$. This is a functional equation known as the "translation equation" (Aczél, 1966, 8.2.2). Below we derive a general form for the solution of this equation in the infinite dimensional case, following the finite-dimensional solution in (Aczél, 1966, 8.2.2, Theorem 1). In the next theorem we assume that the Banach space splits as a direct sum of the type $B = V \oplus \mathbb{R}u_0$. We use the notation $u = [v, c]$ to denote the splitting of individual elements of the Banach space $B$ accordingly.

**Theorem 1** *Let $P$ and $B$ be Banach spaces and suppose that $F\colon P \times B \to P$ is a function satisfying*

$$F(F(p, u), u') = F(p, u + u') \tag{5}$$

*for all $p \in P$ and $u, u' \in B$. Assume that $B$ splits as $V \oplus \{cu_0 : c \in \mathbb{R}\}$ for a fixed $u_0 \in B$, and there exists $\pi \in P$ so that $G_c(\cdot) = F(\pi, [\cdot, c])\colon V \to P$ is a bijection for every fixed $c \in \mathbb{R}$. Then all continuous solutions of* (5) *are of the form*

$$F(p, u) = G_0(G_0^{-1}(p) + L(u)),$$

*where $L\colon B \to V$ is linear.*

**Proof.** Fix $\pi \in P$ so that $G_c(\cdot) = F(\pi, [\cdot, c])\colon V \to P$ is a bijection for every fixed $c \in \mathbb{R}$. For $v \in V$, denote $h(v) = F(\pi, [v, 0]) = G_0(v)$.

We next set $p = \pi$, $u = [v, c]$ and $u' = [v', -c]$ in (5):

$$\begin{aligned} F(F(\pi, [v, c]), [v', -c]) &= F(\pi, [v + v', 0]) \\ &= h(v + v'). \end{aligned}$$

We denote $p = F(\pi, [v, c])$, and note that then $G_c^{-1}(p) = G_c^{-1}(F(\pi, [v, c])) = v$. We have thus derived that

$$F(p, [v', -c]) = h(G_c^{-1}(p) + v') \tag{6}$$

for every $p \in P$ (here we use that $G_c$ is a surjection) and $[v', -c] \in B$.

We substitute this expression for $F$ in the original functional equation (with $u = [v, c]$ and $u' = [v', c']$) and get

$$\begin{aligned} &h(G_{-c'}^{-1}(h(G_{-c}^{-1}(\pi) + v)) + v') \\ &\quad = h(G_{-(c+c')}^{-1}(\pi) + v + v'). \end{aligned}$$

Cancelling the outermost $h$ and setting $c = 0$ gives

$$G_{-c'}^{-1}(h(G_0^{-1}(\pi) + v)) = G_{-c'}^{-1}(\pi) + v.$$

We define $q = h(G_0^{-1}(\pi) + v)$ and a function $l\colon \mathbb{R} \to V$ by $l(c') = G_{-c'}^{-1}(\pi) - G_0^{-1}(\pi)$. Then the previous equation becomes

$$G_{-c'}^{-1}(q) = l(c') + h^{-1}(q).$$

Using this in (6) gives

$$F(q, [v, c]) = h(l(c) + h^{-1}(q) + v).$$

Since $h$ is just a special form of $F$ (which is assumed to be continuous), it is continuous, and so the above equation implies that $l$ is continuous. Substituting this back in (5) gives

$$\begin{aligned} &h\big\{l(c') + h^{-1}\big[h(l(c) + h^{-1}(p) + v)\big] + v'\big\} \\ &\quad = h\big(l(c + c') + h^{-1}(p) + v + v'\big). \end{aligned}$$

After cancelling the outermost $h$, $h^{-1} \circ h$ and common terms, all that remains is

$$l(c') + l(c) = l(c + c').$$

Since $l$ is continuous this Cauchy equation has only linear solutions

$$l(c) = \kappa c$$

where $\kappa$ is an element in $V$. Thus we have shown that the solution is of the form

$$F(p, u) = h\big(h^{-1}(p) + L(u)\big),$$

where $L([v, c]) = v + \kappa c$ is a linear function mapping of $B$ to $V$. It is easy to see that every $F$ of this form is a solution. $\square$

The previous theorem is not directly applicable to our setting. The problem is that, in $F(p, u)$, the range in the second variable $u$ depends on the value of the first variable $p \in \mathcal{M}_\mu$, because $u$ is in the unit ball of $\exp L(p)$. The way to get around this is to notice that $L^\infty \subset L^\Phi(p)$ for every $p \in \mathcal{M}_\mu$, and, moreover, it is dense. So we consider only functions $F\colon \mathcal{M}_\mu \times L^\infty \to M_\mu$. Since we know all continuous solutions in a dense subset of our space, we obviously know all continuous solutions in the whole space, as well.

With the understanding that we restrict ourselves to dense subsets when necessary, we can see immediately how Theorem 1 relates to the charting of a statistical manifold. The inverse of $F(p, u)$, while keeping $p$ fixed, gives an atlas $\phi_p : \mathcal{M}_\mu \to V \subset B$:

$$\phi_p(q) = L^{-1}(h^{-1}(q) - h^{-1}(p)).$$

Let us next see how the exponential model fits into this general scheme. Recall that Theorem 1 involves a co-dimension 1 splitting of the Banach space $B$ in the form $B = V \oplus \{cu_0 : c \in \mathbb{R}\}$. Set $u_0 = 1$, the constant function, so that the splitting becomes $\exp L = \exp L_0 \oplus \mathbb{R}$, where $\exp L_0$ is (as defined in Section 1) the space of exponentially integrable functions of zero mean, and $\mathbb{R}$ denotes the space of constant functions. Then choose $\kappa = 0$, which amounts to requiring $f(\pi, u) \equiv \pi, \forall u \in \mathbb{R}$, i.e., for all constant functions $u$. Consequently, $L(u) \equiv 0$ if and only if $u$ is a constant function. Choose $G_0^{-1}(p) = L(\log p)$. Then

$$\begin{aligned} p \oplus u &= \exp\left(L^{-1}(L(\log p) + L(u))\right) \\ &= \exp\left(L^{-1}(L(\log(p) + u))\right). \end{aligned}$$

We know that $L^{-1}(L(u)) = u + c$ for some constant $c$, since constants are the only elements in the kernel of $L$. Let us define a functional (interpreting the constant function as a real number) by $c[u] = L^{-1}(L(u)) - u$. Then our previous equation gives

$$\begin{aligned} p \oplus u &= \exp(u + \log(p) + c[u + \log p]) \\ &= pe^u e^{c[\log(p \exp u)]}. \end{aligned}$$

Thus we get the Pistone–Sempi model if we choose $L^{-1}$ so that

$$c[u] = -\log \int_X \exp u \, d\mu. \tag{7}$$

Let us look at some obvious generalizations: we fix a non-zero $u_0 \in L^\Phi$ and define $L$ to be a linear transformation so that $L(u) = 0$ if and only if $u = cu_0$. We define $G_0^{-1}(p) = L(\Phi^{-1}(p))$. We require that $G_0^{-1}$ be an injection, so $\Phi^{-1}(p) = \Phi^{-1}(q) + cu_0$ if and only if $p = q$ and $c = 0$. We find that

$$p \oplus u = \Phi\big(u + \Phi^{-1}(p) + c[u + \Phi^{-1}(p)]\big),$$

where the functional $c[u] = L^{-1}(L(u)) - u$ as before. Now we should choose

$$c[v] = \inf\Big\{t > 0\colon \int_X \Phi(v + t)\, d\mu = 1\Big\},$$

in order for $p \oplus u$ to be in the manifold. For $\Phi(t) = e^t$, the above reduces to (7).

Lastly, we note that our affine space approach in extending the exponential model is not an all inclusive one. Burdet et al. (2001) gave the following chart for mapping the open unit ball of $L^2(p)$ to $\mathcal{M}_\mu$:

$$p \oplus u = p\Big(u + \sqrt{1 - \mathrm{E}_p(u^2)}\Big)^2$$

with the inverse given by (for fixed $p$)

$$\phi_p(q) = \sqrt{\frac{q}{p}} - \mathrm{E}_p\left(\sqrt{\frac{q}{p}}\right).$$

Clearly

$$(p \oplus u) \oplus u' \neq p \oplus (u + u').$$

The reason for violating the right association is actually quite simple: the open unit ball of $L^2(p)$ is no longer a vector space. As a generalization of the translation equation (5), one should consider the so-called "transformation equation" (Aczél, 1966, 8.2.1)

$$(p \oplus u) \oplus u' = p \oplus (u \circ u')$$

where $\circ$ is some operator satisfying associativity

$$(u \circ u') \circ u'' = u \circ (u' \circ u'').$$

The corresponding functional equation is

$$F(F(p, u), v) = F(p, G(u, v)).$$

This is out of the scope of the current note and is the subject for future research.

## A Orlicz Spaces

In this appendix we give a short recap of the theory of Orlicz spaces. For more details the reader is referred to one of the numerous books on the subject, e.g., (Krasnosel'skiĭ & Rutickiĭ, 1961; Rao & Ren, 1991).

Recall that a Young function is a convex increasing function $\Phi\colon [0, \infty) \to [0, \infty)$ with $\Phi(0) = 0$. For a Young function $\Phi$ we define a *modular* on the set of measurable functions by

$$\varrho_\Phi(u) := \int_X \Phi(|u(x)|)\, d\mu(x).$$

Using the modular we can define the *Luxemburg norm* on the same set by

$$\|u\|_\Phi := \inf\{t > 0 \colon \varrho_\Phi(u/t) \le 1\}.$$

Using these concept we define the Orlicz space by

$$L^\Phi(X) := \{u \colon \|u\|_\Phi < \infty\}.$$

The best-known example of an Orlicz space is given by the Young functions $\Phi(t) = t^p$ for some real number $p$ greater than or equal to $1$. In this case we have

$$\varrho_\Phi(u) := \int_X |u(x)|^p \, d\mu(x).$$

and the norm is just the Lebesgue norm:

$$\begin{aligned} \|u\|_\Phi &:= \inf\{t > 0 \colon \varrho_\Phi(u/t) \le 1\} \\ &= \left( \int_X |u(x)|^p \, d\mu(x) \right)^{1/p}. \end{aligned}$$

Another important case is $\Phi(t) = \exp(|t|) - 1$, and the corresponding Orlicz space $\exp L$ is the space of exponentially integrable functions.

## References


Aczél, J. (1966). *Lectures on Functional Equations and Their Applications*. Academic Press, New York–London.

Amari, S.-I. (1985). *Differential-Geometrical Methods in Statistics*. Lecture Notes in Statistics, 28. Springer-Verlag, New York.

Amari, S. and Nagaoka, H. (1993/2000). *Methods of Information Geometry*. Translations of Mathematical Monographs, v.191. Oxford University Press.

Burdet, G., Combe, Ph. & Nencka, H. (2001). On real Hilbertian info-manifolds, *Disordered and Complex Systems* (London, 2000), 153–158, AIP Conf. Proc., 553, Amer. Inst. Phys., Melville, NY.

Giblisco, P. and Pistone, G. (1998). Connections on non-parametric statistical manifolds by Orlicz space geometry. *Infin. Dimens. Anal. Quantum Probab. Relat. Top.* **1**, no. 2, 325–347.

Grasselli, M. (2001). Classical and quantum information geometry. Ph.D thesis, Department of Mathematics, King's College London. Available at `www.math.mcmaster.ca/grasselli/thesis.html`.

Krasnosel'skiĭ, M. A. and Rutickiĭ, Ya. B. (1961). *Convex Functions and Orlicz Spaces*. Translated by Boron, L.F. Published by P.Noordhoff, Ltd, Groningen, The Netherlands.

Murray, M.K. and Rice, J.W. (1993). *Differential Geometry and Statistics*. Chapman & Hall, London.

Pistone, G. (2001). New ideas in non-parametric estimation, *Disordered and Complex Systems* (London, 2000), 159–164, AIP Conf. Proc., 553, Amer. Inst. Phys., Melville, NY.

Pistone, G. and Rogantin, M. P. (1999). The exponential statistical manifold: mean parameters, orthogonality and space transformations. *Bernoulli* **5**, no. 4, 721–760.

Pistone, G. and Sempi, C. (1995). An infinite-dimensional geometric structure on the space of all the probability measures equivalent to a given one, *Ann. Statist.* **23**, no. 5, 1543–1561.

Rao, M. and Ren, Z. (1991). *Theory of Orlicz spaces*, Monographs and Textbooks in Pure and Applied Mathematics, 146. Marcel Dekker, Inc., New York.

Townsend, J.T., Solomon, B., and Smith, J.S. (2001). The perfect Gestalt: Infinite dimensional Riemannian face spaces and other aspects of face perception. In Wenger, M.J. and Townsend, J.T. (Eds). *Computational, Geometric, and Process Perspectives of Facial Cognition: Contexts and Challenges.* Mahwah, NJ: Lawrence Erlbaum Associates (pp 39-82).