Geometric structures on the non-parametric statistical manifold

Alberto Cena

Direttore del corso di dottorato
prof. Antonio Lanteri

Tutore: prof. Giovanni Pistone

Anno Accademico 2002-2003

XIII ciclo
# Table of contents

Summary ................................. ii  

1 Analytical framework ................ 1  
   1.1 Young functions and Orlicz Spaces .............. 1  
   1.2 Orlicz spaces in Information Geometry ........ 4  

2 Regularity ................................. 9  
   2.1 Analytic mappings between Banach spaces .......... 9  
   2.2 Exponential function exp_{p,a} .................. 12  
   2.3 Analyticity of the Moment Generating Functional .... 15  

3 Structures of manifold ................. 18  
   3.1 Exponential manifold ...................... 18  
   3.2 Mixture manifold ....................... 19  

4 Christoffel symbols and connections .... 26  
   4.1 Connection in a vector bundle ................. 26  
   4.2 Splitting .................................. 35  
   4.3 Exponential Connection .................... 37  
   4.4 Mixture Connection ....................... 41  
   4.5 Geometry of the sphere in $L^a(\mu)$ ........... 42  
   4.6 Amari $\alpha$-embedding ................... 51  
   4.7 $\mathcal{F}^\alpha$-vector bundle ............... 53  
   4.8 An example in the parametric case ............. 57  

5 Finsler structure ...................... 61
Summary

Statistical Manifolds are manifolds such that each point can be identified with a probability density with respect to a given measure. The theory of Statistical Manifolds, the so called Information Geometry, started with a paper of Rao (1945) where a statistical model was considered as a Riemannian manifold with the metric tensor given by the Fisher information matrix. Seminal work of S.-I. Amari added some insight to Information Geometry. Until quite recently the theory was developed only in the parametric, commutative setting (namely finite dimensional manifolds). For references see: Amari (1985); Amari and Nagaoka (2000). The first rigorous infinite dimensional extension has been developed by Pistone and Sempi (1995); Pistone and Rogantin (1999).

Let \((X, \mathcal{X}, \mu)\) be a measure space and let

\[ M_\mu := \left\{ p : X \to \mathbb{R} \text{ measurable} : p > 0 \mu\text{-a.e.}, \int p d\mu = 1 \right\} \]

be the set of all \(\mu\)-almost surely strictly positive probability densities. In the sequel \(M_\mu\) will always refer to this set.

The set \(M_\mu\) can be endowed with a structure of \(C^\infty\)-Banach manifold using the Orlicz space based on an exponentially growing function. With this structure it is called the exponential statistical manifold. In a neighborhood of each density \(p\), \(M_\mu\) is modeled by the subspace \(B_p\) of all the centered random variables of the Orlicz space \(L^{\Phi_1}(p \cdot \mu)\) associated to the Young function \(\Phi_1 = \cosh^{-1}\), with respect to the probability measure \(p \cdot \mu\) and with the Luxembourg norm \(\|\cdot\|_{\Phi_1, p}\).

We review the construction of the exponential statistical manifold in Section 3.1. In Section 3.2 we propose to enlarge the set supporting the manifold structure and we consider

\[ \mathcal{P} := \left\{ p \in L^1(\mu) : \int p d\mu = 1 \right\}. \]

For each \(p \in M_\mu\), if \(\mathcal{E}(p)\) is the connected component of \(M_\mu\) containing \(p\), then

\[ \mathcal{E}(p) \subset {}^*\mathcal{E}(p) := \left\{ q \in \mathcal{P} : \frac{q}{p} \in L^{\Phi_3}(p \cdot \mu) \right\} \]

II
where $L^{\Phi_3}(p \cdot \mu)$ is the Orlicz space associated to the Young function $\Phi_3(x) = (1 + |x|) \log(1 + |x|) - |x|$ with respect to the probability measure $p \cdot \mu$. In Theorem 37 we construct a $C^\infty$-Banach structure on $*\mathcal{E}(p)$. In a neighborhood of each $f \in *\mathcal{E}(p)$ it is modeled by the closed subspace $*B_f$ of all the centered random variables of $L^{\Phi_3}(f \cdot \mu)$.

In Proposition 30 we are able to characterize all the probability densities $q$ with finite Kullback-Leibler divergence with respect to $p$,

$$K(q | p) = \mathbb{E}_p \left[ \frac{q}{p} \log \left( \frac{q}{p} \right) \right] < \infty.$$  

In fact we prove that for each probability density $q$

$$K(q | p) < \infty \iff q \in *\mathcal{E}(p).$$

In Section 1.1 we briefly introduce Young functions and Orlicz spaces. It is known that the Orlicz spaces $L^{\Phi_1}(p \cdot \mu)$ and $L^{\Phi_1}(q \cdot \mu)$ are equal for each pair of densities $p$ and $q$ belonging to a one-dimensional exponential model:

$$p(\vartheta) = e^{\vartheta u - \psi(\vartheta)}r$$

where $r \in \mathcal{M}_\mu$, $u \in L^{\Phi_1}(r \cdot \mu)$ and $\vartheta \in (-\delta, \delta) \subset \mathbb{R}$. Using the fact that an Orlicz space is a Banach ideal space, in Proposition 4 we present a new proof of the equivalence between the Luxembourg norms $\| \cdot \|_{\Phi_1,p}$ and $\| \cdot \|_{\Phi_1,q}$.

The aim of Chapter 2 is the study of the regularity of the moment generating functional

$$M_p : L^{\Phi_1}(p \cdot \mu) \ni u \mapsto \mathbb{E}_p (e^u) = \int e^u p d\mu \in [0, +\infty]$$

In Theorem 22 we prove that the functional $M_p$ is analytic on the open unit ball of $L^{\Phi_1}(p \cdot \mu)$. For each $u$ in a neighborhood of $u_0 \in B(0,1)$, $M_p$ has the power expansion:

$$M_p (u) = \sum_{n=0}^{\infty} \frac{1}{n!} \mathbb{E}_p [\lambda_{1,n} (u_0) \cdot (u - u_0)^n]$$

where $\lambda_{a,n} (u_0) \in L^a_s (L^{\Phi_1} (p \cdot \mu); L^a (p \cdot \mu))$ for $a \geq 1$ is the $n$-multilinear symmetric function defined by

$$\lambda_{a,n} (u_0) (w_1, \ldots, w_n) = \frac{w_1}{a} \cdots \frac{w_n}{a} e^{u_0/w_1},$$

and $\lambda_{a,0} (u_0) = e^{u_0/a} \in L^a (p \cdot \mu)$. With these multilinear maps, in Definition 19, we introduce the exponential map $\exp_{p,a}$ between the open unit ball of the Orlicz space $L^{\Phi_1}(p \cdot \mu)$ and the Lebesgue space $L^a(p \cdot \mu)$ for each $a \geq 1$ and for each density III.
\( p \in \mathcal{M}_\mu \). Using \( \exp_{p,a} \), in Proposition 75, we are able to prove that the Amari \( \alpha \)-embedding

\[ \mathcal{M}_\mu \ni p \mapsto p^a \in L^a(\mu). \]

is analytic.

Amari (1982) and, independently, Čencov (1982) introduced a family of affine connections \( \nabla^\alpha \) with \( \alpha \in [-1,1] \) in the parametric case. Gibilisco and Pistone (1998) show that the pretangent and the tangent bundle over the Exponential Statistical Manifold are the natural domains to define, using parallel transport, respectively the mixture and the exponential connection in the non-parametric case. Then they define the infinite-dimensional version of the \( \alpha \)-connections on a suitable family of vector bundles, the so called \( \alpha \)-bundle. In this infinite dimensional contest, we evaluate the Christoffel symbols of the exponential and mixture connection (see respectively Definition 61 and 64). In Section 4.5 we study the regularity of the sphere in the Lebesgue spaces and in Proposition 74 we find the Christoffel symbols of the natural connection on its tangent bundle. In Proposition 78 we give the Christoffel symbols of the \( \alpha \)-connection on the \( \mathcal{F}^\alpha \)-bundle over \( \mathcal{M}_\mu \) then, in Theorem 82, we prove that they define the connection which is the pull-back of the natural connection on the sphere by the Amari embedding.

In Chapter 5 we show that the exponential statistical manifold admits also Finsler structures. We hope that Finsler structure can give new insight into Information Geometry.
Chapter 1

Analytical framework

1.1 Young functions and Orlicz Spaces

Young functions are a generalization of the family of mappings \( x \mapsto |x|^a / a \) with \( a \geq 1 \) and Orlicz spaces are a generalization of Lebesgue spaces \( L^a \). We begin by recalling their definitions and some material relevant in the construction of the non-parametric statistical manifold. For a general reference see Rao and Ren (1991).

A Young function \( \Phi \) is a convex function \( \Phi : \mathbb{R} \to \mathbb{R} = \mathbb{R} \cup \{+\infty\} \) such that:

i) \( \Phi (0) = 0; \)

ii) \( \Phi (x) = \Phi (-x), \forall x \in \mathbb{R}; \)

iii) \( \lim_{x \to \infty} \Phi (x) = +\infty. \)

The case \( \Phi (x) = +\infty \) for all \( x \geq x_0 > 0 \) is permitted.

Any Young function \( \Phi \) admits an integral representation

\[
\Phi (x) = \int_0^x \varphi (t) \, dt \quad x \in [0, \infty)
\]

where \( \varphi : [0, \infty) \to [0, \infty] \) is a nondecreasing, left continuous function such that \( \varphi (0) = 0 \) and \( \varphi (x) = +\infty \) for \( x \geq x_0 \) when \( \Phi (x) = +\infty \) for \( x \geq x_0 > 0 \).

Let \( \psi \) be the generalized inverse of \( \varphi \), that is the function

\[
\psi (s) = \inf \{ t : \varphi (t) \geq s \} \quad s \in [0, \infty),
\]

then the **conjugate function** of \( \Phi \) is the Orlicz function \( \Psi \) defined by

\[
\Psi (y) = \int_0^y \psi (t) \, dt \quad y \geq 0.
\]
Functions \(| \cdot|^a/a\) and \(| \cdot|^b/b\) with \(a, b > 1\) such that \(a + b = ab\) are a pair of conjugate Young functions. In the extreme case \(a = 1\) and \(b = \infty\), the conjugate of the Young function \(x \mapsto |x|\) is the mapping

\[
x \mapsto \begin{cases} 
0 & \text{if } |x| < 1 \\
+\infty & \text{if } |x| \geq 1.
\end{cases}
\]

The classical Young Inequality \(|xy| \leq |x|^a/a + |y|^b/b\) is generalized by

\[
|xy| \leq \Phi (x) + \Psi (y) \quad x, y \in \mathbb{R}
\]

where equality holds if \(y = \varphi (x)\) or \(x = \psi (y)\).

Let \((X, \mathcal{X}, \mu)\) be a measure space. The Orlicz class \(\tilde{L}^\Phi (\mu)\) associated to the Young function \(\Phi\) is the set defined by

\[
\tilde{L}^\Phi (\mu) := \left\{ u : X \to \mathbb{R}, \text{measurable} : \int \Phi (u) \, d\mu < \infty \right\}.
\]

\(L^a\) is the Orlicz class of the Young function \(|\cdot|^a/a\). However integrability of \(\Phi (u)\) is not enough to generalize the Lebesgue spaces since \(\tilde{L}^\Phi (\mu)\) may not be a vector space.

The Orlicz space \(L^\Phi (\mu)\) associated to the Young function \(\Phi\) is the set of all the measurable functions \(u : X \to \mathbb{R}\) such that \(\Phi (\alpha u)\) is \(\mu\)-integrable for some \(\alpha > 0\):

\[
L^\Phi (\mu) := \left\{ u : X \to \mathbb{R}, \text{measurable} : \exists \alpha > 0 \text{ s. t. } \int \Phi (\alpha u) \, d\mu < \infty \right\}.
\]

\(L^\Phi (\mu)\) is a convex vector space. Inclusion \(\tilde{L}^\Phi (\mu) \subseteq L^\Phi (\mu)\) easily follows by their definitions. In Lemma 5 we will see a condition, the \(\Delta_2\)-condition, sufficient for their equality.

If functions which differ only on sets of measure zero are identified, \(L^\Phi (\mu)\) is turned into a Banach space by the Luxembourg norm defined by

\[
\|u\|_\Phi := \inf \left\{ k > 0 : \int \Phi \left( \frac{u}{k} \right) \, d\mu \leq 1 \right\}.
\]

For each \(u \in L^\Phi (\mu) \setminus \{0\},\)

\[
\int \Phi \left( \frac{u}{\|u\|_\Phi} \right) \, d\mu \leq 1. \tag{1.1}
\]

In fact, there exists a sequence \(k_n \to \|u\|_\Phi\) such that \(k_n \neq 0\) and

\[
\int \Phi \left( \frac{u}{k_n} \right) \, d\mu \leq 1. \tag{1.2}
\]
and, using the Fatou Theorem, one can see that the Inequality (1.1) follows by the bound (1.2).

Let $\Phi$ be invertible when restricted to the positive axis. If $u \in L^\Phi(\mu) \setminus \{0\}$ is fixed, then the mapping $k \mapsto \int \Phi \left( \frac{u}{k} \right) \, d\mu$ is strictly decreasing in the interval $(0, \infty)$ and the following equality holds

$$\int \Phi \left( \frac{u}{\|u\|_\Phi} \right) \, d\mu = 1.$$ 

In particular, in the case of $\Phi = |\cdot|^a / a$, for each $u \in L^a(\mu) \setminus \{0\}$ we have:

$$\int \frac{1}{a} \frac{|u|}{\|u\|_\Phi} \, d\mu = 1 \quad \Rightarrow \quad \|u\|_\Phi = \left( \int \frac{|u|^a}{a} \, d\mu \right)^{\frac{1}{a}} = a^{-\frac{1}{a}} \|u\|_a.$$

If $\Phi$ and $\Psi$ are conjugate Young functions, then for each $u \in L^\Phi(\mu) \setminus \{0\}$ and $v \in L^\Psi(\mu) \setminus \{0\}$ by the Young Inequality we have

$$\frac{|u|}{\|u\|_\Phi} \frac{|v|}{\|v\|_\Psi} \leq \Phi \left( \frac{u}{\|u\|_\Phi} \right) + \Psi \left( \frac{v}{\|v\|_\Psi} \right)$$

and, using Eq. (1.1) after integrating each side of the above inequality, we have

$$\int \frac{|uv|}{\|u\|_\Phi \|v\|_\Psi} \, d\mu \leq \int \Phi \left( \frac{u}{\|u\|_\Phi} \right) \, d\mu + \int \Psi \left( \frac{v}{\|v\|_\Psi} \right) \, d\mu \leq 2$$

so, finally, we obtain the generalized Hölder Inequality

$$\int |uv| \, d\mu \leq 2 \|u\|_\Phi \|v\|_\Psi \quad u \in L^\Phi(\mu), \quad v \in L^\Psi(\mu). \quad (1.3)$$

By (1.3) we see that conjugacy implies that conjugate Orlicz spaces are in a duality relation, that is there exists a continuous bilinear form

$$L^\Phi(\mu) \times L^\Psi(\mu) \ni (u,v) \mapsto \int uvd\mu \in \mathbb{R}.$$ 

Hence we have $L^\Psi(\mu) \subseteq (L^\Phi(\mu))'$ but, in general, $L^\Phi(\mu)$ and $L^\Psi(\mu)$ are not dual.

If $\Phi$ and $\Psi$ are conjugate Young functions, then the Orlicz norm in $L^\Phi(\mu)$ is defined by

$$N_\Phi (u) := \sup \left\{ \int |uv| \, d\mu : v \in L^\Psi(\mu), \int \Psi(v) \, d\mu \leq 1 \right\}.$$ 

Luxemburg and Orlicz norms are equivalent.
Young functions $\Phi_1$ and $\Phi_2$ are said to be *equivalent* if there exist two constants $0 < c_1 \leq c_2 < \infty$ and $x_0 > 0$ such that

$$\Phi_1 (c_1 x) \leq \Phi_2 (x) \leq \Phi_1 (c_2 x) \quad x \geq x_0.$$

If $\Phi_1$ and $\Phi_2$ are equivalent, the Orlicz spaces $L^{\Phi_1} (\mu)$ and $L^{\Phi_2} (\mu)$ are equal as sets and have equivalent norms as Banach spaces.

Young functions are classified according to their growth properties. In the sequel we will need the following condition. A Young function $\Phi : \mathbb{R} \to \mathbb{R}^+$ satisfies the $\Delta_2$-condition if

$$\Phi (2x) \leq k \Phi (x) \quad x \geq x_0 \geq 0$$

for some constant $k > 0$. For example, function $|\cdot|^a / a$ satisfies the $\Delta_2$-condition with Inequality (1.4) holding for every real number with constant equal, for example, to $2^a$.

### 1.2 Orlicz spaces in Information Geometry

We introduce the Orlicz spaces used in the non-parametric Information Geometry. For $i = 1, 2, 3$ let $\Phi_i : \mathbb{R} \to [0, \infty)$ be the following Young functions

- $\Phi_1 : x \mapsto \cosh (x) - 1$
- $\Phi_2 : x \mapsto \exp (|x|) - |x| - 1$
- $\Phi_3 : x \mapsto (1 + |x|) \log (1 + |x|) - |x|$

Henceforth $\Phi_1$, $\Phi_2$ and $\Phi_3$ will always refer to these particular functions.

$\Phi_1$ and $\Phi_2$ are equivalent. $\Phi_2$ and $\Phi_3$ are a pair of conjugate functions. $\Phi_3$ satisfies the $\Delta_2$-condition with Inequality (1.4) holding for all $x \in \mathbb{R}$ if we take a constant equal to 4.

For each positive density $p \in \mathcal{M}_\mu$, the Orlicz space $L^{\Phi_1} (p \cdot \mu)$ can be characterized by the following Proposition 2. Let $u$ be a random variable on $(X, \mathcal{X}, p \cdot \mu)$. The *moment generating function* of $u$ is the map $\hat{u}_p : \mathbb{R} \to [0, +\infty]$ defined by

$$\hat{u}_p (t) := \int e^{tu} p d\mu = E_p (e^{tu})$$

i.e. it is the Laplace transform of the distribution of the random variable $u$ with respect to the probability measure $p \cdot \mu$.

**Definition 1.** For each density $p \in \mathcal{M}_\mu$, the Cramér class at $p$ is the set of all the random variables $u$ on $(X, \mathcal{X}, p \cdot \mu)$ such that moment generating function $\hat{u}_p$ with respect to the probability measure $p \cdot \mu$ is finite in a neighborhood of 0.
\( \mu \)-integrability of \( e^{tu} \) for \( t \) in a neighborhood of 0 implies that \( u \) has finite expectation. The subset of all the random variables with zero expectation is called the centered Cramér class.

**Proposition 2.** For each density \( p \in \mathcal{M}_\mu \), \( L^{\Phi_1}(p \cdot \mu) \) coincides with the Cramér class at \( p \). The centered Cramér class is a closed subspace.

*Proof.* See Pistone and Rogantin (1999, Proposition 3).

**Definition 3.** For each density \( p \in \mathcal{M}_\mu \), we shall denote by \( B_p \) the closed subspace of all the centered random variables in \( L^{\Phi_1}(p \cdot \mu) \), that is

\[
B_p = \left\{ u \in L^1(p \cdot \mu) : 0 \in \text{dom } \hat{u}_p, \mathbb{E}_p(u) = 0 \right\} = \left\{ u \in L^{\Phi_1}(p \cdot \mu) : \mathbb{E}_p(u) = 0 \right\}.
\]

It inherits the structure of Banach space with the Luxembourg norm

\[
\|u\|_{\Phi_1,p} = \inf \left\{ k > 0 : \mathbb{E}_p\left[ \cosh\left( \frac{u}{k} \right) - 1 \right] \leq 1 \right\}
\]

for each \( u \in B_p \).

Let \( F \subseteq S \) be a subset of the set \( S \) of all measurable functions on a measure space. \((F, \| \cdot \|)\) is a Banach ideal space if

i) it is a Banach space;

ii) for each \( u \in S \) and \( v \in F \), \( |u| \leq |v| \) implies \( u \in F \);

iii) the norm is monotone: for each \( u, v \in F \), \( |u| \leq |v| \) implies \( \|u\| \leq \|v\| \).

Orlicz spaces are Banach ideal spaces.

From the main result of Pistone and Sempi (1995) follows the equality (as sets and Banach spaces) between \( L^{\Phi_1}(p \cdot \mu) \) and \( L^{\Phi_1}(q \cdot \mu) \) for each pair of densities \( p \) and \( q \) belonging to a one-dimensional exponential model:

\[
p(\vartheta) = e^{\vartheta u - \psi(\vartheta)} r
\]

where \( r \in \mathcal{M}_\mu \), \( u \in L^{\Phi_1}(r \cdot \mu) \) and \( \vartheta \in (-\delta, \delta) \subseteq \mathbb{R} \). Pistone and Rogantin (1999, Proposition 5) give a direct proof of the equality only as sets. In the following proposition we use the fact that Orlicz spaces are Banach ideal space to prove the equivalence between the norms \( \| \cdot \|_{\Phi_1,p} \) and \( \| \cdot \|_{\Phi_1,q} \).
Proposition 4. Let $p$ and $q$ be densities connected by one-dimensional exponential model, then the identity map

$$
\text{id} : \left( L^{\Phi_1}(p \cdot \mu), \| \cdot \|_{\Phi_1,p} \right) \to \left( L^{\Phi_1}(q \cdot \mu), \| \cdot \|_{\Phi_1,q} \right)
$$

is an homeomorphism.

Proof. We know that $L^{\Phi_1}(p \cdot \mu) = L^{\Phi_1}(q \cdot \mu)$ so the identity map $\text{id}$ is defined. We have to prove its continuity.

Let $\{u_n\} \subset L^{\Phi_1}(p \cdot \mu)$ be a sequence converging in norm $\| \cdot \|_{\Phi_1,p}$ to 0. We suppose it doesn’t converge in norm $\| \cdot \|_{\Phi_1,q}$. In particular, if necessary considering a subsequence, we suppose

$$
\| u_n \|_{\Phi_1,q} > \varepsilon. \tag{1.5}
$$

By the convergence in norm $\| \cdot \|_{\Phi_1,p}$ there is a subsequence $\{u_{n_k}\}$ such that

$$
\| u_{n_k} \|_{\Phi_1,p} < \frac{1}{2^k}.
$$

Since

$$
\sum_{k=1}^{\infty} k \| u_{n_k} \|_{\Phi_1,p} \leq \sum_{k=1}^{\infty} \frac{k}{2^k} < \infty
$$

and $(L^{\Phi_1}(p \cdot \mu), \| \cdot \|_{\Phi_1,p})$ is complete, series $\sum k |u_{n_k}|$ converges in norm and let $r \in L^{\Phi_1}(p \cdot \mu)$ be such that $\sum k |u_{n_k}| = r$.

Claim. Let $\{x_n\}_{n \in \mathbb{N}}$ and $\{y_n\}_{n \in \mathbb{N}}$ be two sequences in a Banach ideal space such that $|x_n| \leq |y_n|$ for each $n \in \mathbb{N}$. If $x_n \to x$ and $y_n \to y$ in norm, then $|x| \leq |y|$ (see Kantorovich and Akilov (1980, p. 142)).

For each $m \in \mathbb{N}$, let $r_m \in L^{\Phi_1}(p \cdot \mu)$ be the partial sum $r_m = \sum_{k=1}^{m} k |u_{n_k}|$. Sequence $r_m$ converges to $r$ in norm $\| \cdot \|_{\Phi_1,p}$. For each $k > 0$, $k |u_{n_k}| \leq r_m$ with $m \geq k$.

Fix $k > 0$, considering the constant sequence $\{x_m := |u_{n_k}|\}_{m \geq k}$ and the sequence $\{r_m\}_{m \geq k}$, by the claim we have

$$
|u_{n_k}| < \frac{r}{k}.
$$

By the monotony of the norm $\| \cdot \|_{\Phi_1,q}$ we have the bound

$$
\| u_{n_k} \|_{\Phi_1,q} < \frac{1}{k} \| r \|_{\Phi_1,q} \to 0 \quad \text{for } k \to \infty
$$

in contrast with the condition (1.5). Hence the identity map $\text{id}$ is continuous and, by symmetry, it is an homeomorphism. \qed
Now we treat the Orlicz space $L_{\Phi}^3$.

**Lemma 5.** Let $(X, \mathcal{X}, \mu)$ be a finite measure space. If a Young function $\Phi$ satisfies the $\Delta_2$-condition then the Orlicz class $\tilde{L}_{\Phi}^3 (\mu)$ coincides with the Orlicz space $L_{\Phi}^3 (\mu)$.

**Proof.** We have to prove only the inclusion $\tilde{L}_{\Phi}^3 (\mu) \supseteq L_{\Phi}^3 (\mu)$. Let $u \neq 0$ be an element of the Orlicz space $L_{\Phi}^3 (\mu)$ with Luxembourg norm $\| u \|_{\Phi}$. There exists $n \in \mathbb{N}$ such that $\| u \|_{\Phi} \leq 2^n$. Since $\Phi$ is even, increasing for positive real values and it satisfies Inequality (1.4) for $|x| \geq x_0 \geq 0$, we have
\[
\int \Phi (u) \, d\mu = \int \Phi \left( \| u \|_{\Phi} \frac{u}{\| u \|_{\Phi}} \right) \, d\mu \\
\leq \int \Phi \left( \frac{2^n u}{\| u \|_{\Phi}} \right) \, d\mu \\
\leq k^n \int \Phi \left( \frac{u}{\| u \|_{\Phi}} \right) \, d\mu + \mu (X) \Phi (2^n x_0) \\
\leq k^n + \mu (X) \Phi (2^n x_0) < \infty.
\]
Hence $\Phi (u)$ is $\mu$-integrable and $u$ belongs to the Orlicz class $\tilde{L}_{\Phi}^3 (\mu)$. $\square$

If the measure space is not finite but the Young function $\Phi$ satisfies the $\Delta_2$-condition with Inequality (1.4) holding for all the real numbers, then Lemma 5 is also true. In fact, in this case let $u \in L_{\Phi}^3 (\mu)$ be a measurable function such that $\Phi (\alpha u)$ is $\mu$-integrable for $\alpha > 0$, if we take $m \in \mathbb{N}$ such that $2^m \alpha \geq 1$, since $\Phi$ is even, increasing for positive real values and it satisfies Inequality (1.4) for every real value we have
\[
\Phi (u) \leq \Phi (2^m \alpha u) \leq k^m \Phi (\alpha u).
\]
Hence $\int \Phi (\alpha u) \, d\mu < \infty$ implies that $\Phi (u)$ is $\mu$-integrable.

**Proposition 6.** Let $(X, \mathcal{X}, p \cdot \mu)$ be a probability space. A random variable $u$ belongs to $L_{\Phi}^3 (p \cdot \mu)$ if and only if $(1 + |u|) \log (1 + |u|)$ is $p \cdot \mu$-integrable.

**Proof.** Let $u$ be a random variable. Since $\Phi_3$ satisfies the $\Delta_2$-condition, by Lemma 5 we have
\[
u \in L_{\Phi_3}^3 (p \cdot \mu) = \tilde{L}_{\Phi_3}^3 (p \cdot \mu) \iff \mathbb{E}_p [(1 + |u|) \log (1 + |u|) - |u|] < \infty.
\]
Since $(1 + x) \log (1 + x) - x > x$ for $x > x_0 > 0$, we have
\[
\mathbb{E}_p [\Phi_3 (u)] < \infty \iff \begin{cases} 
\mathbb{E}_p (|u|) < \infty \quad \text{and} \\
\mathbb{E}_p [(1 + |u|) \log (1 + |u|)] < \infty
\end{cases} \quad (1.6)
\]
\[
\iff \mathbb{E}_p [(1 + |u|) \log (1 + |u|)] < \infty.
\]
$\square$
Remark. Equivalence in line (1.6) shows in particular that if \( u \in L^{\Phi_3} (p \cdot \mu) \) then \( u \in L^1 (p \cdot \mu) \). A basic result in the theory of the Lebesgue space is, in the case of a finite measure space \((\Omega, \Sigma, m)\), the chain of inclusions

\[
L^\infty (m) \subseteq L^r (m) \subseteq L^s (m) \subseteq L^1 (m)
\]

where \( 1 \leq s < r \leq \infty \). Symbol \( \subseteq \) marks the continuity of the inclusion. In fact we have the bound

\[
\| \cdot \|_s \leq k \| \cdot \|_r
\]

where

\[
k (r,s) = \begin{cases} m(\Omega)^{r-s} & \text{if } r < \infty \\ m(\Omega)^{1-s} & \text{if } r = \infty \end{cases}
\]

In our case we can complete with the following chain of inclusion:

\[
L^\infty (p \cdot \mu) \subseteq L^{\Phi_1} (p \cdot \mu) \subseteq L^r (p \cdot \mu) \subseteq L^s (p \cdot \mu) \subseteq L^{\Phi_3} (p \cdot \mu) \subseteq L^1 (p \cdot \mu).
\]

Continuity follows from the existence of constants \( k_1, k_2, k_3 \) and \( k_4 \) such that

\[
\| \cdot \|_1 \leq k_1 \| \cdot \|_{\Phi_3} \leq k_2 \| \cdot \|_s \leq k_2 \| \cdot \|_r \leq k_3 \| \cdot \|_{\Phi_1} \leq k_4 \| \cdot \|_\infty.
\]

Such inequalities depend on the different growth at \( \infty \) of the Young functions involved.

Definition 7. For each density \( p \in \mathcal{M}_\mu \), we shall denote by \( ^*B_p \) or by \( L \log L_0 (p) \) the so called \( x \log x \)-class, that is the subspace of all the centered random variables in \( L^{\Phi_3} (p \cdot \mu) \):

\[
^*B_p = L \log L_0 (p) := \{ v \in L^{\Phi_3} (p \cdot \mu) : \mathbb{E}_p (v) = 0 \}.
\]

Proposition 8. For each \( p \in \mathcal{M}_\mu \) we have the following two statements:

i) All the elements \( ^*u \in ^*B_p \) are identified with an element \( u^* \) of the dual space \( B_p^* \) of \( B_p \) by the formula: \( u^*(v) = \mathbb{E}_p (^*uv) \), with \( v \in B_p \). In particular, \( ^*B_p \) is identified with a proper subset of \( B_p^* \) and the injection of \( ^*B_p \) into \( B_p^* \) is continuous; we write

\[
^*B_p \subseteq B_p^*.
\]

ii) All the elements \( u \in B_p \) are identified with an element \( \overline{u} \) of the dual space \( (^*B_p)^* \) of \( ^*B_p \) by the formula \( \overline{u}(^*u) = \mathbb{E}_p (^*u)^*u \), with \( ^*u \in ^*B_p \). This identification is onto, that is \( B_p \) is identified with \( (^*B_p)^* \); we write:

\[
(^*B_p)^* \simeq B_p.
\]

Proof. See Pistone and Rogantin (1999, Proposition 8). \( \square \)
Chapter 2

Regularity

Before recalling the construction of the Exponential Statistical Manifold in the following Chapter we define the exponential mapping and we study the regularity of the moment generating functional.

Definition 9. The moment generating functional

\[ M_p : L^{p\cdot\mu} \to [0, +\infty] \]

is defined by

\[ M_p(u) := \mathbb{E}_p(e^u) \].

Pistone and Sempi (1995) prove that \( M_p \) is Gâteaux-differentiable in the interior of its proper domain. This result comes from smoothness of the Laplace transform \( \hat{u}_p \) for each \( u \in \text{dom}(M_p) \). Then they prove that \( M_p \) is infinitely Fréchet-differentiable on the open unit ball \( B(0,1) \subset L^{p\cdot\mu} \). Pistone (2001) suggests that the map \( M_p \) is also analytic \( B(0,1) \). Before proving this, we recall some basic facts regarding analytic functions between real Banach spaces (at introductory level, see Prodi and Ambrosetti (1973) and Upmeier (1985, Sections I.1-I.2); for series in Banach spaces see Kadets and Kadets (1997)).

2.1 Analytic mappings between Banach spaces

Let \( E \) and \( F \) be Banach spaces. We denote by \( \mathcal{L}(E;F) \) the Banach space of the continuous linear maps of \( E \) into \( F \), by \( \mathcal{L}^n(E;F) \) the Banach space of the continuous \( n \)-multilinear maps of \( E^n \) into \( F \). The subset of those which are symmetric is denoted by \( \mathcal{L}^n_s(E;F) \). In the case \( F = \mathbb{R} \) these spaces are denoted by \( \mathcal{L}(E) \), \( \mathcal{L}^n(E) \) and \( \mathcal{L}^n_s(E) \).
Definition 10. A function \( \hat{\lambda} : E \to F \) between Banach spaces is called a continuous \( n \)-homogeneous polynomial if there exists \( \lambda \in L^n_s(E; F) \) such that
\[
\hat{\lambda}(x) = \lambda x^n \quad \forall x \in E.
\]
Let \( P^n(E; F) \) be the vector space of all continuous \( n \)-homogeneous polynomials from \( E \) to \( F \). Let \( P^0(E; F) := F \).

The \( n \)-multilinear symmetric function \( \lambda \) is uniquely determined by \( \hat{\lambda} \) (in fact, \( D^n \hat{\lambda} = n! \lambda \)) and it is called the polar form of \( \hat{\lambda} \).

We endow \( P^n(E; F) \) with the norm:
\[
\|\hat{\lambda}\|_{P^n} := \sup_{x \neq 0} \frac{\|\hat{\lambda}(x)\|_F}{\|x\|_E^n} = \sup_{\|x\|_E = 1} \|\hat{\lambda}(x)\|_F.
\]
Clearly \( \|\hat{\lambda}\|_{P^n} \leq \|\lambda\|_{L^n} \). There is an inequality in the opposite direction:
\[
\|\hat{\lambda}\|_{P^n} \leq \|\lambda\|_{L^n} \leq \frac{n^n}{n!} \|\hat{\lambda}\|_{P^n}. \tag{2.1}
\]

Definition 11. Let \( \{\hat{\lambda}_n\}_{n \in \mathbb{N}} \) be a sequence with \( \hat{\lambda}_n \in P^n(E; F) \), then
\[
\sum_{n=0}^{\infty} \hat{\lambda}_n \tag{2.2}
\]
is called a power series from \( E \) to \( F \). The radius of convergence \( \hat{\rho} \) of (2.2) is defined as:
\[
\hat{\rho} := \sup \left\{ r \in [0, \infty) : \sum_{n=0}^{\infty} \|\hat{\lambda}_n\|_{P^n} r^n < \infty \right\}. \tag{2.3}
\]
The radius of restricted convergence \( \rho \) of (2.2) is defined as
\[
\rho := \sup \left\{ r \in [0, \infty) : \sum_{n=0}^{\infty} \|\lambda_n\|_{L^n} r^n < \infty \right\}.
\]

For the Cauchy-Hadamard formula,
\[
\frac{1}{\hat{\rho}} = \limsup_{n \to \infty} \|\hat{\lambda}_n\|_{P^n}^{\frac{1}{n}} \quad \text{and} \quad \frac{1}{\rho} = \limsup_{n \to \infty} \|\lambda_n\|_{L^n}^{\frac{1}{n}}.
\]
By (2.1) and recalling that \( \lim_{n \to \infty} \left( \frac{n^n}{n!} \right)^{1/n} = e \), we have
\[
\frac{\hat{\rho}}{e} \leq \rho \leq \hat{\rho}.
\]
2.1– Analytic mappings between Banach spaces

The inequality \( \| \hat{\lambda}_n (x) \|_F \leq \| \hat{\lambda}_n \|_{P^n} \| x \|_E^n \) implies that the series \( \sum_{n=0}^{\infty} \hat{\lambda}_n (x) \) converges absolutely and uniformly in the closed ball \( \overline{B}(0,r) \) for each \( r < \hat{\rho} \), moreover (Upmeier, 1985, Proposition 1.4)

\[
\hat{\rho} = \sup \left\{ r : \sum_{n=0}^{\infty} \hat{\lambda}_n (x) \text{ conv. uniformly for } \| x \|_E \leq r \right\}.
\] (2.4)

Similarly, series \( \sum_{n=0}^{\infty} \lambda_n (x_1, \ldots, x_n) \) converges uniformly for every sequence \( \{x_m\}_{m \geq 1} \) such that \( \sup \{ \| x_m \|_E \} \leq r \) with fixed \( r < \rho \) and

\[
\rho = \sup \left\{ r : \sum_{n=0}^{\infty} \lambda_n (x_1, \ldots, x_n) \text{ conv. unif. } \forall \{x_m\}_{m \geq 1} \text{ s.t. } \sup \| x_m \|_E \leq r \right\}.
\]

**Definition 12.** Let \( E \) and \( F \) be Banach spaces and \( U \subset E \) be open. A mapping \( f : U \to F \) is called analytic if for each \( x_0 \in U \) there exists a convergent power series \( \sum \hat{\lambda}_n \) with positive radius of convergence such that, for each \( x \) in a neighborhood of \( x_0 \),

\[
f (x) = \sum_{n=0}^{\infty} \hat{\lambda}_n (x - x_0).
\] (2.5)

Series \( \sum_{n=0}^{\infty} \hat{\lambda}_n (x) \) is an analytic function in the ball \( B(0,\rho) \) of restricted convergence.

An analytic function \( f \) is infinitely often differentiable. If \( f \) has the power expansion (2.5), then \( D^k f (x_0) = k! \lambda_k \) and \( D^k f \) has the following power expansion about \( x_0 \)

\[
D^k f (x) = \sum_{n=0}^{\infty} \frac{(k+n)!}{n!} \lambda_{k+n} (x - x_0)^n.
\] (2.6)

Series on the right-hand side of Equations (2.5) and (2.6) have the same radius of restricted convergence.

Like in the real case, there are the following theorems for the representation in power series of smooth functions and for the composition of analytic functions (Prodi and Ambrosetti, 1973, Theorem 10.5 and 11.1).

**Theorem 13.** Let \( E \) and \( F \) be Banach spaces and \( U \subset E \) be open. Let \( f : U \to F \) be a smooth function. If there exists a constant \( M \) such that for all \( x \in B(x_0,r) \subset U \)

\[
\| D^n f (x) \|_{L^n} \leq \frac{M n!}{r^n}
\]

then \( f \) is represented by the Taylor series:

\[
f (x) = \sum_{n=0}^{\infty} \frac{1}{n!} D^n f (x_0) (x - x_0)^n
\]
and \( f \) is analytic in \( B(x_0, r) \).

**Theorem 14.** Let \( E, F \) and \( G \) be Banach spaces, \( U \subset E \) and \( V \subset F \) be open. If \( f : U \to F \) and \( g : V \to G \) are analytic functions such that \( f(U) \subset V \) then \( g \circ f \) is also analytic.

### 2.2 Exponential function \( \exp_{p,a} \)

We shall define an analytic function between the open unit ball \( B(0,1) \) of the Orlicz space \( L^\Phi_1(p \cdot \mu) \) and the Lebesgue space \( L^a(p \cdot \mu) \).

**Lemma 15.** In the Orlicz space \( L^\Phi_1(p \cdot \mu) \) with the Luxembourg norm,

\[
\|u\|_{\Phi_1} \leq 1 \iff E_p(\cosh u) \leq 2.
\]

**Proof.** As we observed in Section 1.1, since \( \Phi_1 \) is strictly increasing and continuous on \([0, \infty)\), the mapping \((0, \infty) \ni k \mapsto E_p(\Phi_1(u/k)) \in \mathbb{R} \) is strictly decreasing for each \( 0 \neq u \in L^\Phi_1(p \cdot \mu) \) and

\[
\|u\|_{\Phi_1} \leq 1 \iff E_p(\Phi_1(u)) \leq 1 \iff E_p(\cosh u) \leq 2.
\]

**Lemma 16.** For each \( a \geq 1, n \in \mathbb{N}^* \) and \( u \in B(0,1) \), let \( \lambda_{a,n}(u) \) be defined by:

\[
\lambda_{a,n}(u) : L^\Phi_1(p \cdot \mu) \times \cdots \times L^\Phi_1(p \cdot \mu) \to L^a(p \cdot \mu)
\]

\[
(w_1, \ldots, w_n) \mapsto \frac{w_1}{a} \cdots \frac{w_n}{a} e^{u/a}
\]

Then \( \lambda_{a,n}(u) \) are continuous, symmetric, \( n \)-multilinear mappings.

**Proof.** Put \( r = n^{-1} (1 - \|u\|_{\Phi_1}) \). For each \( v_1, \ldots, v_n \in L^\Phi_1(p \cdot \mu) \) with \( \|v_i\|_{\Phi_1} = 1 \), we have

\[
\left\| u + r \sum_{i=1}^n |v_i| \right\|_{\Phi_1} \leq \|u\|_{\Phi_1} + rn = 1
\]

so \( u + r \sum_{i=1}^n |v_i| \in B(0,1) \).

By the inequality \( \frac{|x|^n}{a^n} < e^{|x|} \) for each \( x \in \mathbb{R} \) and by Lemma 15, we have

\[
\int \left| \frac{v_1}{a} \cdots \frac{v_n}{a} e^{u \cdot pd\mu} \right|^a \leq E_p[e^{u + r \sum_{i=1}^n |v_i|}]
\]

\[
\leq 2E_p\left[ \cosh \left( u + r \sum_{i=1}^n |v_i| \right) \right] \leq 4
\]
hence
\[
\int \left| \sum_{n=1}^\infty \frac{v_1 \cdots v_n}{a^n} e^{u/a} \right|^a \, a pd\mu \leq \frac{4n^a}{R^n} = C_n(u) < \infty. \tag{2.7}
\]

For arbitrary \( w_1, \ldots, w_n \in L^{p, \mu} \) \( \setminus \{0\} \) if \( v_i = w_i / \| w_i \|_{p, \mu} \), then
\[
\int \left| \sum_{n=1}^\infty \frac{v_1 \cdots v_n}{a^n} e^{u/a} \right|^a \, a pd\mu = \| w_1 \|_{p, \mu} \cdots \| w_n \|_{p, \mu} \frac{C_n(u)}{n^a} (\| u \|_{p, \mu} / \| u \|_{p, \mu}) \tag{2.8}
\]

By Equations (2.8) and (2.7) we have
\[
\| \lambda_{a,n}(u) (w_1, \ldots, w_n) \|_a = \left( \int \left| \sum_{n=1}^\infty \frac{v_1 \cdots v_n}{a^n} e^{u/a} \right|^a \, a pd\mu \right)^{1/a} \leq (C_n(u))^{1/a} \| w_1 \|_{p, \mu} \cdots \| w_n \|_{p, \mu} \tag{2.9}
\]

hence \( \lambda_{a,n}(u) \in L^n_a (L^{p, \mu} (\mathbb{R}^d); L^a (\mathbb{R}^d)) \).

If \( u \in B(0,1) \), then \( e^{u/a} \in L^a (\mathbb{R}^d) \). In fact, since \( \| u \|_{p, \mu} < 1 \),
\[
\int |e^{u/a}|^a \, a pd\mu \leq E_p [e^{u}] \leq 2E_p [\cosh (|u|)] \leq 2E_p \left[ \cosh \left( \frac{|u|}{\| u \|_{p, \mu}} \right) \right] \leq 4.
\]

**Definition 17.** For each \( a \geq 1 \) and \( u \in B(0,1) \), let
\[
\hat{\lambda}_{a,0}(u) := e^{u/a}
\]

and let \( \hat{\lambda}_{a,n}(u) \in P^n (L^{p, \mu}, L^a (\mathbb{R}^d)) \) with \( n \geq 1 \) be the \( n \)-homogeneous polynomial determined by the polar form \( \lambda_{a,n}(u) \) defined in Lemma 16.

**Proposition 18.** Let \( p \in M_\mu \) and \( a \geq 1 \). Series
\[
A(v) = \sum_{n=0}^\infty \frac{1}{n!} \left( \frac{v}{a} \right)^n
\]
is a power series from \( L^{p, \mu} \) to \( L^a (\mathbb{R}^d) \) with radius of convergence \( \hat{\rho} \geq 1 \).

**Proof.** By Lemma 16
\[
A(v) = \sum_{n=0}^\infty \frac{1}{n!} \left( \frac{v}{a} \right)^n = \sum_{n=0}^\infty \frac{1}{n!} \hat{\lambda}_{a,n}(0) v^n
\]
and \(A = \frac{1}{n!} \hat{\lambda}_{a,n} (0)\) is a power series from \(L^{\Phi_1} (p \cdot \mu)\) to \(L^a (p \cdot \mu)\).

Let \(v \in L^{\Phi_1} (p \cdot \mu)\) be such that \(\|v\|_{\Phi_1} = 1\). By the inequality \(\frac{1}{n!} \left( \frac{|v|}{a} \right)^n \leq e^{\|v\|/a}\), we have

\[
\left\| \frac{1}{n!} \hat{\lambda}_{a,n} (0) (v) \right\|_a = \left( \int \left| \frac{1}{n!} \left( \frac{v}{a} \right)^n \right| a \, pd\mu \right)^{\frac{1}{a}} \leq \left( \mathbb{E}_p \left[ e^{\|v\|} \right] \right)^{\frac{1}{a}} \leq \left( 2 \mathbb{E}_p [\cosh (|v|)] \right)^{\frac{1}{a}} \leq 4^{\frac{1}{a}}
\] (2.10)

and \(\left\| \frac{1}{n!} \hat{\lambda}_{a,n} (0) \right\|_{p_n} \leq 4^{1/a}\). Hence \((1/\hat{\rho}) \leq \lim_{n \to \infty} 4^{1/(an)} = 1\) and \(\hat{\rho} \geq 1\).

**Definition 19.** For each \(a \geq 1\) and for each density \(p \in \mathcal{M}_\mu\), the exponential function \(\exp_{p,a}\) between the open unit ball of the Orlicz space \(L^{\Phi_1} (p \cdot \mu)\) and the Lebesgue space \(L^a (p \cdot \mu)\) is defined by

\[
\exp_{p,a} : B (0,1) \to L^a (p \cdot \mu)
\]

\[
v \mapsto \sum_{n=0}^{\infty} \frac{1}{n!} \left( \frac{v}{a} \right)^n.
\]

Let \(v \in B (0,1) \subset L^{\Phi_1} (p \cdot \mu)\). We know that, pointwise, \(\sum_{k=0}^{\infty} \frac{1}{k!} \left( \frac{v}{a} \right)^k = e^{v/a}\). Since \(p \cdot \mu\) is a finite measure, \(\sum_{k=0}^{\infty} \frac{1}{k!} \left( \frac{v}{a} \right)^k \to e^{v/a}\) also in \(p \cdot \mu\)-probability. Proposition 18 shows that it converges to \(\exp_{p,a} (v)\) in norm \(\|\cdot\|_a\) so it is mean convergent to \(\exp_{p,a} (v)\) in \(L^a (p \cdot \mu)\). Since convergence in mean implies convergence in measure (see Sempi (1986)), \(\sum_{k=0}^{\infty} \frac{1}{k!} \left( \frac{v}{a} \right)^k \to \exp_{p,a} (v)\) in \(p \cdot \mu\)-probability and, by uniqueness,

\[
\exp_{p,a} (v) = e^{\frac{v}{a}}.
\]

**Proposition 20.** Exponential function \(\exp_{p,a}\) satisfies the following properties:

(i) \(\exp_{p,a} (0) = 1\);

(ii) for each \(u,v \in B (0,1)\) such that \(u + v \in B (0,1)\)

\[
\exp_{p,a} (u + v) = \exp_{p,a} (u) \exp_{p,a} (v);
\]

(iii) for each \(u \in B (0,1)\), \(\exp_{p,a} (u)\) has an inverse \((\exp_{p,a} (u))^{-1}\) in \(L^a (p \cdot \mu)\) and

\[
(\exp_{p,a} (u))^{-1} = \exp_{p,a} (-u).
\]
2.3– Analyticity of the Moment Generating Functional

Proof. (i) \( \exp_{p,a} (0) = \lambda_{a,0} (0) = e^0 = 1. \)

(ii) Since in a Banach space absolute convergence implies unconditional convergence (converse is true in \( \mathbb{R}^n \), but not in the general case), with a rearrangement of terms we have:

\[
\exp_{p,a} (u + v) = \sum_{n=0}^{\infty} \frac{1}{n!} \sum_{m=0}^{n} \binom{n}{m} \left( \frac{u}{a} \right)^{n-m} \left( \frac{v}{a} \right)^m = \exp_{p,a} (u) \exp_{p,a} (v).
\]

(iii) It comes from (i) and (ii) putting \( v = -u. \)

Theorem 21. Let \( p \in \mathcal{M} \) and \( a \geq 1. \) Mapping \( \exp_{p,a} \) is an analytic function. In a neighborhood of each \( u_0 \in B (0,1) \) it can be expanded in the Taylor series

\[
\exp_{p,a} (u_0 + u) = \sum_{n=0}^{\infty} \frac{1}{n!} \left( \frac{u}{a} \right)^n e^{u_0/a}.
\]  

(2.11)

Proof. The upper bound of \( \| \lambda_{a,n} (0) \|_{\mathcal{L}^n} \) coming from (2.9) suffices to demonstrate analyticity of \( \exp_{p,a} \) only in a neighborhood of 0. In fact,

\[
\left\| \frac{1}{n!} \lambda_{a,n} (0) \right\|_{\mathcal{L}^n} = \sup_{\| w_i \|_{\Phi_1} \neq 0} \left\| \frac{1}{n!} \lambda_{a,n} (0) (w_1, \ldots, w_n) \right\|_{\mathcal{L}^n} \leq \frac{(C_n (0))^2}{n!} = \frac{4^{1/2} n^n}{n!}
\]

and, since \( \lim_{n \to \infty} \left( \frac{1}{n!} \lambda_{a,n} (0) \right)^{1/n} \leq \lim_{n \to \infty} \left( 4^{1/2} n^n \right)^{1/n} = e, \) for the Cauchy Hadamard formula, we find that the radius of restricted convergence \( \rho \) is not lower than \( 1/e. \)

Let \( u_0 \in B (0,1). \) Repeating previous argument for the \( n \)-homogeneous polynomial \( \frac{1}{n!} \hat{\lambda}_{a,n} (u_0), \) we find that the power series \( \sum_{n=0}^{\infty} \frac{1}{n!} \hat{\lambda}_{a,n} (u_0) \) has radius of restricted convergence \( \rho \geq (1 - \| u_0 \|_{\Phi_1}) / e > 0. \) Hence in a neighborhood of \( u_0, \) mapping \( u \mapsto \sum_{n=0}^{\infty} \frac{1}{n!} \hat{\lambda}_{a,n} (u_0) (u) \) is analytic and

\[
\sum_{n=0}^{\infty} \frac{1}{n!} \hat{\lambda}_{a,n} (u_0) (u) = \sum_{n=0}^{\infty} \frac{1}{n!} \left( \frac{u}{a} \right)^n e^{u_0/a} = e^{u_0/a} = \exp_{p,a} (u_0 + u).
\]

\( \square \)

2.3 Analyticity of the Moment Generating Functional

We are now able to improve the results of Pistone and Sempi (1995) about the regularity of the Moment Generating Functional. We first note that \( M_p, \) restricted to the open unit ball \( B (0,1), \) coincides with \( \mathbb{E}_{p} \circ \exp_{p,1}. \)
Theorem 22. The moment generating functional $M_p$ satisfies the following properties:

(i) $M_p(0) = 1$ otherwise, for each $u \neq 0$, $M_p(u) > 1$;

(ii) it is convex and lower semicontinuous, its proper domain

$$\text{dom}(M_p) = \{ u \in L^\Phi_1(p \cdot \mu) : M_p(u) < \infty \}$$

is a convex set which contains the open unit ball $B(0,1) \subset L^\Phi_1(p \cdot \mu)$;

(iii) it is infinitely Gâteaux-differentiable in the interior of its proper domain, the $n$th-derivative at $u \in \text{dom}(M_p)$ in the direction $v \in L^\Phi_1(p \cdot \mu)$ is

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

(iv) it is bounded, infinitely Fréchet-differentiable and analytic on the open unit ball of $L^\Phi_1(p \cdot \mu)$, the $n$th-derivative at $u \in B(0,1)$ evaluated in $(v_1, \ldots, v_n) \in L^\Phi_1(p \cdot \mu) \times \cdots \times L^\Phi_1(p \cdot \mu)$ is

$$D^n M_p(u)(v_1, \ldots, v_n) = E_p(v_1 \cdots v_n e^u).$$

In particular, $DM_p(0) = E_p$.


(iii) If $u \in \text{dom}(M_p)$, for every $v \in L^\Phi_1(p \cdot \mu)$, $u + tv \in \text{dom}(M_p)$ for $t$ small enough

$$M_p(u + tv) = \int e^{tv} e^u p d\mu = M_p(u) E_q(e^{tv})$$

where $q = e^u p / M_p(u)$. $E_q(e^{tv})$ is the Laplace transform of $v$ with respect to the probability density $q$, it is analytic in the interior of its proper domain (see Pistone and Sempi (1995)) and

$$E_q(e^{tv}) = \sum_{k=0}^\infty \frac{E_q(v^k)}{k!} t^k$$

hence

$$\left. \frac{d^n}{dt^n} M_p(u + tv) \right|_{t=0} = M_p(u) \sum_{k=n}^\infty \frac{E_q(v^k)}{(k-n)!} t^{k-n}$$

and

$$\left. \frac{d^n}{dt^n} M_p(u + tv) \right|_{t=0} = M_p(u) E_q(v^n) = E_p(v^n e^u).$$
(iv) For each \( u \in B(0,1) \) and \( n \in \mathbb{N} \), we have \( \mathbb{E}_p \left( \hat{\lambda}_{1,n} (u) \right) \in \mathcal{P}^n \left( L^{\Phi_1} (p \cdot \mu); \mathbb{R} \right) \) and \( \sum_{n=0}^{\infty} \frac{1}{n!} \mathbb{E}_p \left( \hat{\lambda}_{1,n} (u) \right) \) is a power series from \( L^{\Phi_1} (p \cdot \mu) \) to \( \mathbb{R} \) with positive radius of convergence. In a neighborhood of each \( u_0 \in B(0,1) \), by (2.11) we have

\[
M_p (u) = \mathbb{E}_p (\exp_{p,1} (u)) = \int \sum_{n=0}^{\infty} \frac{1}{n!} \lambda_{1,n} (u_0) (u - u_0)^n \, p \, d\mu.
\]

Integrating term by term we obtain the following expansion in power series about \( u_0 \):

\[
M_p (u) = \sum_{n=0}^{\infty} \frac{1}{n!} \mathbb{E}_p \left( \hat{\lambda}_{1,n} (u_0) (u - u_0) \right) = \sum_{n=0}^{\infty} \frac{1}{n!} \mathbb{E}_p \left( (u - u_0)^n e^{u_0} \right).
\]

Hence \( M_p \) is an analytic function. Its \( n \)th-derivative at \( u \) in the directions \((v_1, \ldots, v_n) \in L^{\Phi_1} (p \cdot \mu) \times \cdots \times L^{\Phi_1} (p \cdot \mu)\) is

\[
D^n M_p (u) (v_1, \ldots, v_n) = \mathbb{E}_p (\lambda_{1,n} (u) (v_1, \ldots, v_n)) = \mathbb{E}_p (v_1 \cdots v_n \, e^u).
\]

**Definition 23.** The cumulant generating functional

\[ K_p : B_p \to [0, +\infty] \]

is defined by

\[ K_p (u) := \log (M_p (u)). \]

**Corollary 24 (of Theorem 22).** For each density \( p \in \mathcal{M}_\mu \), the cumulant generating functional \( K_p : B_p \cap B(0,1) \to [0,\infty) \) is an analytic function.

**Proof.** It is a composition of analytic functions. \( \square \)
Chapter 3

Structures of manifold

3.1 Exponential manifold

First we review the construction, due to Pistone and Sempi (1995), of the structure of manifold over $\mathcal{M}_\mu$.

We refer to the presentation of infinite dimensional manifold of Lang (1995).

For each $p \in \mathcal{M}_\mu$ let $\mathcal{V}_p$ be the open unit ball in the Orlicz space $B_p$, that is

$$\mathcal{V}_p := \left\{ u \in B_p : \|u\|_{\Phi_1(p)} < 1 \right\}.$$

It is a subset of the proper domain of the cumulant generating functional $K_p$, and we can define the mapping

$$e_p : \mathcal{V}_p \to \mathcal{M}_\mu, \quad u \mapsto e^{u-K_p(u)}.$$

This function is one to one. In fact if $e_p(u_1) = e_p(u_2)$, then $u_1 - K_p(u_1) = u_2 - K_p(u_2)$ and $u_1 - u_2$ is constant. Since $u_1, u_2 \in B_p$ this constant has to be 0.

The range of $e_p$ is denoted by $U_p$. The inverse of $e_p$ on $U_p$ is the function

$$s_p : U_p \ni q \mapsto \log \left( \frac{q}{p} \right) - E_p \left[ \log \left( \frac{q}{p} \right) \right] \in \mathcal{V}_p.$$

Each pair $(U_p, s_p)$ for $p \in \mathcal{M}_\mu$ is a chart. The centered log-likelihood $s_p$ is the coordinate function.

If $\mathcal{U}_{p_1} \cap \mathcal{U}_{p_2} \neq \emptyset$ for a pair of densities $p_1, p_2 \in \mathcal{M}_\mu$, then the transition mapping $s_{p_2} \circ e_{p_1}$ is the affine function

$$s_{p_2} \circ e_{p_1} : s_{p_1}(\mathcal{U}_{p_1} \cap \mathcal{U}_{p_2}) \to s_{p_2}(\mathcal{U}_{p_1} \cap \mathcal{U}_{p_2})$$

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

The derivative of the overlap map \( s_{p_2} \circ c_{p_1} \) is

\[
B_{p_1} \ni u \mapsto u - E_{p_2}(u) \in B_{p_2}
\]

which is a toplinear isomorphism between \( B_{p_1} \) and \( B_{p_2} \).

**Definition 25.** The sequence \( \{p_n\}_{n \in \mathbb{N}} \) in \( \mathcal{M}_\mu \) is e-convergent (exponentially convergent) to \( p \) if \( \{p_n\}_{n \in \mathbb{N}} \) tends to \( p \) in \( \mu \)-measure as \( n \to \infty \) and moreover the sequence \( \{\frac{p}{p_n}\}_{n \in \mathbb{N}} \) and \( \{\frac{p_n}{p}\}_{n \in \mathbb{N}} \) are eventually bounded in each \( L^\alpha(p \cdot \mu) \), \( \alpha > 1 \), that is

\[
\forall \alpha > 1 : \limsup_{n \to \infty} E_p\left[\left(\frac{p_n}{p}\right)^\alpha\right] < \infty, \quad \limsup_{n \to \infty} E_p\left[\left(\frac{p}{p_n}\right)^\alpha\right] < \infty.
\]

We report the main Theorem of Pistone and Sempi (1995)

**Theorem 26.** The collection of pairs \( \{U_p, s_p\} : p \in \mathcal{M}_\mu \) is an affine \( C^\infty \) atlas on \( \mathcal{M}_\mu \). The induced topology is equivalent to the e-convergence.

**Definition 27.** The exponential statistical manifold is the manifold defined by the atlas in Theorem 26.

**Definition 28.** For every density \( p \in \mathcal{M}_\mu \), the maximal exponential model at \( p \) is defined to be the family of densities

\[
\mathcal{E}(p) := \left\{ e^{u-K_p(u)}p : u \in \text{dom}^\circ K_p \right\} \subseteq \mathcal{M}_\mu.
\]

Pistone and Sempi (1995, Theorem 4.1) prove that \( \mathcal{E}(p) \) is the connected component of the manifold \( \mathcal{M}_\mu \) containing \( p \).

### 3.2 Mixture manifold

Let \( p \in \mathcal{M}_\mu \) be a probability density. For each \( u \in \mathcal{V}_p \) and \( q = e^{u-K_p(u)}p \), the derivative of \( K_p \) at \( u \), \( DK_p(u) \in B_p^* \), is the linear mapping

\[
DK_p(u) \cdot v = E_p\left[\left(\frac{q}{p} - 1\right) v\right], \quad v \in B_p
\]

and \( DK_p(u) \) is identified to \( q/p - 1 \in *B_p \) according to Proposition 8 (see Pistone and Rogantin (1999, Proposition 16-d)). Mapping \( U_p \ni q \mapsto q/p - 1 \in *B_p \) cannot be a chart because the values are bounded below by \(-1\).
We enlarge the maximal statistical model $\mathcal{M}_\mu$ not assuming the condition $p > 0$ any more. Let us consider the set of all the probability densities relative to the measure $\mu$

$$\mathcal{P}_\mu := \left\{ p \in L^1(\mu) : p \geq 0, \int pd\mu = 1 \right\}$$

and the set

$$\mathcal{P} := \left\{ p \in L^1(\mu) : \int pd\mu = 1 \right\}.$$ 

Observe that $\mathcal{M}_\mu \subseteq \mathcal{P}_\mu \subseteq \mathcal{P}$. For each $q \in \mathcal{P}$ there exists an element $\tilde{q} \in \mathcal{P}_\mu$ defined by

$$\tilde{q} := \frac{|q|}{\int |q| d\mu}.$$

For each probability density $p \in \mathcal{P}_\mu$, let us introduce the subset $^*\mathcal{U}_p$ of $\mathcal{P}$ defined by

$$^*\mathcal{U}_p := \left\{ q \in \mathcal{P} : \frac{q}{p} \in L^\Phi(\mu) \right\}.$$ 

Then consider the following map $\eta_p$ defined on $^*\mathcal{U}_p$

$$\eta_p : ^*\mathcal{U}_p \to ^*B_p$$

$$q \mapsto \frac{q}{p} - 1.$$ 

This mapping is bijective and its inverse is:

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

The collection of sets $\{^*\mathcal{U}_p\}_{p \in \mathcal{P}_\mu}$ is a covering of $\mathcal{P}$. In fact, for each $q \in \mathcal{P}$ we have $q \in ^*\mathcal{U}_q$.

Let us characterize the elements of $^*\mathcal{U}_p \cap \mathcal{P}_\mu$: they are all the probability densities with definite Kullback-Leibler divergence with respect to $p$.

**Definition 29.** Let $p \in \mathcal{M}_\mu$ and $q \in \mathcal{P}_\mu$ be given. If $(q/p) \log (q/p)$ is $p \cdot \mu$ integrable, then the Kullback-Leibler divergence or relative entropy of $q$ with respect to $p$ is the number:

$$K(q | p) = \mathbb{E}_p \left[ \frac{q}{p} \log \left( \frac{q}{p} \right) \right].$$

**Proposition 30.** Let $p \in \mathcal{M}_\mu$ be given. For each $q \in \mathcal{P}$, the relative entropy $K(\tilde{q} | p)$ of the probability density $\tilde{q}$ with respect to $p$ is definite if and only if $q \in ^*\mathcal{U}_p$:

$$K(\tilde{q} | p) = \mathbb{E}_p \left[ \frac{\tilde{q}}{p} \log \left( \frac{\tilde{q}}{p} \right) \right] < \infty \iff q \in ^*\mathcal{U}_p.$$
Proof. Let \( q \in \mathcal{P} \). We assume that \( K ( \tilde{q} | p ) < \infty \). By the inequality \( \Phi_3 (x) \leq 1 + x \log (x) \) for \( x > 0 \) we have

\[
\mathbb{E}_p \left[ \Phi_3 \left( \left( \int |q| d\mu \right)^{-1} \frac{q}{p} \right) \right] = \mathbb{E}_p \left[ \Phi_3 \left( \frac{\tilde{q}}{p} \right) \right] \\
\leq 1 + \mathbb{E}_p \left[ \frac{\tilde{q}}{p} \log \left( \frac{\tilde{q}}{p} \right) \right] = 1 + K ( \tilde{q} | p ) < \infty.
\]

Hence \( \frac{q}{p} \in L^{\Phi_3} (p \cdot \mu) \) and, by definition, \( q \in \ast \mathcal{U}_p \).

Conversely, let \( q \in \ast \mathcal{U}_p \), that is \( \frac{q}{p} \in L^{\Phi_3} (p \cdot \mu) \). Since \( L^{\Phi_3} (p \cdot \mu) \) is linear, \((\int |q| d\mu)^{-1} \frac{q}{p} \in L^{\Phi_3} (p \cdot \mu) \) and, by Proposition 6, \((1 + \frac{q}{p}) \log (1 + \frac{q}{p}) \) is \( p \cdot \mu \) integrable. From the inequality

\[
x \log^+ (x) \leq (1 + x) \log (1 + x), \quad x > 0
\]

where \( \log^+ (x) := \max \{0, \log (x)\} \), we have:

\[
\mathbb{E}_p \left[ \frac{\tilde{q}}{p} \log \left( \frac{\tilde{q}}{p} \right) \right] \leq \mathbb{E}_p \left[ \frac{\tilde{q}}{p} \log^+ \left( \frac{\tilde{q}}{p} \right) \right] + 1 \\
\leq \mathbb{E}_p \left[ \left( 1 + \frac{\tilde{q}}{p} \right) \log \left( 1 + \frac{\tilde{q}}{p} \right) \right] + 1 < \infty
\]

hence \( K ( \tilde{q} | p ) \) is defined.

**Corollary 31.** The relative entropy \( K (q | p) \) of a probability density \( q \in \mathcal{P}_\mu \) with respect to \( p \in \mathcal{M}_\mu \) is defined if and only if \( \log \left( \frac{q}{p} \right) \in L^1 (q \cdot \mu) \).

**Proof.** If \( q \in \mathcal{P}_\mu \), then \( q = \tilde{q} \) and we have

\[
K (q | p) < \infty \iff \mathbb{E}_p \left[ \frac{q}{p} \log \left( \frac{q}{p} \right) \right] < \infty \iff \mathbb{E}_q \left[ \left| \log \left( \frac{q}{p} \right) \right| \right] < \infty.
\]

**Proposition 32.** Let \( p \in \mathcal{M}_\mu \) be given, then \( \mathcal{U}_p \subset \ast \mathcal{U}_p \).

**Proof.** If \( q \in \mathcal{U}_p \), there exists \( u \in \mathcal{V}_p \subset B_p \) such that \( q = e^{u-K_p(u)}p \). Since \( L^{\Phi_1} (q \cdot \mu) = L^{\Phi_1} (p \cdot \mu) \) the random variable \( u \) is \( q \cdot \mu \)-integrable and we have

\[
\mathbb{E}_q \left[ \log \left( \frac{q}{p} \right) \right] = \mathbb{E}_q (u) - K_p (u) < \infty.
\]

From the Corollary 31 follows that \( q \in \ast \mathcal{U}_p \).
Using results of Section 2.3 and Equation (3.1), we can conclude that for each \( p \in M \), the function

\[
\mathcal{V}_p \ni u \mapsto K(q|p) \in \mathbb{R}
\]

where \( q = e^{u-K_p(u)}p \) is smooth. In fact

\[
K(q|p) = E_q(u) - K_p(u) \\
= E_p(ue^u) - K_p(u) \\
= DM_p(u) \cdot u - K_p(u) \\
= DK_p(u) \cdot u - K_p(u).
\]

**Proposition 33.** Let \( p_1 \) and \( p_2 \) be two densities in the same connected component \( \mathcal{E}(p) \subset M_\mu \) for some density \( p \in M_\mu \). Then the mapping

\[
U_{p_1p_2} : L^{\Phi_3}(p_1 \cdot \mu) \rightarrow L^{\Phi_3}(p_2 \cdot \mu) \quad u \mapsto \frac{u}{p_1/p_2}
\]

is a toplinear isomorphism.

**Proof.** Let \( u \in L^{\Phi_3}(p_1 \cdot \mu) \). The Orlicz norm \( N_{\Phi_3,p_2} \) of \( up_1/p_2 \) is, if it exists, the number

\[
N_{\Phi_3,p_2} \left( u \frac{p_1}{p_2} \right) = \sup \left\{ \int \left| \left( u \frac{p_1}{p_2} \right) v \right| p_2d\mu : v \in L^{\Phi_2}(p_2 \cdot \mu), E_{p_2} [\Phi_2(v)] \leq 1 \right\} \\
= \sup \left\{ E_{p_1} (|uv|) : v \in L^{\Phi_2}(p_2 \cdot \mu), \|v\|_{\Phi_2,p_2} \leq 1 \right\}.
\]

Since \( \Phi_1 \) and \( \Phi_2 \) are equivalent norms and the Banach spaces \( L^{\Phi_1}(p_1 \cdot \mu) \) and \( L^{\Phi_2}(p_2 \cdot \mu) \) coincide, also the Banach spaces \( L^{\Phi_2}(p_1 \cdot \mu) \) and \( L^{\Phi_2}(p_2 \cdot \mu) \) coincide and there exists a constant \( c = c(p_1,p_2) > 0 \) such that

\[
c \|v\|_{\Phi_2,p_1} \leq \|v\|_{\Phi_2,p_2}
\]

for each \( v \in L^{\Phi_2}(p_1 \cdot \mu) = L^{\Phi_2}(p_2 \cdot \mu) \).
If \( \|v\|_{\Phi_2,p_2} \leq 1 \) then \( \|cv\|_{\Phi_2,p_1} \leq 1 \). We have

\[
N_{\Phi_3,p_2} \left( \frac{p_1}{p_2} \right) = \sup \left\{ E_{p_1} (|uv|) : v \in L^{\Phi_2} (p_2 \cdot \mu), \|v\|_{\Phi_2,p_2} \leq 1 \right\}
\]

\[
= \frac{1}{c} \sup \left\{ c E_{p_1} (|uv|) : v \in L^{\Phi_2} (p_2 \cdot \mu), \|cv\|_{\Phi_2,p_1} \leq 1 \right\}
\]

\[
\leq \frac{1}{c} \sup \left\{ E_{p_1} (|u(cv)|) : v \in L^{\Phi_2} (p_1 \cdot \mu), \|cv\|_{\Phi_2,p_1} \leq 1 \right\}
\]

\[
= \frac{1}{c} \sup \left\{ E_{p_1} (|uv'|) : v' \in L^{\Phi_2} (p_1 \cdot \mu), \|v'\|_{\Phi_2,p_1} \leq 1 \right\}
\]

\[
= \frac{1}{c} N_{\Phi_3,p_1} (u) < \infty.
\]

Since \( N_{\Phi_3,p_2} (u/p_2) \) is bounded, \( u/p_2 \) is an element of \( L^{\Phi_3} (p_2 \cdot \mu) \) and, since

\[
N_{\Phi_3,p_2} \left( \frac{p_1}{p_2} \right) \leq c^{-1} N_{\Phi_3,p_1} (u),
\]

the linear map \( U_{p_1/p_2} \) is continuous. Simmetrically one can see that the inverse

\[
L^{\Phi_3} (p_2 \cdot \mu) \ni w \mapsto U_{p_2/p_1} (w) = u \frac{p_2}{p_1} \in L^{\Phi_3} (p_1 \cdot \mu)
\]

is a continuous linear map.

**Corollary 34.** Let \( p_1 \) and \( p_2 \) be two positive densities in the same connected component \( E(p) \subset M_\mu \) for some density \( p \in M_\mu \). Then

i) \( \frac{p_1}{p_2} \in L^{\Phi_3} (p_2 \cdot \mu) \);

ii) the linear map \( P_{p_1/p_2}^m : \ast B_{p_1} \to \ast B_{p_2} \)

\[
\ast B_{p_1} \ni u \mapsto P_{p_1/p_2}^m (u) = u \frac{p_1}{p_2} \in \ast B_{p_2}
\]

is a toplinear isomorphism.

**Proof.** i) We have \( U_{p_1/p_2} (1) = \frac{p_1}{p_2} \in L^{\Phi_3} (p_2 \cdot \mu) \).

ii) \( P_{p_1/p_2}^m \) is the restriction of \( U_{p_1/p_2} \) to the \( x \log x \) class of \( L^{\Phi_3} (p_1 \cdot \mu) \). If \( u \in \ast B_{p_1} \) then \( E_{p_2} (u p_1/p_2) = E_{p_1} (u) = 0 \) and \( u p_1/p_2 \in \ast B_{p_2} \). 

**Proposition 35.** Let \( p_1 \) and \( p_2 \) be a pair of positive densities in the same connected component \( E(p) \subset M_\mu \) for some density \( p \in M_\mu \), then \( \ast U_{p_1} = \ast U_{p_2} \).
Proof. Let \( p_1,p_2 \in \mathcal{E}(p) \). First we note that \( p_1,p_2 \in \mathcal{U}_{p_1} \cap \mathcal{U}_{p_2} \). In fact, by Corollary 34 we have \( \frac{p_1}{p_2} \in L^{\Phi_3}(p_2 \cdot \mu) \) and we conclude that \( p_1 \in \mathcal{U}_{p_2} \). Similarly, we see that \( p_2 \in \mathcal{U}_{p_1} \).

Every \( q \in \mathcal{U}_{p_1} \) can be written as \( (u + 1) p_1 \) where \( u = \eta_{p_1}(q) \in \mathcal{B}_{p_1} \) and we have
\[
\frac{q}{p_2} = \frac{u p_1}{p_2} + \frac{p_1}{p_2} = \frac{p_1}{p_2} (u) + \frac{p_1}{p_2} \in L^{\Phi_3}(p_2 \cdot \mu).
\]
Hence \( q \in \mathcal{U}_{p_2} \) and \( \mathcal{U}_{p_1} \subseteq \mathcal{U}_{p_2} \). In the same way we prove the opposite inclusion. \( \square \)

For each pair \( p_1,p_2 \in \mathcal{E}(p) \) we can define the overlap map
\[
\eta_{p_2} \circ \eta^{-1}_{p_1} : \mathcal{B}_{p_1} \to \mathcal{B}_{p_2}, \quad u \mapsto u \frac{p_1}{p_2} + \frac{p_1}{p_2} - 1.
\]

The function \( \eta_{p_2} \circ \eta^{-1}_{p_1} \) is a \( C^\infty \)-affine map since it can be written as the sum of the continuous linear map \( p_{p_1,p_2}^\mu \) and the constant \( \frac{p_1}{p_2} - 1 \in \mathcal{B}_{p_2} \) so it a \( C^\infty \)-affine map.

**Definition 36.** Let \( p \in M_\mu \) be fixed. \( \mathcal{E}(p) \) is the subset of \( \mathcal{P} \) defined by
\[
\mathcal{E}(p) = \left\{ q \in \mathcal{P} : \frac{q}{p} \in L^{\Phi_3}(p \cdot \mu) \right\}
\]

\( \mathcal{E}(p) = \mathcal{U}_{p_\alpha} \) for each \( p_\alpha \in \mathcal{E}(p) \).

\( \mathcal{E}(p) \) has the structure of a manifold modeled on the Banach space \( \mathcal{B}_p \).

**Theorem 37.** Let \( p \in M_\mu \) be given. The collection of charts
\[
\{ (\mathcal{U}_{p_\alpha}, \eta_{p_\alpha}) : p_\alpha \in \mathcal{E}(p) \}
\]
is an affine \( C^\infty \)-atlas on \( \mathcal{E}(p) \).

**Proof.** The collection of sets \( \{ \mathcal{U}_{p_\alpha} : p_\alpha \in \mathcal{E}(p) \} \) covers \( \mathcal{E}(p) \). For each pair \( p_1,p_2 \in \mathcal{E}(p) \) the set \( \eta_{p_1}^{-1}(\mathcal{U}_{p_1} \cap \mathcal{U}_{p_2} = \mathcal{E}(p)) = \mathcal{B}_{p_1} \) is clearly open in \( \mathcal{B}_{p_1} \) and we have just observed that the transition mapping \( \eta_{p_2} \circ \eta^{-1}_{p_1} \) is an \( C^\infty \)-affine function. \( \square \)

We conclude with a local Pythagorean-type relation.

Let \( p \in M_\mu \) be given and let \( s_p : \mathcal{U}_p \to \mathcal{V}_p \) and \( \eta_p : \mathcal{U}_p \to \mathcal{B}_p \) be charts respectively in \( \mathcal{E}(p) \) and \( \mathcal{E}(p) \). Let \( q \in \mathcal{U}_p \), \( u = s_p(q) \) and \( 0 \leq r \in \mathcal{U}_p \) be given. Consider the duality
\[
\langle \eta_p(r), s_p(q) \rangle = \mathcal{E}_p[\eta_p(r) s_p(q)] = \mathcal{E}_p \left[ \left( \frac{r}{p} - 1 \right) u \right] = \mathcal{E}_r(u).
\]
As
\[ u = \log \left( \frac{q}{p} \right) - \mathbb{E}_p \left[ \log \left( \frac{q}{p} \right) \right] = \log \left( \frac{q}{p} \right) + K(p | q), \]
we have
\[
\mathbb{E}_r(u) = \mathbb{E}_r \left[ \log \left( \frac{q}{p} \right) \right] + K(p | q)
= \mathbb{E}_r \left[ \log \left( \frac{q}{r} \right) + \log \left( \frac{r}{p} \right) \right] + K(p | q)
= -K(r | q) + K(r | p) + K(p | q).
\]
In particular, \( \langle \eta_p(r), s_p(q) \rangle = 0 \) implies the relation
\[ K(r | q) = K(r | p) + K(p | q). \]
Chapter 4

Christoffel symbols and connections

4.1 Connection in a vector bundle

Let $X$ be a manifold and $\pi : E \to X$ be a vector bundle respectively modeled on the Banach spaces $F$ and $E$. Let us denote $E_x$ the fiber $\pi^{-1}(x)$ over $x \in X$. Let $\pi_E : TE \to E$ be the tangent bundle of $E$. We assume that manifolds are objects of the category $\text{Man}^n$ of all $C^n$-manifolds with $n \geq 1$ so a mapping of class $C^n$ between them will be denoted morphism. Note that if $X$ is a $C^n$-manifold, then the tangent bundle $\pi : TX \to X$ lies in $\text{Man}^{n-1}$.

We refer to Klingenberg (1978) and Lang (1995).

$X$ has an atlas such that over each chart $(U, \varphi)$ there is a trivialization which determines a local representation of the bundle, i.e. a chart $(\pi^{-1}(U), \Phi)$ of $E$ such that the following diagram

\[
\begin{array}{ccc}
\pi^{-1}(U) & \xrightarrow{\Phi} & \varphi(U) \times E \\
\downarrow \pi & & \downarrow \text{pr}_1 \\
U & \xrightarrow{\varphi} & \varphi(U)
\end{array}
\]

is commutative.

Example 38 (The tangent bundle over the maximal exponential model).

By Definition 28, for every density $f \in \mathcal{M}_\mu$ the maximal exponential model at $f$ is the family of densities

\[
\mathcal{E}(f) := \left\{ e^{u - K_f(u)} f : u \in \text{dom} \, K_f \right\} \subset \mathcal{M}_\mu.
\]
Given the $C^\infty$-atlas $\{(U_p, s_p) : p \in E(f)\}$ of $E(f)$, the tangent bundle

$$TE(f) = \bigcup_{q \in E(f)} B_q$$

has a trivializing atlas $\{(TU_p, Ts_p) : p \in E(f)\}$ such that the diagram

$$
\begin{array}{ccc}
TU_p & \xrightarrow{Ts_p} & V_p \times B_p \\
\downarrow{} & & \downarrow{} \\
U_p & \xrightarrow{s_p} & V_p \\
\end{array}
$$

is commutative.

For all $w \in B_q \subset TU_p = \bigcup_{g \in U_p} B_g$, we have:

$$Ts_p(w) = (s_p(q), w - E_p(w)).$$

**Example 39 (The pre-tangent bundle over $E(f)$).** Given the $C^\infty$-atlas of $E(f)$, $\{(U_p, s_p)\}_{p \in E(f)}$, the pre-tangent bundle over $E(f)$ is defined by

$$*(T^E(f)) := \bigcup_{q \in E(f)} *B_q.$$

It has a trivializing atlas $\{(*TU_p, *Ts_p) : p \in E(f)\}$ such that the diagram

$$
\begin{array}{ccc}
*TU_p & \xrightarrow{*Ts_p} & V_p \times *B_p \\
\downarrow{} & & \downarrow{} \\
*U_p & \xrightarrow{*s_p} & V_p \\
\end{array}
$$

is commutative.

For all $w \in *B_q \subset *TU_p = \bigcup_{g \in U_p} *B_g$, we have:

$$*Ts_p(w) = \left( s_p(q), P^m_{qp}(w) = w \frac{q}{p} \right)$$

where $P^m_{qp} : *B_q \to *B_p$ is the toplinear isomorphism defined in Corollary 34.

**Definition 40.** A connection in the vector bundle $\pi : E \to X$ is a vector bundle morphism $H : \pi_E \to \pi$ over $\pi$ (i.e. $H$ is a morphism making the diagram

$$
\begin{array}{ccc}
TE & \xrightarrow{H} & E \\
\downarrow{} & & \downarrow{} \\
E & \xrightarrow{\pi} & X \\
\end{array}
$$

27
commutative) such that for each local representation induced by a local chart \((U,\varphi)\) of \(X\) with \(\varphi(U) = V \subset F\) the local representative \(H_V := \Phi \circ H \circ (T\Phi)^{-1}\)

\[
\begin{array}{ccc}
T\pi^{-1}(U) & \xrightarrow{H} & \pi^{-1}(U) \\
\downarrow T\Phi & & \downarrow \Phi \\
V \times E \times F \times E & \xrightarrow{H_V} & V \times E
\end{array}
\]

is given by:

\[
H_V : (x,\xi,y,\eta) \mapsto (x,\eta + \Gamma_V(x)(y,\xi))
\] (4.1)

where \(\Gamma_V : V \to \mathcal{L}^2(F,E;E)\) is a morphism. Mappings \(\Gamma_V\) are the Christoffel symbols of the connection. A connection in the tangent bundle \(\pi_X : TX \to X\) is called an affine connection.

If \(X\) and \(E\) are objects of \(\text{Man}^n\) with \(n \geq 1\), we shall assume that \(H\) and \(\Gamma_V\) are \(C^{n-1}\)-morphisms. In the case of an affine connection, \(E = TX\) is a \(C^{n-1}\)-manifold and we shall assume that \(n \geq 2\) and that \(H\) and \(\Gamma_V\) are \(C^{n-2}\)-morphisms.

Let us compute the transformation rule for Christoffel symbols under a change of coordinates. If \((U_1,\varphi_1)\) and \((U_2,\varphi_2)\) is a pair of charts of \(X\) such that \(U := U_1 = U_2\) and if \(V_1 = \varphi_1(U)\) and \(V_2 = \varphi_2(U)\), then \(f := \varphi_2 \circ \varphi_1^{-1} : V_1 \to V_2\) is an isomorphism. Rising to the upper level, there are two charts \((\pi^{-1}(U),\Phi_1)\) and \((\pi^{-1}(U),\Phi_2)\) of \(E\) and there is the isomorphism \(F = \Phi_2 \circ \Phi_1^{-1} = (f,L)\) such that

\[
F : V_1 \times E \ni (x,\xi) \mapsto \left(f(x),L(x) \cdot \xi\right) \in V_2 \times E
\]

where \(L : V_1 \to \text{Aut}(E)\) is a morphism (\(\text{Aut}(E) \subset \mathcal{L}(E;E)\) is the open subset of all the toplinear isomorphisms). Note that \(DL : V_1 \to \mathcal{L}^2(F,E;E)\). When we lift \(F\) to the tangent bundles we obtain \(TF : V_1 \times E \times F \times E \to V_2 \times E \times F \times E\) such that

\[
TF : (x,\xi,y,\eta) \mapsto \left(f(x),L(x) \cdot \xi,Df(x) \cdot y,DL(x) (y,\xi) + L(x) \cdot \eta\right).
\]

28
4.1– Connection in a vector bundle

We may summarize with the following diagram.

\[
\begin{array}{c}
V_1 \times E \times F \times E \xrightarrow{TF} V_2 \times E \times F \times E \\
\downarrow \quad \downarrow \\
V_1 \times E \xrightarrow{\Phi_1^{-1}} V_1 \\
\downarrow \quad \downarrow \\
V_2 \times E \xrightarrow{\Phi_2} V_2 \\
\end{array}
\]

(4.2)

The local representations \(H_{V_1}\) and \(H_{V_2}\) of \(H\) are related by the relation: \(H_{V_1} = F^{-1} \circ H_{V_2} \circ TF\)

\[
\begin{array}{c}
V_1 \times E \times F \times E \xrightarrow{H_{V_1}} V_1 \times E \\
\downarrow \quad \downarrow \\
V_2 \times E \times F \times E \xrightarrow{H_{V_2}} V_2 \times E \\
\end{array}
\]

(4.3)

Hence under a change of variables the Christoffel symbols satisfy the following identity:

\[
\Gamma_{V_2} (f (x)) (Df (x) \cdot y, L (x) \cdot \xi) = -DL (x) (y, \xi) + L (x) \cdot (\Gamma_{V_1} (x) (y, \xi))
\]

(4.3)

for each \((x, \xi, y, \eta) \in V_1 \times E \times F \times E\).

Let \(\mathcal{X}_E (X)\) denote the \(C^n (X)\)-module of all the sections of class \(C^n\) of the bundle \(\pi\), that is the morphism \(\xi : X \to E\) such that \(\pi \xi = \text{id}_X\). Let \(\mathcal{X} (X)\) denote the set of all \(C^n\)-vector fields of \(X\).
Definition 41. Given a connection \( H \) in a vector bundle \( \pi : E \to X \), the covariant derivative of a section \( \xi \in \mathfrak{X}_E(X) \) is the morphism \( \nabla \xi : TX \to E \) defined by \( \nabla \xi := H \circ T\xi \).

In a local chart \((U, \varphi)\) of \( X \), a section \( \xi \in \mathfrak{X}_E(X) \) is represented by a morphism of \( V = \varphi(U) \) into \( V \times E \) which has two component \((\text{id}_V, \xi_V)\) See the following diagram

\[
\begin{array}{ccc}
U & \xrightarrow{\xi} & \pi^{-1}(U) \\
\varphi \downarrow & & \downarrow \Phi \\
V & \xrightarrow{(\text{id}_V, \xi_V)} & V \times E
\end{array}
\]

The morphism \( \xi_V : V \to E \) is called the principal part of the local representation of the section. Sometimes, to simplify the notation we agree that \( \xi = \xi_V \).

In coordinates, \( T\xi : TU \to T\pi^{-1}(U) \) is given by the local representation:

\[
V \times F \ni (x,y) \mapsto (x, \xi_V(x), y, D\xi_V(x) \cdot y) \in V \times E \times F \times E
\]

and, by Eq. (4.1), for each \((x,y) \in \varphi(U) \times F\) the covariant derivative of \( \xi \) is given by the following formula:

\[
\nabla \xi(x,y) = \left( x, D\xi_V(x) \cdot y + \Gamma_V(x)(y, \xi_V(x)) \right).
\]

(4.4)

Definition 42. Given a vector bundle \( \pi : E \to X \), a covariant derivative \( \nabla \) on the vector bundle is a \( \mathbb{R} \)-linear map

\[
\nabla : \mathfrak{X}(X) \times \mathfrak{X}_E(X) \to \mathfrak{X}_E(X)
\]

\[
(v, \xi) \mapsto \nabla_v \xi
\]

satisfying the following conditions:

1. it is \( C^n(X) \)-linear in the first variable, that is, for each \( f \in C^n(X) \),

\[
\nabla_{fv} \xi = f \nabla_v \xi
\]

2. it is a derivation on \((C^n(X), \mathfrak{X}_E(X))\) in the second variable, that is, for each \((f, \xi) \in C^n(X) \times \mathfrak{X}_E(X)\)

\[
\nabla_{v} f \xi = (\nabla_v f) \xi + f \nabla_v \xi
\]

where \( \nabla_v f \) is the Lie derivative of \( f \) along \( v \), \( \nabla_v f := v(f) \).

From a connection \( H \) defined by gluing together a set of Christoffel symbols as in Definition 40 we always obtain a covariant derivative according to Definition 42.
**Definition 43.** Given a connection $H$ in a vector bundle $\pi : E \to X$, a covariant derivative $\nabla : \mathfrak{X}(X) \times \mathfrak{X}_E(X) \to \mathfrak{X}_E(X)$ is defined by setting: $\nabla_v \xi := \nabla \xi \circ v$ where $\nabla \xi$ is the covariant derivative of the section $\xi$ as in Definition 41.

Definition 43 is consistent. In fact, by Eq. (4.4) we have the local formula

$$(\nabla_v \xi)_V (x) = D\xi_V (x) \cdot v_V (x) + \Gamma_V (x) \left( v_V (x), \xi_V (x) \right)$$

(4.5)

and, with an easy calculation, we can check that the two conditions of Def. 42 are satisfied.

In finite dimension, the association of a covariant derivative to a connection is bijective, in our more general case this is not always true (see Lang (1995, VIII-2)).

**Definition 44.** A bundle-connection pair $(E, \nabla)$ is a pair composed of a vector bundle $\pi : E \to X$ and a covariant derivative $\nabla$ on $E$.

**Connection on manifolds in $\mathbb{R}^n$**

For helping geometric understanding of the Christoffel symbols it is useful to look at the case of a manifold embedded in $\mathbb{R}^n$. We refer to Boothby (1975, VII 2).

Let $M \subset \mathbb{R}^n$ be a submanifold of dim $m$ and $\xi \in \mathfrak{X}(M)$ be a vector field to $M$. If $\sigma : (-\varepsilon, \varepsilon) \to M$ is a $C^1$-curve, then usual differentiation in $\mathbb{R}^n$ induces an “intrinsic” (on $M$) derivative of $\xi$ along it (Boothby, 1975, Definition 2.2).

**Definition 45.** The covariant derivative $\frac{D\xi}{dt}$ of the tangent vector field $\xi$ along $\sigma$ is the orthogonal projection on $TqM \subset \mathbb{R}^n$ where $q = \varphi^{-1}(u)$ and, since $\xi(q) \in TqM$, there exist $m$ regular functions $\xi_i : V \to \mathbb{R}$ such that the local representation $\tilde{\xi} = \xi \circ \varphi^{-1}$ can be written as a linear combination:

$$\tilde{\xi}(u) = \sum_{i=1}^{m} \xi_i(u) F_i(u).$$

(4.6)
Note that $\xi_1 (u), \ldots, \xi_m (u)$ is the principal part of the local representation of $\xi$:

$$
\begin{array}{c}
U \\
\pi \\
V
\end{array} \xrightarrow{\xi} \begin{array}{c}
TU \\
T\varphi \\
V \times \mathbb{R}^m
\end{array}
$$

Put $\xi (t) = \xi (\sigma (t))$ and $u (t) = \varphi \circ \sigma (t)$, then

$$
\xi (t) = \sum_{i=1}^m \xi_i (u (t)) F_i (u (t))
$$

and

$$
\dot{\xi} = \sum_{i=1}^m D\xi_i (u) \cdot \dot{u} F_i (u) + D^2 \varphi^{-1} (u) (\xi (u), \dot{u}) \tag{4.7}
$$

where $D\xi_i (u) \cdot \dot{u} = \sum_{j=1}^m \frac{\partial \xi_i}{\partial u_j} \bigg|_u \dot{u}_j$ and

$$
D^2 \varphi^{-1} (u) (\xi, \dot{u}) = \sum_{\alpha=1}^n \sum_{i,j=1}^m \frac{\partial^2 g_{\alpha}}{\partial u_i \partial u_j} \bigg|_u \xi_i (u) \dot{u}_j \frac{\partial}{\partial x_\alpha}.
$$

Applying projection $\Pi_q : \mathbb{R}^n \to T_q M$ with $q = \varphi^{-1} (u)$, we find

$$
\frac{D\xi}{dt} = \sum_{i=1}^m D\xi_i (u) \cdot \dot{u} F_i (u) + \sum_{\alpha=1}^n \sum_{i,j=1}^m \frac{\partial^2 g_{\alpha}}{\partial u_i \partial u_j} \bigg|_u \xi_i (u) \dot{u}_j \Pi_q \frac{\partial}{\partial x_\alpha}. \tag{4.8}
$$

There exist $mn$ regular functions $a^k_\alpha : V \to \mathbb{R}$ such that:

$$
\Pi_q \frac{\partial}{\partial x_\alpha} = \sum_{k=1}^m a^k_\alpha (u) F_k (u) \quad \alpha = 1, \ldots, n.
$$

We may now introduce the classical Christoffel symbols:

$$
\Gamma_{ij}^k (u) := \sum_{\alpha=1}^n \frac{\partial^2 g_{\alpha}}{\partial u_i \partial u_j} \bigg|_u a^k_\alpha (u) \quad 1 \leq i, j, k \leq m
$$

and Eq. (4.8) become:

$$
\frac{D\xi}{dt} = \sum_{k=1}^m \left[ D\xi_k (u) \cdot \dot{u} + \sum_{i,j=1}^m \Gamma_{ij}^k (u) \xi_i (u) \dot{u}_j \right] F_k (u). \tag{4.9}
$$
4.1– Connection in a vector bundle

In particular, if \( \xi = F_i \) and \( u(t) \) is defined by \( u_l = \text{const.} \) for \( l \neq j \) and \( u_j = t \), then

\[
\frac{DF_i}{dt} = \sum_{k=1}^{m} \Gamma_{ij}^k F_k.
\]

Hence \( \Gamma_{ij}^k \) is the \( k \)th component relative to the coordinate frame of the projection of the rate of change of \( F_i \) along a coordinate curve.

The covariant derivative \( \frac{d\xi}{dt} \) is the intrinsic derivative of \( \xi \), that is as seen from the viewpoint of \( M \). Bilinear mappings \( \Gamma(u) \in \mathcal{L}^2(T_q M; T_q M) \)

\[
\Gamma(u) \cdot (v,w) = \sum_{i,j,k=1}^{m} \Gamma_{ij}^k (u) v_i w_j F_k(u)
\]

are understood to be the term in Eq. (4.9) which takes into account the motion along the curve of the tangent spaces.

A covariant derivative along a curve define a covariant derivative \( \nabla : \mathfrak{X}(M) \times \mathfrak{X}(M) \rightarrow \mathfrak{X}(M) \) as we may assume

\[
(\nabla_{\eta} \xi)(q) = (\nabla_{\eta_0} \xi)(q) := \left. \frac{D\xi}{dt} \right|_{t_0}
\]

where in the right-hand side the covariant derivative is along an arbitrary curve such that \( \sigma(t_0) = q \) and \( \dot{\sigma}(t_0) = \eta(t_0) = \eta_0 \) (Boothby, 1975, Theorem 2.11). Finally, in finite dimension, formula (4.5) become the classical one

\[
(\nabla_{\eta} \xi)_V = \sum_{k=1}^{m} \left[ \eta(\xi_k) + \sum_{i,j=1}^{m} \Gamma_{ij}^k \xi_i \eta_j \right] F_k.
\]

**Parallel transport**

Let \( \pi : E \rightarrow X \) be a vector bundle and \( \alpha : J \rightarrow X \) be a \( \mathcal{C}^1 \)-curve. A lift \( \gamma \) of \( \alpha \) to \( E \) is a \( \mathcal{C}^1 \)-curve \( \gamma : J \rightarrow E \) s.t. \( \pi \gamma = \alpha \). The set Lift\(_E\) (\( \alpha \)) of all lifts of \( \alpha \) to \( E \) is a \( \mathbb{R} \)-vector space. Following proposition is stated by Lang (1995, Theorem VIII-3.1) in the particular case of an affine connection determined by a spray on \( X \).

**Proposition 46.** Given a covariant derivative \( \nabla : \mathfrak{X}(X) \times \mathfrak{X}_E(X) \rightarrow \mathfrak{X}_E(X) \) associated to a connection, there exists a unique linear map \( \nabla_\alpha : \text{Lift}_E(\alpha) \rightarrow \text{Lift}_E(\alpha) \) which in a chart \( (U,\varphi) \) has local representation:

\[
(\nabla_\alpha \gamma)_V(t) = \dot{\gamma}_V(t) + \Gamma_V(\alpha_V(t)) (\dot{\alpha}_V(t),\gamma_V(t)).
\]
Proof. Lang’s proof can be adapted using the transformation rule (4.3) for Christoffel symbols to show that Eq. (4.10) transforms in the proper way under a change of charts.

To be more precise: the local representation \((\nabla_\dot{\alpha} \gamma)_V\) is a mapping \(J \to V \times E\) and in Eq. (4.10) we are confounding it with the second component, its principal part.

Likewise the case of a submanifold of \(\mathbb{R}^n\), by Proposition 46 one may introduce an intrinsic derivative of sections along curves. If \(\xi \in \mathfrak{X}(X)\) is a section and \(\gamma(t) := \xi(\alpha(t))\) for \(t \in J\), then \(\frac{d\xi}{dt} := \nabla_\alpha \gamma\). Note that if \(\eta \in \mathfrak{X}(X)\) is a vector field such that \(\dot{\alpha}(t_0) = \eta(\alpha(t_0))\) for some \(t_0 \in J\), then \(\nabla_\alpha \gamma(t_0) = \nabla_\eta \xi(\alpha(t_0))\).

In the following we assume that \(\pi : E \to X\) is a vector bundle in the category \(\text{Man}^n\) with \(n > 1\).

**Definition 47.** Let \(\alpha : J \to X\) be a \(C^2\)-curve. A lift \(\gamma \in \text{Lift}_E(\alpha)\) is \(\alpha\)-parallel if \(\nabla_\dot{\alpha} \gamma = 0\).

By Eq. (4.10), \(\alpha\)-parallelism is locally equivalent to the first order linear differential equation for \(\gamma_V\):
\[
\dot{\gamma}_V = -\Gamma_V(\alpha_V)(\dot{\alpha}_V, \gamma_V).
\]

**Definition 48.** Given an affine connection in the tangent bundle of \(X\), a \(C^2\)-curve \(\alpha\) is said geodesic if \(\dot{\alpha}\) is \(\alpha\)-parallel, that is \(\nabla_\alpha \dot{\alpha} = 0\).

Differential Equation (4.11) can be rewrite for geodesics in the following manner:
\[
\dot{\alpha}_V = -\Gamma_V(\alpha_V)(\dot{\alpha}_V, \dot{\alpha}_V).
\]

**Theorem 49.** Let \(\alpha : J \to X\) be a \(C^2\)-curve. Let \(t_0 \in J\). Given \(v \in E_{\alpha(t_0)}\), there exists a unique lift \(\gamma(t,v) : J \to E\) which is \(\alpha\)-parallel and such that \(\gamma(t_0,v) = v\).

**Proof.** It follows from the Theorem of existence and uniqueness of solutions of linear differential equations. See, for exemple, Lang (1995, Proposition IV-1.9)

**Definition 50.** Let
\[
\mathcal{P}(X) := \{\alpha : I = [0,1] \to X, \alpha \text{ is piecewise } C^2\}
\]
be the space of all piecewise \(C^2\) curves of \(X\). A parallel transport on a vector bundle \(\pi : E \to X\) is a function
\[
P : \mathcal{P}(X) \to \bigcup_{p,q \in X} \text{Iso}(E_p, E_q)
\]
such that
4.2 Splitting

1. $P_\alpha \in \text{Iso} \left( E_{\alpha(0)}, E_{\alpha(1)} \right)$;
2. $P$ is parametrization independent;
3. $P_{\alpha^{-1}} = P_{\alpha}^{-1}$;
4. $P_\alpha P_\beta = P_{\beta \alpha}$.

A connection in a vector bundle $E$ define a parallel transport on $E$. In fact, let $\alpha : I \to X$ be a $C^2$-curve. By Theorem 49, we can define a mapping

$$P_\alpha : E_{\alpha(0)} \to E_{\alpha(1)}$$

with $P_\alpha (v) := \gamma(1,v)$ where $\gamma(\cdot, v)$ is the unique lift in $E$ which is $\alpha$-parallel and $\gamma(0, v) = v$. $P_\alpha$ is a linear isomorphism (Lang, 1995, Theorem VIII-3.4) and $P$ is a parallel transport.

**Theorem 51.** Given a vector bundle $\pi : E \to X$ there exists a bijection between covariant derivatives and parallel transports on $E$.

**Proof.** See Gibilisco (1997, Theorem 6.1).

**Definition 52.** A connection is globally flat if for any loop $\sigma$ at any fixed point $p \in X$, $P_\sigma = \text{id}_{E_p}$.

For a globally flat connection, the parallel transport $P_\alpha$ depends only on the end points of the curve $\alpha$, so for each $p, q \in X$ we can put $P_{p,q} := P_\alpha \in \text{Iso} (E_p, E_q)$ where $\alpha$ is an arbitrary path from $p$ to $q$.

### 4.2 Splitting

An Orlicz space’s splitting plays an important role in the definition of the connections on the Statistical Manifold $\mathcal{M}_\mu$ so, before introducing them, we shall show how at each density $p \in \mathcal{M}_\mu$ the tangent space splits.

**Definition 53.** The closed subspace $F$ of a Banach space $E$ is said to be split if there exists a closed subspace $G \subset E$ such that $E = F \oplus G$.

Since a continuous linear isomorphism of Banach spaces is a homeomorphism, $F$ splits in $E$ iff $E$ is the algebraic direct sum of the closed subspaces $F$ and $G$.

**Definition 54.** A continuous linear function $\Pi \in \mathcal{L}(E; E)$ is a projection if $\Pi^2 = \Pi$.

Splittings and projections are related.
**Theorem 55.** Let \( E \) be a Banach space and \( F \) a closed subspace of \( E \). Then \( F \) splits iff there exists a projection \( \Pi \in \mathcal{L}(E;E) \) and \( F = \{ x \in E : \Pi x = x \} = \text{im} \Pi \) and \( E = F \oplus \ker \Pi \).

**Proof.** See Abraham et al. (1988, Corollary 2.2.18). \( \square \)

**Proposition 56.** Let \( E \) be a Banach space and \( \lambda \in E^* \). For each \( v \not\in \ker \lambda \) there exists a splitting

\[
E = \ker \lambda \oplus \langle v \rangle
\]

where \( \langle v \rangle \) is the 1-dimensional linear subspace generated by \( v \). Every \( w \in E \) can be decomposed as a sum

\[
w = (w - \lambda w \frac{v}{\lambda v}) + \lambda w \frac{v}{\lambda v}.
\]

(4.13)

**Proof.** Fix \( v \not\in \ker \lambda \) and define the mapping \( i_v : \mathbb{R} \to E \) by

\[
\mathbb{R} \ni \alpha \mapsto \alpha \frac{v}{\lambda v} \in \langle v \rangle \subset E.
\]

It is an injective linear functional with range the closed subspace generated by \( v \). Since \( i_v \) is continuous, \( \Pi := \text{id} - i_v \circ \lambda \in \mathcal{L}(E;E) \). Linear functional \( \Pi \) satisfies \( \Pi^2 = \Pi \) and it is a projection. Since:

1. \( \ker \Pi = \langle v \rangle \), in fact:
   \[
w \in \ker \Pi \iff w = \lambda w \frac{v}{\lambda v} \in \langle v \rangle ;
   \]

2. \( \text{im} \Pi = \{ w \in E : \Pi w = w \} = \ker \lambda \), in fact:
   \[
w \in \text{im} \Pi \iff w = \Pi w = w - \lambda w \frac{v}{\lambda v} \iff \lambda w = 0;
   \]

by Theorem 55, \( E = \ker \lambda \oplus \langle v \rangle \). \( \square \)

**Definition 57.** A Young function \( \Phi : \mathbb{R} \to [0,\infty) \) is an N-function if it is a continuous Young function such that \( \Phi(x) = 0 \) iff \( x = 0 \), \( \lim_{x \to 0} \Phi(x)/x = 0 \) and \( \lim_{x \to \infty} \Phi(x)/x = \infty \).

\( \Phi_1, \Phi_2, \Phi_3 \) and \( |x|^a \) for \( a > 1 \) are N-functions.

**Proposition 58.** Let \((X,\mathcal{X},p\cdot \mu)\) be a probability space, \( \Phi : \mathbb{R} \to [0,\infty) \) be an N-function, then expectation \( E_p : L^\Phi(p\cdot \mu) \to \mathbb{R} \) is a continuous linear map. Subspace

\[
L_0^\Phi(p\cdot \mu) = \{ u \in L^\Phi(p\cdot \mu) : E_p(u) = 0 \}
\]

is closed.
4.3– Exponential Connection

Proof. Expectation $\mathbb{E}_p$ is linear. We show it is bounded. Since $\lim_{x \to \infty} \frac{\Phi (x)}{x} = \infty$, there exists $N > 0$ such that for $|x| > N$, $\Phi (x) > |x|$. Let $u \in L^\Phi (p \cdot \mu) \setminus \{0\}$,

$$\mathbb{E}_p \left( \frac{u}{\| u \|_\Phi} \right) = \int_{\| u \|_\Phi \leq N} \frac{u}{\| u \|_\Phi} p d\mu + \int_{\| u \|_\Phi > N} \frac{u}{\| u \|_\Phi} p d\mu \leq N + \mathbb{E}_p \left( \Phi \left( \frac{u}{\| u \|_\Phi} \right) \right) \leq N + 1,$$

implies $\mathbb{E}_p (u) \leq C \| u \|_\Phi$. $L^\Phi_0 (p \cdot \mu)$ is the kernel of $\mathbb{E}_p$. \hfill \square

In the case of $\Phi_1 = \cosh -1$, continuity also comes from the identity $\mathbb{E}_p = DM_p (0)$.

Proposition 59. Let $(X, \mathcal{X}, p \cdot \mu)$ be a probability space, $\Phi : \mathbb{R} \to [0, \infty)$ be an $N$-function, then the subspace $L^\Phi_0 (p \cdot \mu)$ of all the random variables of $L^\Phi (p \cdot \mu)$ having zero expectation splits in $L^\Phi (p \cdot \mu)$:

$$L^\Phi (p \cdot \mu) = L^\Phi_0 (p \cdot \mu) \oplus \mathbb{R}.$$

Projection $\Pi$ on $L^\Phi_0 (p \cdot \mu)$ is

$$\Pi := \text{id} - i \circ \mathbb{E}_p,$$

where $i : \mathbb{R} \hookrightarrow L^\Phi (p \cdot \mu)$ is the inclusion mapping. Every $f \in L^\Phi (p \cdot \mu)$ is decomposed as a sum

$$f = (f - \mathbb{E}_p (f)) + \mathbb{E}_p (f).$$

Proof. Note that the set of all constant random variables $\langle 1 \rangle$ is closed in the Orlicz space $L^\Phi (p \cdot \mu)$ and it is the range of the inclusion mapping $i : \mathbb{R} \hookrightarrow L^\Phi (p \cdot \mu)$.

Putting $v = 1$ and $\lambda = \mathbb{E}_p$, everything comes from Proposition 56. \hfill \square

Corollary 60. The tangent space $T_p \mathcal{M}_\mu = B_p$ at each density $p \in \mathcal{M}_\mu$ splits in $L^{\Phi_1} (p \cdot \mu)$ as $L^{\Phi_1} (p \cdot \mu) = B_p \oplus \mathbb{R}$.

4.3 Exponential Connection

We shall define an affine connection in the tangent bundle of the maximal exponential model $E$ at an arbitrary point of the exponential statistical manifold $\mathcal{M}_\mu$.

Definition 61. The exponential connection in the tangent bundle $TE$ of a maximal exponential model $E \subseteq \mathcal{M}_\mu$ is the affine connection such that, for each chart $(U_p, s_p)$ and for each $u \in U_p$, the Christoffel symbols $\Gamma^k_p (u) \in L^2 (B_p; B_p)$ are the null operator.
The exponential connection is well defined. Let \((U_{p_1}, s_{p_1})\) and \((U_{p_2}, s_{p_2})\) be a pair of charts such that \(U_{p_1} \cap U_{p_2} \neq \emptyset\). We put \(f := s_{p_2} \circ e_{p_1}\) and \(F := Ts_{p_2} \circ T e_{p_1} = (f, L = Df)\) like in the diagram (4.2). For each \(u \in V_{p_1}\) and \(w \in B_{p_1}\)

\[
L(u) \cdot w = Df(u) \cdot w = w - E_{p_2}(w).
\]

Since \(DL \equiv 0\), for each \(u \in V_{p_1}\) and \(w_1, w_2 \in B_{p_1}\), the change of variables formula (4.3) for the Christoffel symbols simplifies to

\[
\Gamma^e_{p_2}(f(u)) \left( Df(u) \cdot w_1, Df(u) \cdot w_2 \right) = Df(u) \cdot \Gamma^e_{p_1}(u)(w_1,w_2) = 0
\]

and the Christoffel symbols \(\Gamma^e_{p_1}\) and \(\Gamma^e_{p_2}\) glue together in \(U_{p_1} \cap U_{p_2}\).

Let \(\xi, \eta \in X(\mathcal{E})\) be two vector fields and \(\xi_p, \eta_p\) be, respectively, the principal part of their local representation \(Ts_p \circ \xi \circ e_p\) and \(Ts_p \circ \eta \circ e_p\) with respect to a chart \((U_p, s_p)\):

\[

\begin{align*}
\mathcal{V}_p & \xrightarrow{\xi} T U_p \xrightarrow{T s_p} \mathcal{V}_p \times B_p \\
U_p & \xrightarrow{s_p} \mathcal{V}_p
\end{align*}
\]

By Eq. (4.4) and since \(\Gamma^e_p(u) \equiv 0\), the covariant derivative \(\nabla^e \xi : T \mathcal{E} \to T \mathcal{E}\) of the section \(\xi\) has the following local representation: for each \((u,w) \in \mathcal{V}_p \times B_p\)

\[
(\nabla^e \xi)_p(u,w) = (u, D \xi_p(u) \cdot w).
\]

Likewise, by Eq. (4.5), the covariant derivative \(\nabla^e_p \eta \in X(\mathcal{E})\) has the following local representation:

\[
\mathcal{V}_p \ni u \mapsto (\nabla^e_p \eta)_p = (u, D \xi_p(u) \cdot \eta_p(u)) \in \mathcal{V}_p \times B_p.
\]

**Proposition 62.** The exponential connection pair \((T \mathcal{E}, \nabla^e)\) is globally flat and the associated parallel transport \(P^e\) is:

\[
P^e_{p,q} : B_p \ni w \mapsto w - E_q(w) \in B_q.
\]

**Proof.** Let \(\alpha : I \to \mathcal{E}\) be a curve of class \(C^2\) with \(\alpha(0) = p\) and \(\alpha(1) = q\). We may assume that \(p,q \in U_p\), the domain of the chart \((U_p, s_p)\). For \(w \in B_p\), let \(\gamma : I \to T \mathcal{E}\) the lift of \(\alpha\) defined by

\[
\gamma(t) := w - E_{\alpha(t)}(w) \in B_{\alpha(t)}.
\]

In the chart \((U_p, s_p)\), the principal part of the local representation of \(\gamma\) is the constant curve \(\gamma_p(t) = w\). Derivative \(\dot{\gamma}_p(t) = 0\), so \(\gamma\) satisfies Eq. (4.11). The curve \(\gamma\) is \(\alpha\)-parallel and the parallel transport \(P_{\alpha}(w) = w - E_q(w)\) is independent of \(\alpha\). □
The curve $\alpha : I \to \mathcal{U}_p$ given by $\alpha (t) = e^{tw - K_p(tw)p}$ is the geodesic starting in $\alpha (0) = p$ with velocity vector $\dot{\alpha} (0) = w \in B_p$. In fact $\dot{\alpha} (t) = w - \mathcal{E}_\alpha (t) (w)$ and $\nabla \dot{\alpha} = 0$.

$\nabla^e$ is precisely the exponential connection introduced by Gibilisco and Pistone (1998, Definition 23) starting from the associated parallel transport. Since for each $q_1, q_2 \in \mathcal{E}$, $L^{\Phi_1} (q_1 \cdot \mu) = L^{\Phi_1} (q_2 \cdot \mu)$ we may identify the tangent bundle $T \mathcal{E}$ with a subset of the Banach space $L^{\Phi_1} := L^{\Phi_1} (q \cdot \mu)$ for some $q \in \mathcal{E}$. Exploiting this identification, one can compute exponential covariant derivative like in the following proposition (Gibilisco and Pistone, 1998, Proposition 25).

**Proposition 63.** Let $(T \mathcal{E}, \nabla^e)$ be the exponential connection pair, $\xi, \eta \in \mathfrak{X} (\mathcal{E})$, $q \in \mathcal{E}$ and $v \in B_q$ such that $\eta (q) = v$, then

\[
(\nabla^e_\eta \xi) (q) = d_v \xi (q) - \mathcal{E}_q (d_v \xi (q))
\]

where $d_v \xi (q) := \lim_{h \to 0} \frac{1}{h} [\xi (\sigma (h)) - \xi (\sigma (0))]$ with $\sigma : (-\varepsilon, \varepsilon) \to \mathcal{E}$ a $C^1$-curve such that $\sigma (0) = q$ and $\dot{\sigma} (0) = v$.

**Remark about Proposition 63.** Let $\xi \in \mathfrak{X} (\mathcal{E})$. We consider $\xi : \mathcal{E} \to T \mathcal{E} \subset L^{\Phi_1}$. Let a chart $(\mathcal{U}_p, s_p)$ be given, we put $\tilde{\xi}_p := \xi \circ e_p : V_p \to L^{\Phi_1}$ and $\tilde{\sigma} := s_p \circ \sigma$.

If $\sigma (0) = q = e_p (u)$, $\tilde{\sigma} (0) = v \in B_q$ and $w = v - \mathcal{E}_p (v)$, then we have

\[
d_v \xi (q) = \lim_{h \to 0} \frac{\tilde{\xi}_p (\tilde{\sigma} (h)) - \tilde{\xi}_p (\tilde{\sigma} (0))}{h} = D \tilde{\xi}_p (u) \cdot w.
\]

Since

\[\Pi_q : L^{\Phi_1} \ni f \mapsto f - \mathcal{E}_q (f) \in B_q\]

is the projection on $B_q$ in the splitting $L^{\Phi_1} = B_q \oplus \mathbb{R}$,

\[
(\nabla^e_\eta \xi) (q) = D \tilde{\xi}_p (u) \cdot w - \mathcal{E}_q \left( D \tilde{\xi}_p (u) \cdot w \right) = \Pi_q D \tilde{\xi}_p (u) \cdot w.
\]

That is, the exponential covariant derivative $\nabla^e_\eta \xi$ can be viewed as the natural one which assigns at each $q \in \mathcal{E}$ the projection on $B_q$ of the derivative of $\xi : \mathcal{E} \to L^{\Phi_1}$ at $q$ in the direction $\eta (q)$. 

39
\[ \tilde{\xi}_p (u) = \xi_p (u) - \mathbb{E}_q (\xi_p (u)) = \xi_p (u) - DK_p (u) \cdot \xi_p (u) \quad (4.15) \]

where \( \xi_p : \mathbb{V}_p \rightarrow B_p \) is, as usual, the principal part of the local representation \( T_{S_p} \circ \xi \circ e_p \) given by the trivialization (see diagram (4.14)).

Differentiating (4.15), one obtains
\[
D \tilde{\xi}_p (u) \cdot w = D\xi_p (u) \cdot w - D^2 K_p (u) (\xi_p (u), w) - DK_p (u) \cdot (D\xi_p (u) \cdot w) = D\xi_p (u) \cdot w - D^2 K_p (u) (\xi_p (u), w) - \mathbb{E}_q [D\xi_p (u) \cdot w] \quad (4.16)
\]

and, since
\[
D^2 K_p (u) (\xi_p (u), w) = Cov_q (\xi_p (u), w) = \mathbb{E}_q [\xi_p (u) \cdot w] - \mathbb{E}_q [\xi_p (u)] \mathbb{E}_q (w),
\]

Equation (4.16) becomes
\[
D \tilde{\xi}_p (u) \cdot w = \left( D\xi_p (u) \cdot w - \mathbb{E}_q [D\xi_p (u) \cdot w]\right) - \mathbb{E}_q [\xi_p (u) \cdot w]. \quad (4.17)
\]

Compare Equations (4.17) and (4.7) of page 32. In finite dimensional case, \( \tilde{\xi} \) decomposes as a sum with the vector \( D\varphi^{-1} (u) \cdot \left( \sum_{i=1}^m D\xi_i (u) \cdot \dot{u} \frac{\partial}{\partial u_i} \right) \in T_q M \) as addend.

Recall that \( \sum_{i=1}^m \xi_i (u) \frac{\partial}{\partial u_i} \) is the principal part of the local representation of \( \xi \). In this non parametric case, \( D\tilde{\xi}_p (u) \cdot w \) decomposes as a sum of the rate of change at \( u \) in the direction \( w \) of the principal part of the local representation of \( \xi \) (written in the proper way as a vector of \( B_q \)) and a second term in \( L^{q_1} \). We see that this second term is constant, that is it belongs to the complementary of \( B_q \) in the splitting \( L^{q_1} = B_q \oplus \mathbb{R} \).

By Eq. (4.4) of page 30 and above Eq. (4.17), we have
\[
\left( \nabla^e_\eta \xi \right) (q) = D \tilde{\xi}_p (u) \cdot w - \mathbb{E}_q \left( D \tilde{\xi}_p (u) \cdot w\right) = D\xi_p (u) \cdot w - \mathbb{E}_q (D\xi_p (u) \cdot w)
\]

so we find the expected relation
\[
\left( \nabla^e_\eta \xi \right) (q) = \left( \nabla^e_\eta \xi \right) (p) (u) = \mathbb{E}_q \left[ \left( \nabla^e_\eta \xi \right) (p) (u) \right]
\]

between the covariant derivative \( \nabla^e_\eta \xi \) expressed in the absolute coordinates of \( L^{q_1} \) and the principal part of its local representation \( \left( \nabla^e_\eta \xi \right) (p) \) expressed in the relative coordinates given by the local chart \( (TU_p, TS_p) \).

Fix \( w \in B_p \) and define the local vector field \( F_w \in \mathfrak{X} (U_p) \)
\[
F_w (q) := T e_p (u, w) = w - \mathbb{E}_q (w)
\]
4.4 Mixture Connection

for each \( q = e_p(u) \in \mathcal{U}_p \). Note that, in Proposition 62, \( \gamma(t) = F_w(\alpha(t)) \). If \( \nabla \) is an arbitrary affine connection in \( T\mathcal{E} \), then by Eq. (4.5) the principal part of the covariant derivative \( \nabla_{F_{w_1}} F_{w_2} \) for each \( w_1, w_2 \in B_p \) is

\[
\left( \nabla_{F_{w_1}} F_{w_2} \right)_p = D(F_{w_1})_p + \Gamma_p(w_1, w_2)
\]

Taking the derivative of \( F_{w_1} \) (as a function with range in \( L^{\Phi_1} \)) at \( p \) along a curve \( \sigma: (-\varepsilon, \varepsilon) \to \mathcal{U}_p \) such that \( \sigma(0) = p \) and \( \dot{\sigma}(0) = w_2 \) we have

\[
\dot{F}_{w_1}(0) = D F_{w_1}(0) \cdot w_2 = -D^2 K_p(0) \cdot (w_1, w_2).
\]

Projecting on \( B_p \) we find

\[
0 = \Pi_p D F_{w_1}(0) \cdot w_2 = \nabla^e_p F_{w_1}(p) = \left( \nabla^e_{F_{w_1}} F_{w_2} \right)_p(0) = \Gamma^e_p(w_1, w_2) = 0.
\]

Hence identity \( \Gamma^e = 0 \) is equivalent to the following fact: at each density \( p \) and in each direction \( w \in B_p \), the rate of change in \( L^{\Phi_1} \) of a local vector field constant in the trivialization around \( p \) is always complementary to \( B_p \). This suggests a geometric interpretation: tangent spaces \( B_p \) at each \( p \in \mathcal{E} \) are embedded in \( L^{\Phi_1} \) in such a way that their motion along a curve doesn’t affect the natural covariant derivative of a vector field along it.

4.4 Mixture Connection

We shall define a connection in the pre-tangent bundle over \( \mathcal{E} \) defined in Example 39.

**Definition 64.** We shall denote mixture connection in the pre-tangent bundle \( ^*T\mathcal{E} \) the connection such that for each chart \( \{ (\mathcal{U}_p, s_p) : p \in \mathcal{E} \} \) and for each \( u \in \mathcal{U}_p \), the Christoffel symbols \( \Gamma^m_p(u) \in \mathcal{L}^2(B_p, ^*B_p, ^*B_p) \) are the null operator.

Mixture connection is well defined. Let \( (\mathcal{U}_{p_1}, s_{p_1}) \) and \( (\mathcal{U}_{p_2}, s_{p_2}) \) be a pair of chart such that \( \mathcal{U}_{p_1} \cap \mathcal{U}_{p_2} \neq \emptyset \). We put \( f := s_{p_2} \circ e_{p_1}, L = P_{p_2}^{m_{p_1}} \) and we use \( F = (f, L) \) like in the diagram (4.2). Since \( L \) is independent of \( u \), \( DL \equiv 0 \). For each \( u \in \mathcal{V}_{p_1} \) and \( w_1, w_2 \in B_{p_1} \), the change of variables formula (4.3) for Christoffel symbols simplifies to

\[
\Gamma^m_{p_2}(f(u)) \left( Df(u) \cdot w_1, Df(u) \cdot w_2 \right) = D f(u) \cdot \left( \underbrace{\Gamma^m_{p_1}(u)(w_1, w_2)}_{=0} \right) = 0.
\]

41
Proposition 65. The mixture connection pair \((\ast T\mathcal{E}, \nabla^m)\) is globally flat and the associated parallel transport \(P^m\) is:

\[
P^m_{p,q} : \ast B_p \ni w \mapsto w^p_q \in \ast B_q.
\]

Proof. Let \(\alpha : (-\varepsilon, \varepsilon) \to \mathcal{E}\) be a curve of class \(C^2\) with \(\alpha(0) = p\) and \(\alpha(1) = q\). We may assume that \(p, q \in U_p\), the domain of the chart \((U_p, s_p)\). For \(w \in \ast B_p\), let \(\gamma : (-\varepsilon, \varepsilon) \to \ast T\mathcal{E}\) be the lift of \(\alpha\) defined by

\[
\gamma(t) := w^p_{\alpha(t)} \in \ast B_{\alpha(t)}.
\]

In the chart \((U_p, s_p)\), the principal part of the local representation of \(\gamma\) is the constant curve \(\gamma_p(t) = w\). Derivative \(\dot{\gamma}_p(t) = 0\), so \(\gamma_p\) satisfies Eq. (4.11). Curve \(\gamma\) is \(\alpha\)-parallel and the parallel transport \(P_\alpha(w) = \gamma^p_q\) is independent of \(\alpha\). \(\square\)

4.5 Geometry of the sphere in \(L^a(\mu)\)

We show that the sphere of radius \(r\) in the Lebesgue space \(L^a(\mu)\) is a \(C^\infty\)-manifold and it is a submanifold of \(L^a(\mu)\) with less regularity. Then we introduce a set of Christoffel symbols for the sphere.

Bonic and Frampton (1966) show that the \(a\)-th power \(\|\cdot\|^a\) of the canonical norm in the Lebesgue space \(L^a(\mu)\) with \(a > 1\) is differentiable, further

\[
\|\cdot\|^a : L^a(\mu) \to [0, \infty) \text{ is of class } \begin{cases} 
C^{a-1} & \text{if } a \text{ is odd} \\
C^a & \text{if } a \in (1, \infty) \setminus \mathbb{N} \\
C^\infty & \text{if } a \text{ is even}
\end{cases}
\]

(4.18)

where \([a]\) is the greatest integer less or equal to \(a\). For \(1 \leq k \leq a - 1\) if \(a\) is odd otherwise for \(1 \leq k \leq [a]\), the \(k\)-th derivative of \(\|\cdot\|^a\) at \(f \in L^a(\mu)\) is the continuous \(k\)-linear form on \(L^a(\mu)\) defined by:

\[
(D^k \|\cdot\|^a)(f)(h_1, \ldots, h_k)
\]

\[
= a(a - 1) \cdots (a - k + 1) \int (\text{sgn } f)^k |f|^{a-k} h_1 \cdots h_k d\mu
\]

for \((h_1, \ldots, h_k) \in (L^a(\mu))^k\).

If \(a\) is even, then \(\|\cdot\|^a\) is \(C^\infty\) and \(D^k \|\cdot\|^a\) is null for each \(k > a\).

The canonical norm \(\|\cdot\|\) is differentiable away from 0 with the same regularity given
4.5– Geometry of the sphere in $L^a(\mu)$

by (4.18) above. Its differential is the mapping $d\|\cdot\|_a : L^a(\mu) \setminus \{0\} \to L^a(\mu)^*$ such that, evaluated at $f$, it gives the linear functional:

$$d\|\cdot\|_a(f) \cdot = \|f\|_a^{1-a} \int \cdot \text{sgn} f |f|^{a-1} d\mu.$$  

In particular, if for each $f \in L^a(\mu)$ we shall denote

$$f^* := \|f\|_a^{1-a} \text{sgn} f |f|^{a-1} = \left( \int |f|^a d\mu \right)^{-\frac{1}{b}} \text{sgn} f |f|^\frac{a}{b} \in L^b(\mu)$$

where $b > 1$ is the conjugate exponent of $a$, i.e. the real number such that $\frac{1}{a} + \frac{1}{b} = 1$,

then

$$d\|\cdot\|_a(f) \cdot f = \int f f^* d\mu = \|f\|_a.$$  

(4.19)

Note that $\|\cdot\|_a$ is a positively homogeneous function and Eq. (4.19) comes also from Euler’s Theorem (see Theorem 84).

If $\langle , \rangle$ is the pairing

$$\langle , \rangle : L^a(\mu) \times L^b(\mu) \ni (g,h) \mapsto \langle g,h \rangle = \int gh d\mu \in \mathbb{R},$$

then $f^* = \frac{\delta \|\cdot\|_a}{\delta f}$ is the functional derivative of the norm with respect to $f$ that is the unique element of $L^b(\mu)$ such that for each $g \in L^a(\mu)$

$$d\|\cdot\|_a(f) \cdot g = \left\langle g, \frac{\delta \|\cdot\|_a}{\delta f} \right\rangle$$

(see (Abraham et al., 1988, Supplement 2.4C)).

Let

$$S^a(r) = \{ f \in L^a(\mu) : \|f\|_a = r \}$$

be the sphere of radius $r$ in the Lebesgue space $L^a(\mu)$ and let $S^a$ be the unit sphere.

**Proposition 66.** The sphere $S^a(r)$ is a closed submanifold of $L^a(\mu)$ of class $C^{a-1}$ if $a$ is odd, $C^a$ if $a \in (1, \infty) \setminus \mathbb{N}$ and $C^{\infty}$ if $a$ is even. The tangent space at $f \in S^a(r)$ is the closed hyperplane

$$T_fS^a(r) = \left\{ g \in L^a(\mu) : \int g f^* d\mu = 0 \right\}$$

where $f^* = r^{1-a} \text{sgn} f |f|^{a-1}$.  

43
Proof. We shall use the Submersion Theorem (see, for example, Abraham et al. (1988, Theorem 3.5.4)). Since $S^a(r) = \|\cdot\|^{-1}_a(r)$, we have to prove that $r$ is a regular value of $\|\cdot\|_a$.

For each $f \in S^a(r)$, the closed subspace

$$\ker \left( d \|\cdot\|_a (f) \right) = \left\{ g \in L^a(\mu) : \int g f^* d\mu = 0 \right\}$$

splits in $L^a(\mu)$. In fact, by Proposition 56, $L^a(\mu)$ can be decomposed as a direct sum

$$L^a(\mu) = \ker \left( d \|\cdot\|_a (f) \right) \oplus \langle f \rangle$$

and, according to Equation (4.13), each $h \in L^a(\mu)$ can be decomposed as the sum

$$h = \left( h - \int h f^* d\mu \frac{f}{r} \right) + \int h f^* d\mu \frac{f}{r}.$$  \hspace{1cm} (4.20)

where we use $d \|\cdot\|_a (f) \cdot f = r$ (see Eq. (4.19)). As the range of $d \|\cdot\|_a (f)$ is a subspace of $\mathbb{R}$ of dim 1, $d \|\cdot\|_a (f)$ is surjective. Hence $r$ is a regular value of the function $\|\cdot\|_a$ and the sphere $S^a(r)$ is a closed submanifold with $\ker (d \|\cdot\|_a (f))$ as tangent space at $f$.

From Eq. (4.20) one can see that, for each $f \in S^a(r)$, the mapping $\Pi_f$

$$\Pi_f : L^a(\mu) \ni h \mapsto h - \int h f^* d\mu \frac{f}{\|f\|_a} \in T_f S^a(r).$$  \hspace{1cm} (4.21)

is the projection from $L^a(\mu)$ into $T_f S^a(r)$.

Remark. Proposition 66 can be viewed in the light of a more abstract setting. Let $X$ be a real Banach space, $X^*$ be its dual space and $S \subset X$ be the unit sphere. We give the following definitions.

**Definition 67.** $X$ is uniformly convex if for each $\varepsilon > 0$ there exists $\delta > 0$ such that $x, y \in S$ and $\frac{x + y}{2} \not\in S$ implies $\|x - y\| < \varepsilon$.

**Definition 68.** We say that $X$ has the projection property if for any closed convex $M \subset X$ and any $x \in X$ there is a unique $m \in M$ such that

$$\|x - m\| = \inf \{\|x - z\| : z \in M\} = d(x, M).$$

In this case we define $\pi_M(x) := m$.

**Definition 69.** We say that $x$ is orthogonal to $y$ and denote it $x \perp y$ if $\|x\| \leq \|x + \alpha y\|$ for any $\alpha \in \mathbb{R}$.
4.5 - Geometry of the sphere in $L^a(\mu)$

**Definition 70.** The duality mapping $J : X \rightarrow 2^{X^*}$ is defined by

$$J (x) := \{ \lambda \in X^* : \lambda \cdot x = \|x\|^2 = \|\lambda\|^2 \}.$$

We say that $X$ has the duality map property if $J$ is single-valued. In this case we set $\tilde{x} := J (x)$.

Note that by the Hahn-Banach Theorem $J (x) \neq \emptyset$ for each $x \in X$. Following propositions are proved by Gibilisco and Isola (1999).

**Proposition 71.** Let $X$ and $X^*$ be uniformly convex Banach spaces. Then

i) $X$ has the projection property;

ii) $X$ has the duality map property;

iii) $x \perp \ker \tilde{x}$;

iv) if $M := \ker \tilde{x}$, then $\pi_M (v) = v - \frac{\tilde{x} \cdot v}{\tilde{x} \cdot x} x$.

**Proposition 72.** Let $X$ and $X^*$ be uniformly convex Banach spaces. Then

i) $S$ is a Banach submanifold of $X$;

ii) $T_x S$, the tangent space to $S$ at $x \in S$, can be identified with $\ker \tilde{x}$;

iii) the projection operator $\pi_x : T_x X \rightarrow T_x S$ is given by $\pi_x (v) = v - \tilde{x} \cdot v x$.

Lebesgue spaces $L^a(\mu)$ are uniformly convex for every $a > 1$ (see, for example, Habala et al. (1996, Ch. 11)). The duality map is single-valued and $J : L^a(\mu) \rightarrow L^b(\mu)$, with $a + b = ab$, is given by

$$L^a(\mu) \ni f \mapsto \tilde{f} = \|f\|^2_a sgn f |f|^b_a \in L^b(\mu).$$

In fact, one can check that

$$\int \tilde{f} f d\mu = \|f\|^2_a = \|\tilde{f}\|^2_b.$$ 

Note that $\tilde{f} = \|f\|_a f^*$. If $M = \ker \left( \int \tilde{f} \cdot d\mu \right) = \ker \left( \int f^* \cdot d\mu \right)$, then

$$\pi_M (h) = h - \frac{\int \tilde{f} h d\mu}{\int \tilde{f} f d\mu} f = h - \frac{\int f^* h d\mu}{\|f\|^2_a} f = \Pi_f (h).$$
In the sequel it will be convenient to cover $S^a(r)$ with charts.

**Proposition 73.** The sphere $S^a(r)$ for $a > 1$ is a $C^\infty$ manifold. In a neighborhood of each $f \in S^a(r)$, the sphere is modeled by the tangent space to $S^a(r)$ at $f$.

**Proof.** We shall cover $S^a(r)$ with the atlas $\{(U_f, \varphi_f) : f \in S^a(r)\}$. For each $f \in S^a(r)$, the set

$$U_f^+ := \left\{ g \in L^a(\mu) : \int g f^* \, d\mu = d \|f\|_a \cdot g > 0 \right\}$$

is an open subset of $L^a(\mu)$. We put $U_f := S^a(r) \cap U_f^+$ and we define the continuous mapping $\varphi_f : U_f \to T_f S^a(r)$ by

$$\varphi_f : \begin{array}{ccc} U_f & \to & T_f S^a(r) \\ g & \mapsto & \frac{r}{\int g f^* \, d\mu} \cdot g - f \end{array}$$

The chart $\varphi_f$ carries $g$ into the projection $\Pi_f \left( \frac{r}{\int g f^* \, d\mu} \cdot g \right)$ on $T_f S^a(r)$ of the intersection of the linear subspace $\langle g \rangle$ with the affine hyperplane $f + T_f S^a(r)$. See the following figure,

![Figure 4.1. Chart $\varphi_f : U_f \to T_f S^a(r)$](image)

In Proposition 66 we show that $T_f S^a(r)$ is a closed subspace of the Banach space $L^a(\mu)$. The inverse function of $\varphi_f$ gives the parametrization:

$$\varphi_f^{-1} : \begin{array}{ccc} T_f S^a(r) & \to & U_f \\ u & \mapsto & \frac{u + f}{\|u + f\|_a} \cdot r \end{array}$$

Let $f_1, f_2 \in S^a(r)$ be such that $U_{f_1} \cap U_{f_2} \neq \emptyset$. Since $\varphi_{f_1}$ and $\varphi_{f_2}$ are homeomorphisms, $\varphi_{f_i}(U_{f_1} \cap U_{f_2})$ is open in $T_{f_i} S^a(r)$ for $i = 1, 2$. The overlap map $\varphi_{f_2} \circ \varphi_{f_1}^{-1}$ is given by

$$\varphi_{f_1} (U_{f_1} \cap U_{f_2}) \ni u \mapsto \frac{u + f_1}{\int (u + f_1) f_2^* \, d\mu} \cdot r - f_2 \in \varphi_{f_2} (U_{f_1} \cap U_{f_2}) \quad (4.22)$$

and it is $C^\infty$-map (recall that $f_2^* \in L^b(\mu)$).

Then the collection $\{(U_f, \varphi_f) : f \in S^a(r)\}$ satisfies the three axioms for being a $C^\infty$-atlas for $S^a(r)$ (see Lang (1995, Ch. II)).
4.5– Geometry of the sphere in $L^a(\mu)$

We now observe that the system of coordinates for the sphere $S^a(r)$ defined by the stereographic projection is not compatible with the $C^\infty$-structure given by the atlas $\{(U_f,\varphi_f)\}_{f \in S^a(r)}$.

Fix $f \in S^a(r)$. The stereographic projection $\psi_f$ carries $g \in S^a(r) \setminus \{f\}$ into the intersection of $T_f S^a(r)$ with the line

$$l_{gf}(\lambda) = \lambda f + (1 - \lambda) g, \quad \lambda \in \mathbb{R}$$

which connects $g$ to $f$. As

$$l_{gf}(\lambda) \in T_f S^a(r) \iff \int l_{gf}(\lambda) f^*d\mu = \int (\lambda f + (1 - \lambda) g) f^*d\mu = 0$$
$$\iff \lambda r + (1 - \lambda) \int gf^*d\mu = 0$$
$$\iff \lambda = \frac{\int gf^*d\mu}{\int gf^*d\mu - r}$$

we have the chart

$$\psi_f : S^a(r) \setminus \{f\} \to T_f S^a$$

$$g \mapsto \frac{rg - (\int gf^*d\mu)f}{r - \int gf^*d\mu}.$$

If, for example, we consider the pair of charts $(U_f,\varphi_f)$ and $(S^a(r) \setminus \{f\},\psi_f)$ then the change of coordinates $\psi_f \circ \varphi_f^{-1}$ is given by the mapping

$$T_f S^a(r) \setminus \{0\} \ni u \mapsto \psi_f \circ \varphi_f^{-1}(u) = \frac{ru}{\|u + f\|_a - r} \in T_f S^a(r) \setminus \{0\}$$

which is not smooth for each $a$ not even.

If $U_{f_1}$ and $U_{f_2}$ have non-empty intersection for some $f_1,f_2 \in S^a(r)$, then the derivative of $\varphi_{f_2} \circ \varphi_{f_1}^{-1}$ is a toplinear isomorphism between $T_{f_1} S^a(r)$ and $T_{f_2} S^a(r)$. We shall evaluate the derivative of $\varphi_{f_2} \circ \varphi_{f_1}^{-1}$ in Eq. (4.24). When $f_2 \notin T_{f_1} S^a(r)$ also the restriction to $T_{f_1} S^a(r)$ of the projection $\Pi_{f_2}$ given in (4.21) is a toplinear isomorphism between $T_{f_1} S^a(r)$ and $T_{f_2} S^a(r)$. In fact, as $\int f_2 f_1^*d\mu \neq 0$, we define the continuous mapping

$$T_{f_2} S^a(r) \ni g \mapsto g - \frac{\int gf_1^*d\mu}{\int f_2 f_1^*d\mu} f_2 \in T_{f_1} S^a(r)$$

and, with a calculation, we can check that it is the inverse function of $\Pi_{f_2}|_{T_{f_1} S^a(r)}$. 47
The map \( \varphi_f \) can be extended to a function \( \varphi_f : U_f^+ \rightarrow T_fS^a(r) \) constant on the rays. Observe that \((U_f,\varphi_f)\) comes from a chart \((U_f^+,\varphi_f^+)\) of \(L^a(\mu)\) with the submanifold property, namely
\[
U_f^+ \ni g \rightarrow \varphi_f^+(g) = \left( \frac{r}{\int g f^*d\mu} g - f, \|g\|_a \right) \in T_fS^a(r) \times (0,\infty)
\]
and \(\varphi_f^+(U_f) = T_fS^a(r) \times \{r\}\).

The derivative of \(\varphi_f\), \(D\varphi_f : U_f^+ \rightarrow \mathcal{L}(L^a(\mu),T_fS^a(r))\), is given by:
\[
D\varphi_f(g) \cdot h = \frac{r}{\int g f^*d\mu} h - \frac{r}{\int g f^*d\mu} \left( \frac{\int h f^*d\mu}{\sqrt{\int g f^*d\mu}} \right) g
\]
(4.23)
for each \(g \in U_f^+\) and \(h \in L^a(\mu)\). The derivative of its inverse function, \(D\varphi_f^{-1} : T_fS^a(r) \rightarrow \mathcal{L}(T_fS^a(r),L^a(\mu))\), is given by:
\[
D\varphi_f^{-1}(u) \cdot w = \frac{rw}{\|u + f\|_a} - \frac{r}{\|u + f\|_a^2} \left( \frac{\int w f^*d\mu}{\sqrt{\int f f^*d\mu}} \right) (u + f)
\]
for each \(u,w \in T_fS^a(r)\).

Note that \(D\varphi_f(g)\) maps the tangent hyperplane \(T_gS^a(r)\) into \(T_fS^a(r)\), \(D\varphi_f^{-1}(\varphi_f(g))\) maps \(T_fS^a(r)\) into \(T_gS^a(r)\) and \(D\varphi_f^{-1}(\varphi_f(g)) = (D\varphi_f(g))^{-1}\).

If \(U_{f_1} \cap U_{f_2} \neq \emptyset\) for some \(f_1,f_2 \in S^a(r)\), then the first and second derivative of the change of coordinates \(\varphi_{f_2} \circ \varphi_{f_1}^{-1}\) have the following equations:

**I.** \(D(\varphi_{f_2} \circ \varphi_{f_1}^{-1})(u) \in \mathcal{L}(T_{f_1}S^a(r) ; T_{f_2}S^a(r))\) has equation:
\[
D(\varphi_{f_2} \circ \varphi_{f_1}^{-1})(u) \cdot w = \frac{rw}{\int (u + f_1) f_2^*d\mu} - \frac{r}{\int (u + f_1) f_2^*d\mu} \left( \frac{\int w f_2^*d\mu}{\sqrt{\int f_2 f_2^*d\mu}} \right) (u + f_1)
\]
(4.24)
for each \(u \in \varphi_{f_1}(U_{f_1} \cap U_{f_2})\) and \(w \in T_{f_1}S^a(r)\);

**II.** \(D^2(\varphi_{f_2} \circ \varphi_{f_1}^{-1})(u) \in \mathcal{L}^2(T_{f_1}S^a(r) ; T_{f_2}S^a(r))\) has equation:
\[
D^2(\varphi_{f_2} \circ \varphi_{f_1}^{-1})(u) \cdot (w_1,w_2) = \frac{2r}{\sqrt{\int (u + f_1) f_2^*d\mu}} \left( \frac{\int w_1 f_2^*d\mu}{\sqrt{\int f_2 f_2^*d\mu}} \right) w_1 + \frac{r}{\sqrt{\int f_2 f_2^*d\mu}} \left( \frac{\int w_2 f_2^*d\mu}{\sqrt{\int f_2 f_2^*d\mu}} \right) w_2
\]
(4.25)
for each \(u \in \varphi_{f_1}(U_{f_1} \cap U_{f_2})\) and \(w_1,w_2 \in T_{f_1}S^a(r)\).
Associated to the atlas \( \{(U_f, \varphi_f)\}_{f \in S^a(r)} \) we have a natural atlas of the tangent bundle \( T S^a(r) \) over \( S^a(r) \): \( \{(TU_f, T \varphi_f)\}_{f \in S^a(r)} \) where

\[
T \varphi_f = (\varphi_f, D \varphi_f) : TU_f \simeq U_f \times T_f S^a(r) \to T_f S^a(r) \times T_f S^a(r).
\]

We are ready to define a connection on the sphere \( S^a(r) \) with \( a \geq 2 \) by a set of Christoffel symbols. We cover the sphere with the charts \( (U_f, \varphi_f) \) however \( S^a(r) \) is taken with the regularity of Proposition 66.

**Proposition 74.** Let the atlas \( \{(U_f, \varphi_f) : f \in S^a(r)\} \) of \( S^a(r) \) with \( a \geq 2 \) be given, all the mappings \( \Gamma_f : T_f S^a(r) \to \mathcal{L}^2(T_f S^a(r); T_f S^a(r)) \) defined by

\[
\Gamma_f(u)(w_1, w_2) = \frac{\left( \int w_2 (u + f)^* d\mu \right) w_1 + \left( \int w_1 (u + f)^* d\mu \right) w_2}{\|u + f\|_a}
\]

for each \( u, w_1, w_2 \in T_f S^a(r) \) are a set of Christoffel symbols of the tangent bundle \( T S^a(r) \) with regularity \( C^{a-2} \) if \( a \) is odd, \( C^{[a]-1} \) if \( a \in (2, \infty) \setminus \mathbb{N} \) and \( C^\infty \) if \( a \) is even.

The connection defined by \( \Gamma_f \) is the natural one induced by the projection on \( S^a(r) \) of the trivial connection on \( L^a(\mu) \).

**Proof.** If \( U_{f_1} \cap U_{f_2} \neq \emptyset \) for some \( f_1, f_2 \in S^a(r) \), using the change of coordinates formula (4.22) and its derivatives given by Equations (4.24) and (4.25), one can check with a calculation that, for each \( u \in \varphi_{f_1}(U_{f_1} \cap U_{f_2}) \) and \( w_1, w_2 \in T_{f_1} S^a(r) \),

\[
\Gamma_{f_2} (\varphi_{f_2} \circ \varphi_{f_1}^{-1})(u) \left( D (\varphi_{f_2} \circ \varphi_{f_1}^{-1})(u) \cdot w_1, D (\varphi_{f_2} \circ \varphi_{f_1}^{-1})(u) \cdot w_2 \right) = -D^2 (\varphi_{f_2} \circ \varphi_{f_1}^{-1})(u)(w_1, w_2) + D (\varphi_{f_2} \circ \varphi_{f_1}^{-1})(u) \cdot \Gamma_{f_1}(w_1, w_2).
\]

Hence the locally defined continuous bilinear mappings \( \Gamma_f \) glue together according to the change of variables formula (4.3) for Christoffel symbols.

For each \( f \in S^a(r) \) and \( u, w_1, w_2 \in T_f S^a(r) \), we can write

\[
\Gamma_f(u)(w_1, w_2) = \frac{\left( \int D \varphi_f^{-1}(u) \cdot w_1 f^* d\mu \right) w_2 + \left( \int D \varphi_f^{-1}(u) \cdot w_2 f^* d\mu \right) w_1}{\int \varphi_f^{-1}(u) f^* d\mu}. \tag{4.26}
\]

By this equality one can see the regularity of the mapping

\[
T_f S^a(r) \ni u \mapsto \Gamma_f(u) \in \mathcal{L}^2(T_f S^a(r); T_f S^a(r)).
\]

We set \( (T S^a(r), \nabla^a) \) the bundle connection pair defined by this family of Christoffel symbols.

Let \( \xi \in \mathfrak{X}(S^a(r)) \) be a vector field. For each chart \( (U_f, \varphi_f) \), we put \( \tilde{\xi}_f := \xi \circ \varphi_f^{-1} : T_f S^a(r) \to T S^a(r) \subset L^a(\mu) \). The principal part of the local representation of \( \xi \) is
\( \xi_f = D\varphi_f (\varphi_f^{-1}) \cdot \tilde{\xi}_f : T_f S^a (r) \to T_f S^a (r) \).

Let \( \sigma : (-\varepsilon, \varepsilon) \to U_f \) be a \( C^2 \)-curve such that \( \sigma (0) = g = \varphi_f^{-1} (u) \) for an element \( u \in T_f S^a (r) \), \( \dot{\sigma} (0) = v \in T_g S^a (r) \) and \( w = D\varphi_f (g) \cdot v \).

We have to prove:

\[
\nabla^g_a \xi (g) = \Pi_g D\tilde{\xi}_f (u) \cdot w
\]
or, if \( (\nabla^a \xi)_f \) is the principal part of the local representation of the covariant derivative,

\[
(\nabla^a \xi)_f (u) = D\xi_f (u) \cdot w + \Gamma_f (u) (w, \xi_f (u)) = D\varphi_f (g) \cdot \left( \Pi_g D\tilde{\xi}_f (u) \cdot w \right).
\]  \hspace{1cm} (4.27)

Using Equations (4.21) and (4.23), we obtain

\[
D\varphi_f (g) \cdot \left( \Pi_g D\tilde{\xi}_f (u) \cdot w \right) = \frac{r}{\int g f^* d\mu} D\tilde{\xi}_f (u) \cdot w - r \frac{\int D\tilde{\xi}_f (u) \cdot w f^* d\mu}{(\int g f^* d\mu)^2} g
\]

\[
= D\varphi_f (g) \cdot \left( D\tilde{\xi}_f (u) \cdot w \right).
\]

Since the derivative of \( \xi_f \) evaluated at \( u \) in the direction \( w \) is

\[
D\xi_f (u) \cdot w = D^2 \varphi_f (g) \left( D\varphi_f^{-1} (u) \cdot w, \tilde{\xi}_f (u) \right) + D\varphi_f (g) \cdot \left( D\tilde{\xi}_f (u) \cdot w \right),
\]

Equation (4.27) holds if and only if

\[
D^2 \varphi_f (g) (v, \tilde{\xi}_f (u)) = -\Gamma_f (u) (w, \xi_f (u)).
\]

We evaluate \( \Gamma_f (u) (w, \xi_f (u)) \) using Equation (4.26) and replacing

\[
\xi_f (u) = D\varphi_f (g) \cdot \tilde{\xi}_f (u) = \frac{r}{\int g f^* d\mu} \tilde{\xi}_f (u) - \frac{r}{\int g f^* d\mu} \frac{\int \tilde{\xi}_f (u) f^* d\mu}{(\int g f^* d\mu)^2} g
\]

and

\[
w = D\varphi_f (g) \cdot v = \frac{r}{\int g f^*} v - \frac{r}{\int g f^*} \frac{\int v f^* d\mu}{(\int g f^* d\mu)^2} g
\]

we obtain

\[
\Gamma_f (u) (w, \xi_f (u))
\]

\[
= \frac{r}{(\int g f^* d\mu)^2} \tilde{\xi}_f (u) + \frac{r}{(\int g f^* d\mu)^2} \frac{\int \tilde{\xi}_f (u) f^* d\mu}{(\int g f^* d\mu)^2} v
\]

\[
- 2 \frac{\left( \int v f^* d\mu \right) \left( \int \tilde{\xi}_f (u) f^* d\mu \right) g}{(\int g f^* d\mu)^3}
\]

\[
= -D^2 \varphi_f (g) (v, \tilde{\xi}_f (u)).
\]
4.6 Amari $\alpha$-embedding

Let $\alpha := (a - 2) / a \ (a > 1 \Rightarrow \alpha \in (-1, 1))$, the Amari $\alpha$-embedding $A^\alpha : \mathcal{M}_\mu \rightarrow L^a (\mu)$ is defined by:

$$\mathcal{M}_\mu \ni p \mapsto p^{\frac{1}{a}} \in L^a (\mu).$$

Its range lies in $S^a$.

**Proposition 75.** The Amari $\alpha$-embedding $A^\alpha : \mathcal{M}_\mu \rightarrow L^a (\mu)$ is an analytic mapping.

**Proof.** For each pair of charts $(U_{p_1}, s_{p_1})$ and $(U_{p_2}, s_{p_2})$ of $\mathcal{M}_\mu$ with $U_{p_1} \cap U_{p_2} \neq \emptyset$, the transition mapping

$$s_{p_2} \circ e_{p_1} (u) = u - E_{p_2} (u) + \log \frac{p_1}{p_2} \left( \log \frac{p_1}{p_2} \right)$$

is affine so $\{(U_p, s_p) : p \in \mathcal{M}_\mu\}$ is actually a $C^\infty$-atlas.

With respect to a local chart $(U_p, s_p)$, the Amari $\alpha$-embedding is locally defined by $A^\alpha \circ e_p : V_p \rightarrow L^a (\mu)$

$$A^\alpha \circ e_p : V_p \ni u \mapsto q^{\frac{1}{a}} = e^{u - K_p (u)} p^{\frac{1}{a}} \in L^a (\mu).$$

Let $B (0, 1) \subset L^{\Phi_1} (p \cdot \mu)$ be the open unit ball. In Section 2.2 we define the analytic mapping $\exp_{p,a} : B (0, 1) \rightarrow L^a (p \cdot \mu)$ as:

$$v \mapsto \sum_{n=0}^{\infty} \frac{1}{n!} \left( \frac{v}{a} \right)^n = e^\frac{v}{a}.$$

Moreover, in Theorem 22 the analyticity of the restriction to $B (0,1)$ of the moment generating functional $M_p$ is proved. Hence the mapping

$$V_p \ni u \mapsto \left( \frac{1}{M_p (u)} \right)^{\frac{1}{a}} \exp_{p,a} (u) \in L^a (p \cdot \mu)$$

is analytic.

If we resort to the norm-preserving linear function $I_p^{(a-2)/a} \in \mathcal{L} \left( L^a (p \cdot \mu) ; L^a (\mu) \right)$ defined by

$$I_p^{a-2} : L^a (p \cdot \mu) \ni f \mapsto f p^{\frac{1}{a}} \in L^a (\mu),$$

then we can write

$$A^\alpha \circ e_p (u) = I_p^{a-2} \left( \frac{1}{(M_p (u))^{\frac{1}{a}} \exp_{p,a} (u)} \right)$$

and we see the $C^\infty$-regularity of the local representation $A^\alpha \circ e_p$. \qed
Remark. For each \( p \in \mathcal{M}_\mu \), by the continuity of \( M_p : B (0,1) \to [1,\infty) \) there exists a neighborhood \( \mathcal{V}_p' \subset \mathcal{V}_p \) of 0 such that
\[
M_p (u) M_p (-u) < 3. \tag{4.28}
\]
For each \( u \in \mathcal{V}_p' \), by bound (4.28) we obtain
\[
\int \cosh (u - K_p (u)) p d\mu = \frac{1}{2} \left( 1 + M_p (u) M_p (-u) \right) < 2
\]
and, recalling that for each \( 0 \neq f \in L_{p}^\Phi (p \cdot \mu) \) we have
\[
\|f\|_{\Phi_1} < 1 \iff \mathbb{E}_p (\Phi_1 (f)) < 1 \iff \mathbb{E}_p (\cosh f) < 2,
\]
we can conclude that \( \|u - K_p (u)\|_{\Phi_1} < 1 \). We have proved that if \( u \in \mathcal{V}_p' \), then \( u - K_p (u) \) lies in the domain of \( \exp_{p,a} \) so we have
\[
\exp_{p,a} (u - K_p (u)) = e^{\frac{u - K_p (u)}{a}}.
\]
Finally, the restriction of \( A^\alpha \circ e_p \) to \( \mathcal{V}_p' \) can be written as
\[
A^\alpha \circ e_p (u) = I_p^{a-2} \left( \exp_{p,a} \left( u - K_p (u) \right) \right).
\]

The range of \( A^\alpha \circ e_p \) lies in \( U_{p^{1/a}} \), so, composing with the coordinate function \( \varphi_{p^{1/a}} \), one obtains the local representation \( A^\alpha_p := \varphi_{p^{1/a}} \circ A^\alpha \circ e_p \):

\[
\begin{align*}
\mathcal{U}_p & \xrightarrow{A^\alpha} U_{p^{1/a}} \\
\mathcal{V}_p & \xrightarrow{A^\alpha_p} T_{p^{1/a}} S^a
\end{align*}
\]
such that for each \( u \in \mathcal{V}_p 
\]
\[
A^\alpha_p (u) = \frac{q_{\frac{1}{a}}}{\mathbb{E}_p \left[ \left( \frac{q}{p} \right)^{1/a} \right]} - p_{\frac{1}{a}} = \frac{e^{\frac{u - K_p (u)}{a}}}{\mathbb{E}_p \left[ e^{\frac{u - K_p (u)}{a}} \right]} - p_{\frac{1}{a}}.
\]
To obtain the above expression for \( A^\alpha_p \) and in the following calculations, it should be noted that \( (p^{1/a})^* = p^{1/b} \).
Taking the derivative $DA^\alpha_p : V_p \to L(B_p;T_{p^{1/a}}S^a)$, for each $w \in B_p$ we have

$$DA^\alpha_p (u) \cdot w = \frac{q^{1/a}}{a \mathbb{E}_p \left[ \left( \frac{q}{p} \right)^{1/a} \right]} \left[ w - \frac{\mathbb{E}_p \left[ w \left( \frac{q}{p} \right)^{1/a} \right]}{\mathbb{E}_p \left[ \left( \frac{q}{p} \right)^{1/a} \right]} \right]$$

(4.29)

where $q = e^{u-K_p(u)}$.

When we take a chart $(U_{p^{1/a}}, \psi_{p^{1/a}})$ centered in a point $p^{1/a}$ of the unit sphere coming from a density, the coordinate function $\psi_{p^{1/a}}$ evaluated in another point $q^{1/a}$ coming from $M_\mu$ simplifies to

$$\psi_{p^{1/a}} (q^{1/a}) = \frac{q^{1/a}}{\mathbb{E}_p \left[ \left( \frac{q}{p} \right)^{1/a} \right]} - p^{1/a}$$

and the Christoffel symbol $\Gamma_{p^{1/a}}$ of the sphere at $u = \psi_{p^{1/a}} (q^{1/a})$ becomes the bilinear map such that:

$$\Gamma_{p^{1/a}} (u) (v_1,v_2) = -\mathbb{E}_p \left[ \left( \frac{q}{p} \right)^{1/a} \right] \left( \int v_2 q^{1/a} d\mu \ v_1 + \int v_1 q^{1/a} d\mu \ v_2 \right)$$

for each $v_1,v_2 \in T_{p^{1/a}}S^a$.

### 4.7 $\mathcal{F}^\alpha$-vector bundle

Let $\alpha = (a - 2)/a$ with $a > 1$ and let

$$L^a_0 (p \cdot \mu) := \{ w \in L^a (p \cdot \mu) : \mathbb{E}_p (w) = 0 \}$$

be the closed subspace of all centered random variables in the Lebesgue space $L^a (p \cdot \mu)$ with $p \in M_\mu$. We shall define a vector bundle over $M_\mu$ modeled by these Banach spaces.

**Lemma 76.** For each pair $p,q \in M_\mu$, $L^a_0 (p \cdot \mu)$ and $L^a_0 (q \cdot \mu)$ are toplinearly isomorphic.

**Proof.** Linear mapping

$$L^a_0 (p \cdot \mu) \ni w \mapsto w \left( \frac{p}{q} \right)^{1/a} - \mathbb{E}_q \left[ w \left( \frac{p}{q} \right)^{1/a} \right] \in L^a_0 (q \cdot \mu)$$
is continuous with inverse

\[
L_o^a(q \cdot \mu) \ni v \mapsto v \left(\frac{q}{p}\right)^{\frac{1}{a}} - \frac{E_p \left[ v \left(\frac{q}{p}\right)^{\frac{1}{a}} \right]}{E_p \left[ \left(\frac{q}{p}\right)^{\frac{1}{a}} \right]} \in L_o^a(p \cdot \mu). \tag{4.30}
\]

**Definition 77.** Let \(F^\alpha := \bigcup_{p \in M_\mu} L_o^a(p \cdot \mu)\) be a disjoint union. The \(F^\alpha\) vector bundle over \(M_\mu\) is

\[
\pi : F^\alpha \to M_\mu
\]

with fiber \(\pi^{-1}(p) = L_o^a(p \cdot \mu)\) over each density \(p \in M_\mu\).

For each open domain \(U_p\) of a chart \((U_p, s_p)\) of \(M_\mu\), there exists a trivialization:

\[
\pi^{-1}(U_p) \xrightarrow{\tau^\alpha_p} U_p \times L_o^a(p \cdot \mu)
\]

where:

\[
(\tau^\alpha_p)^{-1} : U_p \times L_o^a(p \cdot \mu) \ni (q, w) \mapsto w \left(\frac{p}{q}\right)^{\frac{1}{a}} - \frac{E_q \left[ w \left(\frac{p}{q}\right)^{\frac{1}{a}} \right]}{E_q \left[ \left(\frac{p}{q}\right)^{\frac{1}{a}} \right]} \in L_o^a(q \cdot \mu).
\]

\(\tau^\alpha_p\) is defined by the family of mappings (4.30) according to \(q \in U_p\).

The collection of charts \(\{(\pi^{-1}(U_p), \Phi^\alpha_p) : p \in M_\mu\}\) where

\[
\Phi^\alpha_p := (s_p, \text{id}) \circ \tau^\alpha_p : \pi^{-1}(U_p) \to V_p \times L_o^a(p \cdot \mu)
\]

is an atlas of \(F^\alpha\). If \(U_{p_1} \cap U_{p_2} \neq \emptyset\) for some \(p_1, p_2 \in M_\mu\), then the overlap map \(\Phi_{p_2} \circ \Phi_{p_1}^{-1}\) has principal part given by:

\[
V_{p_1} \times L_o^a(p_1 \cdot \mu) \ni (u, w) \mapsto w \left(\frac{p_1}{p_2}\right)^{\frac{1}{a}} - \frac{E_{p_2} \left[ w \left(\frac{p_1}{p_2}\right)^{\frac{1}{a}} \right]}{E_{p_2} \left[ \left(\frac{p_1}{p_2}\right)^{\frac{1}{a}} \right]} \in L_o^a(p_2 \cdot \mu). \tag{4.31}
\]

Its partial derivative \(D_1 \Phi_{p_2} \circ \Phi_{p_1}^{-1}\) at \(u \in s_{p_1}(U_{p_1} \cap U_{p_2})\) has principal part such that for each \((v, w) \in B_{p_1} \times L_o^a(p_1 \cdot \mu)\)

\[
D_1 \Phi_{p_2} \circ \Phi_{p_1}^{-1}(u)(v, w) = \frac{E_{p_2} \left[ w \left(\frac{p_1}{p_2}\right)^{\frac{1}{a}} \right]}{a E_{p_2} \left[ \left(\frac{p_1}{p_2}\right)^{\frac{1}{a}} \right]} \left[-v + \frac{E_{p_2} \left[ v \left(\frac{p_1}{p_2}\right)^{\frac{1}{a}} \right]}{E_{p_2} \left[ \left(\frac{p_1}{p_2}\right)^{\frac{1}{a}} \right]} \left(\frac{q}{p_2}\right)^{\frac{1}{a}} \right] \in L_o^a(p_2 \cdot \mu). \tag{4.32}
\]

54
Proposition 78. Given the atlas \( \{(U_p,s_p) : p \in \mathcal{M}_\mu\} \) of \( \mathcal{M}_\mu \), all the mappings \( \Gamma_p : \mathcal{V}_p \to L^2(B_p,L_0^a(p \cdot \mu);L_0^a(p \cdot \mu)) \) defined by

\[
\Gamma_p(u)(v,w) = -E_q\left[w\left(\frac{p}{q}\right)\left(\frac{q}{p}\right)\frac{v}{a}\right] + \frac{E_p\left[v\left(\frac{q}{p}\right)\right]E_q\left[w\left(\frac{q}{p}\right)\right]}{aE_p\left(\frac{q}{p}\right)} \left(\frac{q}{p}\right)^\frac{1}{a}
\]

where \( q = e_p(u) \) are a set of Christoffel symbols of the vector bundle \( \mathcal{F}^\alpha \).

Proof. If \( U_{p_1} \cap U_{p_2} \neq \emptyset \) for some \( p_1,p_2 \in \mathcal{M}_\mu \), using the change of coordinates formula (4.31) of \( \mathcal{F}^\alpha \) and its partial derivative (4.32) one can check that for each \( u \in s_{p_1}(U_{p_1} \cap U_{p_2}) \) and \( (v,w) \in B_{p_1} \times L_0^a(p_1 \cdot \mu) \)

\[
\Gamma^\alpha_{p_2}(s_{p_2} \circ e_{p_1}(u))\left(D(s_{p_2} \circ e_{p_1})(u) \cdot v,\Phi^\alpha_{p_2} \circ (\Phi^\alpha_{p_1})^{-1}(u,w)\right) = -D_1\left(\Phi^\alpha_{p_2} \circ (\Phi^\alpha_{p_1})^{-1}\right)(u,w) + \Phi^\alpha_{p_2} \circ (\Phi^\alpha_{p_1})^{-1}(u,\Gamma^\alpha_{p_1}(u)(v,w))
\]

Hence the locally defined continuous bilinear mappings \( \Gamma^\alpha_p \) glue together according to the change of variables formula (4.3) for Christoffel symbols.

Definition 79. \( (\mathcal{F}^\alpha,\nabla^\alpha) \) is the bundle-connection pair defined by the family \( \{\Gamma^\alpha_p\}_{p \in \mathcal{M}_\mu} \) of Christoffel symbols.

Proposition 80. For each density \( p \in \mathcal{M}_\mu \), \( L_0^a(p \cdot \mu) \) and \( T_{p^{1/a}}S^a \) are toplinearly isomorphic by the isometry

\[ I^\alpha_p : L_0^a(p \cdot \mu) \ni w \mapsto wp^{\frac{1}{a}} \in T_{p^{1/a}}S^a. \]

Proof. For each \( w \in L_0^a(p \cdot \mu) \), \( wp^{1/a} \in L^a(\mu) \) and

\[ \left\|wp^{\frac{1}{a}}\right\|_a^a = \int |wp^{1/a}|^a d\mu = \int |w|^a pd\mu = \|w\|_{p,a}^a. \]

Since

\[ \int \left(wp^{\frac{1}{a}}\right)^a d\mu = \int \left(wp^{\frac{1}{a}}\right)p^{\frac{1}{a}} d\mu = \mathbb{E}_p(w) = 0, \]

\( wp^{1/a} \in T_{p^{1/a}}S^a \) and \( I^\alpha_p \) is a norm-preserving linear mapping between \( L_0^a(p \cdot \mu) \) and \( T_{p^{1/a}}S^a \) with inverse \( u \mapsto up^{-1/a}. \)
Proposition 81. There exists a vector bundle morphism $I^\alpha$:

\[
\begin{array}{ccc}
\mathcal{F}^\alpha & \overset{I^\alpha}{\longrightarrow} & TS^a \\
\downarrow & & \downarrow \\
\mathcal{M}_\mu & \overset{A^\alpha}{\longrightarrow} & S^a
\end{array}
\] (4.34)

such that for each density $p \in \mathcal{M}_\mu$ the principal part of $I^\alpha$ coincides with

\[
I^\alpha_p : L^a_0 (p \cdot \mu) \ni w \mapsto \wp^\frac{1}{\beta} \in T_{p^1/a}S^a
\]

when restricted to the fiber $L^a_0 (p \cdot \mu)$ over $p$.

Proof. For each $p \in \mathcal{M}_\mu$, $I^\alpha$ is defined by $I^\alpha_p$ on the fiber $L^a_0 (p \cdot \mu)$ in such a way that diagram (4.34) is commutative. $I^\alpha$ has the local representation $(I^\alpha)^V_p = \left( A^\alpha_p, (I^\alpha)^V_p \right) := T \varphi_{p^1/a} \circ I^\alpha \circ (\Phi^\alpha_p)^{-1}$

\[
\begin{array}{ccc}
\mathcal{V}_p \times L^a_0 (p \cdot \mu) & \overset{(I^\alpha)^V_p}{\longrightarrow} & (T_{p^1/a}S^a)^2 \\
(\varphi^\alpha_p)^{-1} & \downarrow & \downarrow T \varphi_{p^1/a} \\
\bigcup_{q \in \mathcal{M}_\mu} L^a_0 (q \cdot \mu) & \overset{I^\alpha}{\longrightarrow} & TU_{p^1/a}
\end{array}
\]

such that

\[
\mathcal{V}_p \times L^a_0 (p \cdot \mu) \ni (u,w) \mapsto \left( \frac{q^\frac{1}{\beta}}{E_p \left( \frac{2}{p} \right)^{\frac{1}{\alpha}}} - \frac{wp^\frac{1}{\beta}}{E_p \left( \frac{2}{p} \right)^{\frac{1}{\alpha}}} - p^\frac{1}{\beta}, \wp^\frac{1}{\beta} \right) \in (T_{p^1/a}S^a)^2.
\]

Map $u \mapsto (I^\alpha)^V_p$ is a morphism of $\mathcal{V}_p$ into $\mathcal{L} \left( L^a_0 (p \cdot \mu); T_{p^1/a}S^a \right)$ then $I^\alpha$ is a vector bundle morphism (see Lang (1995, Proposition 1.3)).

\[\square\]

Theorem 82. For $\alpha > 0$, the bundle connection pair $(\mathcal{F}^\alpha, \nabla^\alpha)$ is the pull back of the natural connection in the tangent bundle $TS^a$ under the vector bundle morphism $I^\alpha$.

Proof. Let $H^\alpha : TF^\alpha \rightarrow \mathcal{F}^\alpha$ the connection in the vector bundle $\mathcal{F}^\alpha$ defined by the set $\{ \Gamma^\alpha_p \}$ of Christoffel symbols. A local representation $H^\alpha_p$ is given by:

\[
H^\alpha_p : \mathcal{V}_p \times L^a_0 (p \cdot \mu) \times B_p \times L^a_0 (p \cdot \mu) \rightarrow \mathcal{V}_p \times L^a_0 (p \cdot \mu) \\
(u,w_1,v,w_2) \mapsto (u,w_2 + \Gamma^\alpha_p (u) (v,w_1))
\]
4.8 An example in the parametric case

One can check, after a long calculation, that for each \((u,w_1,v,w_2) \in \mathcal{V}_p \times L_p^a (p \cdot \mu) \times B_p \times L_0^a (p \cdot \mu)\)

\[
H^\alpha_p (u,w_1,v,w_2) = (I^\alpha_p)^{-1} H^\alpha_{p^{1/\alpha}} \left(A^\alpha_p (u), (I^\alpha_p)_{\mathcal{V}_p^2} (u,w_1),
DA^\alpha_p (u) v, D (I^\alpha_p)_{\mathcal{V}_p^2} (u,w_1) (v,w_2) \right)
\]

where \(H^\alpha_{p^{1/\alpha}} : (T_{p^{1/\alpha}} S^a)^4 \rightarrow (T_{p^{1/\alpha}} S^a)^2\) is the local representative of the natural connection on the tangent bundle \(TS^a\) as in Proposition 74.

4.8 An example in the parametric case

We illustrate the \(\nabla^\alpha\)-connection in a particular parametric case.

Let \(X\) be a discrete set with finite cardinality 2 and \(\mu\) be the counting measure. \(\mathcal{M}_\mu\) can be identified with the 1-dimensional simplex

\[\mathcal{M}_\mu = \{q = (q_0,q_1) \in \mathbb{R}^2 : q_i > 0, q_0 + q_1 = 1\} .\]

Fix \(p \in \mathcal{M}_\mu\). In such a case the space of the random variables \(L^{\Phi_1} (p \cdot \mu)\) is \(\mathbb{R}^2\) and \(B_p\) is the linear subspace orthogonal to \(p\), that is

\[B_p = \{u = (u_0,u_1) \in \mathbb{R}^2 : u_0 p_0 + u_1 p_1 = 0\} .\]

Putting \(u_1 = \vartheta\), we have \(u_0 = -\vartheta p_1 / p_0\). \(\mathcal{M}_\mu\) is parametrized by:

\[e_p (\vartheta) := e^{u - K_p (u)} p = \begin{cases} q_0 (\vartheta) = \frac{e^{-\vartheta p_1 / p_0}}{e^{-\vartheta p_1 / p_0} p_0 + e^{\vartheta p_1}} \\
q_1 (\vartheta) = \frac{e^{\vartheta p_1}}{e^{-\vartheta p_1 / p_0} p_0 + e^{\vartheta p_1}} \end{cases}
\]

The log likelihood \(\ell\) in \(\vartheta\), that is the logarithm \(\ell := \log q\) of the density \(q\) regarded as a function of the parameter \(\vartheta\) can be written as

\[\ell (\vartheta) = \begin{cases} \ell_0 = -\vartheta p_1 / p_0 + \log p_0 - \log \left(e^{-\vartheta p_1 / p_0} p_0 + e^{\vartheta p_1}\right) \\
\ell_1 = \vartheta + \log p_1 - \log \left(e^{-\vartheta p_1 / p_0} p_0 + e^{\vartheta p_1}\right) \end{cases}
\]

57
The derivative of $\ell$ with respect to $\vartheta$ is

\[
\frac{\partial \ell}{\partial \vartheta} = \begin{cases}
\frac{\partial \ell_0}{\partial \vartheta} = -\frac{p_1}{p_0} - \frac{p_1}{p_0} \left( -e^{-\vartheta p_0 / p_1} + e^{\vartheta} \right) = -\frac{\vartheta}{e^{\vartheta / p_0} p_1} \\
\frac{\partial \ell_1}{\partial \vartheta} = 1 - \frac{p_1}{e^{-\vartheta p_0 / p_1}} \left( -e^{\vartheta} p_0 + e^{\vartheta} p_1 \right)
\end{cases}
\]

and the second derivative is

\[
\frac{\partial^2 \ell}{\partial \vartheta^2} = \left( -\frac{e^{\vartheta / p_0} p_1}{p_0 (p_0 + e^{\vartheta / p_0} p_1)^2}, -\frac{e^{\vartheta / p_0} p_1}{p_0 (p_0 + e^{\vartheta / p_0} p_1)^2} \right).
\]

The Fisher information matrix of $M_\mu$ at $q$ is reduced to

\[
g_{11}(\vartheta) := \mathbb{E}_q \left[ \left( \frac{\partial \ell}{\partial \vartheta} \right)^2 \right] = -\mathbb{E}_q \left[ \frac{\partial^2 \ell}{\partial \vartheta^2} \right] = \frac{\vartheta}{e^{\vartheta / p_0} p_1} p_0 \left( p_0 + e^{\vartheta / p_0} p_1 \right)^2.
\] (4.35)

For each real number $\alpha \in [-1,1]$, consider the function $\Gamma_{11,1}^{(\alpha)}(\vartheta)$ which maps each $\vartheta$ to the following value:

\[
\Gamma_{11,1}^{(\alpha)}(\vartheta) := \mathbb{E}_q \left[ \left( \frac{\partial^2 \ell}{\partial \vartheta^2} + \frac{1 - \alpha}{2} \left( \frac{\partial \ell}{\partial \vartheta} \right)^2 \right) \frac{\partial \ell}{\partial \vartheta} \right] = \frac{1 - \alpha}{2} \frac{e^{\vartheta / p_0} p_1}{p_0 \left( p_0 + e^{\vartheta / p_0} p_1 \right)^3}. \quad (4.36)
\]

The Amari $\alpha$-connection in the tangent bundle of $M_\mu$ is defined by:

\[
\left\langle \nabla^\alpha_{\frac{\partial}{\partial \vartheta}} \frac{\partial}{\partial \vartheta}, \frac{\partial}{\partial \vartheta} \right\rangle = \Gamma_{11,1}^{(\alpha)}
\]

where $\langle , \rangle$ is the Fisher metric (see, for example, Amari and Nagaoka (2000, Section 2.3)). Since

\[
\left\langle \nabla^\alpha_{\frac{\partial}{\partial \vartheta}} \frac{\partial}{\partial \vartheta}, \frac{\partial}{\partial \vartheta} \right\rangle = \Gamma_{11}^{11} \left\langle \frac{\partial}{\partial \vartheta}, \frac{\partial}{\partial \vartheta} \right\rangle = \Gamma_{11}^{11} g_{11},
\]

the Christoffel symbol $\Gamma_{11}^{11}$ of the $\alpha$-connection can be evaluated using Equations (4.35)-(4.36). With respect to the natural atlas of the tangent bundle induced
by the parameter \( \vartheta \) we have:

\[
\Gamma^{\alpha}_{11} (\vartheta) = \frac{p_0 - e^{\vartheta p_0} p_1}{ap_0 (p_0 + e^{\vartheta p_0} p_1)}
\]

where \( a = \frac{2}{1 - \alpha} \in [1, \infty] \).

Let us introduce the change of coordinates \( F^\alpha = (\text{id}, L^\alpha) \) of the tangent bundle of \( M_\mu \):

\[
\mathbb{R} \times T_p M_\mu \ni (\vartheta, \frac{\partial}{\partial \vartheta}) \mapsto (\vartheta, L^\alpha (\vartheta) \frac{\partial}{\partial \vartheta}) \in \mathbb{R} \times T_p M_\mu
\]

such that

\[
L^\alpha (\vartheta) := e^{\vartheta p_0} a \left( p_0 + e^{\vartheta p_0} p_1 \right)^\frac{1}{2} \left( p_0 + e^{\vartheta p_0} p_1 \right).
\]

The derivative of \( L^\alpha \) is:

\[
DL^\alpha (\vartheta) = \frac{e^{\vartheta p_0} \left( p_0^2 - e^{\frac{\vartheta}{a} p_0} p_1^2 \right)}{a^2 p_0 \left( p_0 + e^{\vartheta p_0} p_1 \right)^{\frac{1}{2}} \left( p_0 + e^{\vartheta p_0} p_1 \right)^{\frac{1}{2}}}.
\]

The Christoffel symbol \( \tilde{\Gamma}^\alpha \) of the Amari \( \alpha \)-connection with respect to these coordinates is related to \( \Gamma^\alpha \) by the change of variables formula (4.3):

\[
\tilde{\Gamma}^\alpha (\vartheta) L^\alpha (\vartheta) = -DL^\alpha (\vartheta) + L^\alpha (\vartheta) \Gamma^\alpha (\vartheta)
\]

and we have:

\[
\tilde{\Gamma}^\alpha (\vartheta) = \frac{p_1 \left( e^{\vartheta p_0} - e^{\vartheta p_0} \right)}{a \left( p_0 + e^{\vartheta p_0} p_1 \right) \left( p_0 + e^{\vartheta p_0} p_1 \right)}.
\]

(4.37)

**Proposition 83.** In the case of the 1-dimensional probability simplex \( M_\mu \), the connection \( \nabla^\alpha \) given in Definition 79 coincides (under a change of coordinates) with the Amari \( \alpha \)-connection.

**Proof.** Fix \( p = (p_0, p_1) \in M_\mu \). In this case the space of random variables \( L^\alpha (p \cdot \mu) \) is \( \mathbb{R}^2 \) and, like \( B_p, L^\alpha_0 (p \cdot \mu) \) is the linear subspace orthogonal to \( p \), so \( \nabla^\alpha \) is an affine connection.

Let us compute, with respect to the parameter \( \vartheta \), the Christoffel symbol of \( \nabla^\alpha \) according to the formula (4.33). For each \( v = (-v_1 p_1/p_0, v_1) \in B_p \simeq \mathbb{R} \) and \( w = (-w_1 p_1/p_0, w_1) \in L^\alpha_0 (p \cdot \mu) \simeq \mathbb{R} \) we have

\[
\mathbb{E}_p \left[ v \left( \frac{q}{p} \right)^\frac{1}{2} \right] = \frac{e^{\frac{\vartheta}{a} p_0} - 1}{\left( p_0 + e^{\vartheta p_0} p_1 \right)^\frac{1}{2}} p_1 v_1
\]

(4.38)
and
\[
\mathbb{E}_q \left[ w \left( \frac{p}{q} \right)^{\frac{1}{a}} \right] \left( \frac{q_1}{p_1} \right)^{\frac{1}{a}} = \frac{e^{\frac{\vartheta}{p_0}} - e^{\frac{\vartheta}{p_0} p_1}}{p_0 + e^{\frac{\vartheta}{p_0} p_1}} \frac{p_1 w_1}{(4.39)}
\]

moreover
\[
\mathbb{E}_p \left[ \left( \frac{q}{p} \right)^{\frac{1}{a}} \right] = \frac{p_0 + e^{\frac{\vartheta}{p_0} p_1}}{(p_0 + e^{\frac{\vartheta}{p_0} p_1})} \frac{p_1}{(4.40)}
\]

Taking into account Equation (4.38)-(4.39)-(4.40), the Christoffel symbol \( \Gamma^\alpha_p \) given by Equation (4.33) can be written as
\[
\Gamma^\alpha_p (\vartheta) = \frac{p_1 \left( e^{\frac{\vartheta}{p_0}} - e^{\frac{\vartheta}{p_0} p_1} \right)}{a \left( p_0 + e^{\frac{\vartheta}{p_0} p_1} \right) \left( p_0 + e^{\frac{\vartheta}{p_0} p_1} \right)} \left[ -1 + \frac{p_1 \left( e^{\frac{\vartheta}{p_0}} - 1 \right)}{p_0 + e^{\frac{\vartheta}{p_0} p_1} p_1} \right]
\]

and, comparing with Equation (4.37), \( \Gamma^\alpha_p (\vartheta) = \tilde{\Gamma}^\alpha (\vartheta) \).

Note that \( L^\alpha (\vartheta) \) is the composition of the derivative
\[
DA^\alpha_p (\vartheta) : B_p \rightarrow T_{p^{\rho_a}} S^a
\]
of the local representation of the Amari \( \alpha \)-embedding written in Equation (4.29) with the inverse of the linear map
\[
(I^\alpha)_{\eta_p} (\vartheta, \cdot) : L_0^a (p \cdot \mu) \rightarrow T_{p^{\rho_a}} S^a
\]

defined in Proposition 81.
Chapter 5

Finsler structure

We propose a family of Finsler structures of the exponential manifold $\mathcal{M}_\mu$. Our approach to Finsler manifolds refers to Bao et al. (2000).

First we present two preliminary theorems. In Bao et al. (2000, Paragraph 1.2) they are proved in the finite dimensional case $\mathbb{R}^n$. However their proofs can be easily adapted for a general Banach space.

**Theorem 84 (Euler’s Theorem).** Let $H$ be a real-valued function on a Banach space $E$, differentiable on $E \setminus \{0\}$. The following two statements are equivalent:

- $H$ is positively homogeneous of degree $r$. That is,
  \[ H(\lambda v) = \lambda^r H(v), \quad \forall \lambda > 0, \forall v \in E. \]

- The radial directional derivative of $H$ is $r$ times $H$. Namely,
  \[ DH(v) \cdot v = rH(v) \quad \forall v \in E \setminus \{0\}. \]

If $H$ is homogeneous of degree 1, taking the derivative of each side of the equation:
\[ DH(v) \cdot v = H(v), \]
we have $D^2H(v)(v, \cdot) + DH(v) = DH(v)$ and $D^2H(v)(v, \cdot) = 0$.

**Theorem 85.** Let $E$ be a Banach space and $F : E \to [0, \infty)$ be a $C^p$–morphism on $E \setminus \{0\}$ with $p \geq 2$. Following conditions:

- Positive homogeneity:
  \[ F(\lambda v) = \lambda F(v) \quad \forall \lambda > 0, \forall v \in E. \]
• Strong convexity: if \( L = \frac{1}{2} F^2 \), then
\[
D_2 L : E \setminus \{0\} \to \mathcal{L}_s^2(E)
\]
is positive-definite.

imply

**Positivity:** \( F(v) > 0 \) for all \( v \neq 0 \);

**Triangle Inequality:** \( F(v_1 + v_2) \leq F(v_1) + F(v_2) \) where equality holds if and only if \( y_2 = \alpha y_1 \) for some \( \alpha \geq 0 \);

**Fundamental Inequality:** for all \( w \in E \) and \( v \in E \setminus \{0\} \)
\[
DF(v) \cdot w \leq F(w)
\]
where equality holds iff \( w = \alpha v \) for some \( \alpha \geq 0 \).

We define a Finsler manifold.

**Definition 86.** Let \( M \) be a \( C^\infty \)-differentiable manifold, let \( TM \) be the tangent bundle of \( M \) and \( \pi : TM \to M \) the tangent bundle projection.

A Finsler structure \((M,F)\) of \( M \) is a function

\[
F : TM \to [0,\infty)
\]

with the following properties:

(i) **Regularity:** \( F \) is \( C^\infty \) on the entire slit tangent bundle
\[
TM \setminus 0 = \{(m,v) \in TM : v \in T_m M, v \neq 0\}.
\]

(ii) **Positive homogeneity:** \( F(m,\lambda v) = \lambda F(m,v) \) for all \( \lambda > 0 \), \( m \in M \) and \( v \in T_m M \).

(iii) **Strong convexity:** if \( L = \frac{1}{2} F^2 \), then for each local trivialization \( TU \simeq U \times E \) of \( TM \)
\[
D_2^2 L : U \times (E \setminus \{0\}) \to \mathcal{L}_s^2(E)
\]
is positive-definite.
Theorem 85 implies that if $F$ is also absolutely homogeneous (i.e. $\forall \lambda \in \mathbb{R}, F(m,\lambda v) = |\lambda| F(m,v)$), then it defines smoothly a norm on every fiber $T_m M$. In the finite dimensional case $\mathbb{R}^n$ such $F$ gives rise to the so called Minkowski norm.

Let $F_2,F_4 : T\mathcal{M}_\mu \to [0,\infty)$ be the pair of functions defined in local coordinates by

$$F_i(u,w) = \left( \int (w - \mathbb{E}_q(w))^i q d\mu \right)^{\frac{1}{i}} \quad i = 2,4$$

with $(u,w) \in \mathcal{V}_p \times B_p \simeq T\mathcal{V}_p$ and $q = e_p(u) = e^{u-K_p(u)}p$.

These functions are globally defined on $T\mathcal{M}_\mu$. In fact, if we suppose that $q \in \mathcal{U}_p \cap U_{p_1} \neq \emptyset$, then $q = e^{u_1-K_p(u_1)}p_1$ for $u_1 \in \mathcal{V}_{p_1}$.

The change of coordinates $Ts_{p_1} \circ Te_p : Ts_p(T\mathcal{U}_p \cap T\mathcal{U}_{p_1}) \to Ts_{p_1}(T\mathcal{U}_p \cap T\mathcal{U}_{p_1})$ has equation

$$(u,w) \mapsto \begin{cases} u_1 = u + \ln \frac{p}{p_1} - \mathbb{E}_{p_1} \left( u + \ln \frac{p}{p_1} \right) \\ w_1 = w - \mathbb{E}_{p_1}(w) \end{cases}$$

and we have

$$F_i^i(u_1,w_1) = \mathbb{E}_q \left[ (w_1 - \mathbb{E}_q(w_1))^i \right] = \mathbb{E}_q \left[ (w - \mathbb{E}_{p_1}(w) - \mathbb{E}_q(w - \mathbb{E}_{p_1}(w)))^i \right] = \mathbb{E}_q \left[ (w - \mathbb{E}_q(w))^i \right] = F_i^i(u,w)$$

**Proposition 87.** The functions $F_i : T\mathcal{M}_\mu \to [0,\infty)$, with $i = 2,4$, locally defined in (5.1) by

$$F_i(u,w) = \left( \int (w - \mathbb{E}_q(w))^i q d\mu \right)^{\frac{1}{i}}$$

are $C^\infty$-morphisms (infinitely Fréchet-differentiable) on $T\mathcal{M}_\mu \setminus 0$.

With respect to a local chart $(T\mathcal{U}_p,Ts_p)$, $F_2^2(u,w)$ is the second derivative of the cumulant generating functional $K_p$ at $u \in \mathcal{V}_p$ in the direction $w \in B_p$,

$$F_2^2(u,w) = \mathbb{E}_q \left[ (w - \mathbb{E}_q(w))^2 \right] = D^2 K_p(u) \cdot w^2$$

and $F_4^4(u,w)$ is a polynomial in the second and fourth derivatives of $K_p$ at $u \in \mathcal{V}_p$ in the direction $w \in B_p$,

$$F_4^4(u,w) = \mathbb{E}_q \left[ (w - \mathbb{E}_q(w))^4 \right] = D^4 K_p(u) \cdot w^4 + 3 (D^2 K_p(u) \cdot w^2)^2.$$
5 – Finsler structure

Proof. We have \( F_2^2(u,w) = \mathbb{E}_q \left[ (w - \mathbb{E}_q (w))^2 \right] = \mathbb{E}_q (w^2) - [\mathbb{E}_q (w)]^2 \). From

\[
\mathbb{E}_q (w^2) = \int w^2 e^{-\kappa_p(u)} pd\mu = \frac{D^2 M_p(u) \cdot w^2}{M_p(u)}
\]

and

\[
\mathbb{E}_q (w) = \int we^{-\kappa_p(u)} pd\mu = \frac{DM_p(u) \cdot w}{M_p(u)} = DK_p(u) \cdot w
\]

we obtain

\[
F_2^2(u,w) = \frac{D^2 M_p(u) \cdot w^2}{M_p(u)} - \left( \frac{DM_p(u) \cdot w}{M_p(u)} \right)^2 = D^2 K_p(u)(w,w).
\] (5.3)

With a similar calculation we get

\[
F_4^2(u,w) = \mathbb{E}_q \left[ (w - \mathbb{E}_q (w))^4 \right] = D^4 K_p(u) \cdot w^4 + 3 \left( D^2 K_p(u) \cdot w^2 \right)^2.
\] (5.6)

Since

\[
D^4 K_p(u) \cdot w^4 = \frac{D^4 M_p(u) \cdot w^4}{M_p(u)} - 3 \left( \frac{D^2 M_p(u) \cdot w^2}{M_p(u)} \right)^2 + 12 \left( \frac{D^2 M_p(u) \cdot w^2 (DM_p(u) \cdot w)^2}{(M_p(u))^3} \right)
\]

\[
- 6 \left( \frac{DM_p(u) \cdot w}{M_p(u)} \right)^4,
\]

comparing Equations (5.4) and (5.5) and recalling (5.3) we obtain:

\[
F_4^2(u,w) = D^4 K_p(u) \cdot w^4 + 3 \left( D^2 K_p(u) \cdot w^2 \right)^2.
\] (5.6)

Since the cumulant generating functional \( K_p \) is \( C^\infty \) in \( \mathcal{V}_p \), the two functions

\[
\mathcal{V}_p \times B_p \ni (u,w) \mapsto D^i K_p(u) \cdot w^i \in \mathbb{R} \quad i = 2, 4
\]
evaluating respectively the second and the fourth derivative of \( K_p \) at \( u \) in the direction \( w \) are \( C^\infty \). Hence, looking at Equations (5.3) and (5.6), we conclude that \( F_i^2(u,w) \) are \( C^\infty \) in \( \mathcal{V}_p \times B_p \) too and, globally, they are \( C^\infty \)-morphism on \( TM_p \). Instead \( F_2 \) and \( F_4 \) are smooth only on \( \mathcal{V}_p \times (B_p \setminus \{0\}) \) and, globally, they are smooth on \( TM_p \setminus 0 \). \( \square \)
In local coordinates the first partial derivatives of $F_2^2$ are:

\[
D_1 F_2^2 (\cdot, w) : V_p \rightarrow \mathcal{L}(B_p) \\
u \mapsto D^3 K_p(u)(w, w, \cdot)
\]

and

\[
D_2 F_2^2 (u, \cdot) : B_p \rightarrow \mathcal{L}(B_p) \\
w \mapsto 2D^2 K_p(u)(w, \cdot).
\]

The total first derivative $DF_2^2 : V_p \times B_p \rightarrow \mathcal{L}(B_p^2)$ at $(u, w)$ evaluated in $(w_1, w_2)$ is

\[
DF_2^2(u, w) \cdot (w_1, w_2) = D_1 F_2^2(u, w) \cdot w_1 + D_2 F_2^2(u, w) \cdot w_2 \\
= D^3 K_p(u)(w, w, w_1) + 2D^2 K_p(u)(w, w_2).
\]

Equation (5.2) shows that $E_q(w)$ is the derivative of $K_p$ at $u = s_p(q)$ in the direction $w$ hence mapping $V_p \times B_p \ni (u, w) \mapsto E_q(w) \in \mathbb{R}$, with $q = e^{u - K_p(u)} p$, is $C^\infty$. The total derivative at $(u, w)$ evaluated in $(w_1, w_2)$ is

\[
D E_q(w) \cdot (w_1, w_2) = D^2 K_p(u)(w, w_1) + DK_p(u) \cdot w_2 \\
= E_q(w w_1) - E_q(w) E_q(w_1) + E_q(w_2) \\
= \text{cov}_q(w, w_1) + E_q(w_2).
\]

For each pair of not negative real numbers $(a, b) \in \mathbb{R}^+ \times \mathbb{R}^+$, we consider the function $F_{(a, b)} : T\mathcal{M}_\mu \rightarrow [0, \infty)$ defined, in local coordinate, by

\[
F_{(a, b)}(u, w) = \left[ a \left( \int (w - E_q(w))^4 q d\mu \right)^{\frac{1}{2}} + b \int (w - E_q(w))^2 q d\mu \right]^{\frac{1}{2}} \\
= \left[ a \left( D^4 K_p(u) \cdot w^4 + 3 \left( D^2 K_p(u) \cdot w^2 \right)^2 \right)^{\frac{1}{2}} + b D^2 K_p(u) \cdot w^2 \right]^{\frac{1}{2}}
\]

with $(u, w) \in TV_p \simeq V_p \times B_p$ and $q = e_p(u)$.

**Theorem 88.** For each $(a, b) \in \mathbb{R}^+ \times \mathbb{R}^+$, $(a, b) \neq (0, 0)$, the map $F_{(a, b)} : T\mathcal{M}_\mu \rightarrow [0, \infty)$ define a Finsler structure $(\mathcal{M}_\mu, F_{(a, b)})$ of the exponential statistical manifold $\mathcal{M}_\mu$.

**Proof.** The functions $F_{(a, b)}$ satisfy the regularity condition on $T\mathcal{M}_\mu \setminus 0$ by Proposition 87 and clearly they satisfy the absolute homogeneity. We have to prove the strong convexity.
If we put \( L_{(a,b)} := \frac{1}{2} F_{(a,b)}^2 \), then
\[
D_2 L_{(a,b)} (u,w) \cdot w = a \left( D^4 K_p(u) (w^3, \cdot) + 3 D^2 K_p(u) \cdot w^2 D^2 K_p(u) (w, \cdot) \right) + b D^2 K_p(u) (w, \cdot)
\]
and, after a calculation, we obtain
\[
D_2 L_{(a,b)} (u,w) \cdot w_1 = a \frac{\mathbb{E}_q [(w - \mathbb{E}_q (w))^3 (w_1 - \mathbb{E}_q (w_1))]}{\left( \mathbb{E}_q [(w - \mathbb{E}_q (w))^4] \right)^{\frac{3}{2}}} + b \mathbb{E}_q [(w - \mathbb{E}_q (w))(w_1 - \mathbb{E}_q (w_1))].
\]
The second derivative of \( L_{(a,b)} \) is
\[
D^2_2 L_{(a,b)} (u,w) (w_1, w_2) = a \frac{3 \mathbb{E}_q [(w - \mathbb{E}_q (w))^4] \mathbb{E}_q [(w - \mathbb{E}_q (w))^2 (w_1 - \mathbb{E}_q (w_1))(w_2 - \mathbb{E}_q (w_2))]}{\left( \mathbb{E}_q [(w - \mathbb{E}_q (w))^4] \right)^{\frac{3}{2}}} - a \frac{2 \mathbb{E}_q [(w - \mathbb{E}_q (w))^3 (w_1 - \mathbb{E}_q (w_1)) \mathbb{E}_q [(w - \mathbb{E}_q (w))^3 (w_2 - \mathbb{E}_q (w_2))]}{\left( \mathbb{E}_q [(w - \mathbb{E}_q (w))^4] \right)^{\frac{3}{2}}} + b \mathbb{E}_q [(w_1 - \mathbb{E}_q (w_1))(w_2 - \mathbb{E}_q (w_2))].
\]
Let \( 0 \neq v \in B_p \),
\[
D^2_2 L_{(a,b)} (u,w) \cdot v^2 = a \frac{3 \mathbb{E}_q [(w - \mathbb{E}_q (w))^4] \mathbb{E}_q [(w - \mathbb{E}_q (w))^2 (v - \mathbb{E}_q (v))^2]}{\left( \mathbb{E}_q [(w - \mathbb{E}_q (w))^4] \right)^{\frac{3}{2}}} - a \frac{2 \mathbb{E}_q [(w - \mathbb{E}_q (w))^3 (v - \mathbb{E}_q (v))]}{\left( \mathbb{E}_q [(w - \mathbb{E}_q (w))^4] \right)^{\frac{3}{2}}} + b \mathbb{E}_q [(v - \mathbb{E}_q (v))^2].
\]
By the Hölder Inequality,
\[
2 \mathbb{E}_q [(w - \mathbb{E}_q (w))^3 (v - \mathbb{E}_q (v))]^2 \leq 3 \mathbb{E}_q [(w - \mathbb{E}_q (w))^4] \mathbb{E}_q [(w - \mathbb{E}_q (w))^2 (v - \mathbb{E}_q (v))^2]
\]
and, since
\[
\mathbb{E}_q [(v - \mathbb{E}_q (v))^2] = 0 \implies v \text{ constant in } B_p \implies v = 0,
\]
we see that \( D^2_2 L_{(a,b)} (u,w) \cdot v^2 \geq 0 \) so \( D^2_2 L_{(a,b)} (u,w) \) is positive definite for each \((u,w) \in \mathcal{U}_p \times (B_p \setminus \{0\})\) and \( F_{(a,b)} \) is strong convex. \( \square \)
Bibliography


