Dual Connections in Nonparametric Information Geometry

M. R. Grasselli

SharcNet Chair in Financial Mathematics
Department of Mathematics and Statistics
McMaster University

Supported by NSERC and MITACS

Ann Arbor, July 29, 2004

1. Program

- Given a probability space $(\Omega, \mathcal{F}, \mu)$, construct a Banach manifold $\mathcal{M}$ of all probability measures equivalent to μ.

- Extend the Fisher information

$$g_{ij} = \int \frac{\partial \log p(x,\theta)}{\partial \theta^i} \frac{\partial \log p(x,\theta)}{\partial \theta^j} p(x,\theta) dx \qquad (1)$$

  to a well defined scalar product on $T_p\mathcal{M}$ and *prove Chentsov's theorem*.

- Obtain the infinite dimensional analogues for the exponential and mixture connections acting on the tangent bundle $T\mathcal{M}$ and establish their Amari duality with respect to the Fisher scalar product.

- Define the infinite dimensional α-connections and prove that

$$\nabla^{(\alpha)} = \frac{1+\alpha}{2}\nabla^{(e)} + \frac{1-\alpha}{2}\nabla^{(m)}. \tag{2}$$

- *Define statistical divergences and prove projection/minimization theorems.*

2. Wandering in Orlicz Spaces

Consider Young functions of the form

$$\Phi(x) = \int_0^{|x|} \phi(t)dt, \quad x \geq 0, \tag{3}$$

where $\phi : [0, \infty) \mapsto [0, \infty)$ is nondecreasing, continuous and such that $\phi(0) = 0$ and $\lim_{x\to\infty} \phi(x) = +\infty$. This include the monomials $|x|^r/r$, for $1 < r < \infty$, as well as the following examples:

$$\Phi_1(x) = \cosh x - 1, \tag{4}$$

$$\Phi_2(x) = e^{|x|} - |x| - 1, \tag{5}$$

$$\Phi_3(x) = (1 + |x|) \log(1 + |x|) - |x| \tag{6}$$

The complementary of a Young function Φ of the form (3) is given by

$$\Psi(y) = \int_0^{|y|} \psi(t)dt, \quad y \geq 0, \tag{7}$$

where ψ is the inverse of ϕ. One can verify that (Φ_2, Φ_3) and $(|x|^r/r, |x|^s/s)$, with $r^{-1} + s^{-1} = 1$, are examples of complementary pairs.

We say that $\Psi_1 \prec \Psi_2$ (Ψ_1 is weaker than Ψ_2), if there exist a constant $a > 0$ such that

$$\Psi_1(x) \leq \Psi_2(ax), \quad x \geq x_0, \tag{8}$$

for some $x_0 \geq 0$ (depending on a). For example,

$$|x| \prec \Phi_3 \prec \frac{|x|^r}{r} \prec \frac{|x|^s}{s} \prec \Phi_2 \tag{9}$$

whenever $1 < r \leq s < \infty$.

Two Young functions Ψ_1 and Ψ_2 are equivalent if $\Psi_1 \prec \Psi_2$ and $\Psi_2 \prec \Psi_1$. For example, the functions Φ_1 and Φ_2 are equivalent, both being of exponential type.

A Young function $\Phi : \mathbf{R} \mapsto \mathbf{R}^+$ satisfies the Δ_2-condition if

$$\Phi(2x) \leq K\Phi(x), \quad x \geq x_0 \geq 0, \tag{10}$$

for some constant $K > 0$. Examples of functions in this class are the monomials $|x|^r/r, r \geq 1$ and the function Φ_3.

Now let (Ω, Σ, P) be a probability space. The Orlicz space associated with a Young function Φ defined as

$$L^{\Phi}(P) = \left\{ f : \Omega \mapsto \overline{\mathbf{R}}, \text{measurable} : \int_{\Omega} \Phi(\alpha f) dP < \infty, \text{ for some } \alpha > 0 \right\}. \tag{11}$$

If we identify functions which differ only on sets of measure zero, then L^{Φ} is a Banach space when furnished with the Luxembourg norm

$$N_{\Phi}(f) = \inf \left\{ k > 0 : \int_{\Omega} \Phi(\frac{f}{k}) dP \leq 1 \right\}, \tag{12}$$

or with the equivalent Orlicz norm

$$\|f\|_{\Phi} = \sup \left\{ \int_{\Omega} |fg| d\mu : g \in L^{\Psi}(\mu), \int_{\Omega} \Psi(g) dP \leq 1 \right\}, \tag{13}$$

where Ψ is the complementary Young function to Ψ.

If Φ and Ψ are complementary Young functions, $f \in L^{\Phi}(P)$, $g \in L^{\Psi}(P)$, then we have the generalized Hölder inequality:

$$\int_{\Omega} |fg| dP \leq 2 N_{\Phi}(f) N_{\Psi}(g). \tag{14}$$

It follows that $L^{\Phi} \subset \left(L^{\Psi}\right)^{*}$ for any pair of complementary Young functions.

If $\Psi_2 \prec \Psi_1$ then there exist a constant k such that $\|\cdot\|_{\Psi_2} \leq k \|\cdot\|_{\Psi_1}$ and therefore $L^{\Psi_1}(P) \subset L^{\Psi_2}(P)$.

If two Young functions are equivalent, the Banach spaces associated with them coincide as sets and have equivalent norms.

Now define

$$M^{\Phi}(P) = \left\{ f \in L^{\Phi} \, : \, \int_{\Omega} \Phi(kf) dP < \infty, \text{ for all } k > 0 \right\}. \qquad (15)$$

Lemma 1 *Let* (Φ, Ψ) *be a complementary pair of Young functions,* Φ *continuous,* $\Phi(x) = 0$ *iff* $x = 0$*. Then:*

1. $M^{\Phi}(P)$ *is the closure of* $L^{\infty}(\Omega, \Sigma, P)$ *in the* $L^{\Phi}(P)$*–norm.*

2. $\left(M^{\Phi}(P)\right)^{*}$ *is isometrically isomorphic to* $L^{\Psi}(P)$*.*

If, moreover, Φ *satisfies the* Δ_2*-condition, then* $M^{\Phi}(P) = L^{\Phi}(P)$*.*

3. The Pistone-Sempi Manifold

Consider the set

$$\mathcal{M} \equiv \mathcal{M}(\Omega, \Sigma, \mu) = \{f : \Omega \mapsto \mathbf{R}, f > 0 \text{ a.e. and } \int_\Omega f d\mu = 1\}.$$

For each point $p \in \mathcal{M}$, let $L^{\Phi_1}(p)$ be the exponential Orlicz space over the probability space $(\Sigma, \Omega, pd\mu)$ and consider its closed subspace of p-centred random variables

$$B_p = \{u \in L^{\Phi_1}(p) : \int_\Omega upd\mu = 0\} \tag{16}$$

as the coordinate Banach space.

In probabilistic terms, the set $L^{\Phi_1}(p)$ correspond to random variables whose moment generating function with respect to the probability $pd\mu$ is finite on a neighborhood of the origin.

They define one dimensional exponential models $p(t)$ associated with a point $p \in \mathcal{M}$ and a random variable u:

$$p(t) = \frac{e^{tu}}{Z_p(tu)} p, \qquad t \in (-\varepsilon, \varepsilon). \tag{17}$$

Define the inverse of a local chart around $p \in \mathcal{M}$ as

$$\begin{aligned} e_p : \mathcal{V}_p &\rightarrow \mathcal{M} \\ u &\mapsto \frac{e^u}{Z_p(u)} p. \end{aligned} \tag{18}$$

Denote by $\mathcal{U}_p$ the image of $\mathcal{V}_p$ under e_p. Let e_p^{-1} be the inverse of e_p on $\mathcal{U}_p$. Then a local chart around p is given by

$$\begin{aligned} e_p^{-1} : \mathcal{U}_p &\rightarrow B_p \\ q &\mapsto \log\left(\frac{q}{p}\right) - \int_\Omega \log\left(\frac{q}{p}\right) p d\mu. \end{aligned} \tag{19}$$

For any $p_1, p_2 \in \mathcal{M}$, the transition functions are given by

$$
\begin{aligned}
e_{p_2}^{-1} e_{p_1} : e_{p_1}^{-1}(\mathcal{U}_{p_1} \cap \mathcal{U}_{p_2}) &\rightarrow e_{p_2}^{-1}(\mathcal{U}_{p_1} \cap \mathcal{U}_{p_2}) \\
u &\mapsto u + \log\left(\frac{p_1}{p_2}\right) - \int_\Omega \left(u + \log \frac{p_1}{p_2}\right) p_2 d\mu. \qquad (20)
\end{aligned}
$$

Proposition 2 *For any $p_1, p_2 \in \mathcal{M}$, the set $e_{p_1}^{-1}(\mathcal{U}_{p_1} \cap \mathcal{U}_{p_2})$ is open in the topology of B_{p_1}.*

We then have that the collection $\{(\mathcal{U}_p, e_p^{-1}), p \in \mathcal{M}\}$ satisfies the three axioms for being a C^∞–atlas for $\mathcal{M}$. Moreover, since all the spaces B_p are isomorphic as topological vector spaces, we can say that $\mathcal{M}$ is a C^∞–manifold modeled on $B_p \equiv T_p\mathcal{M}$.

Given a point $p \in \mathcal{M}$, the connected component of $\mathcal{M}$ containing p coincides with the maximal exponential model obtained from p: $\mathcal{E}(p) = \left\{ \frac{e^u}{Z_p(u)} p, u \in B_p \right\}$.

4. The Fisher Information and Dual Connections

Let $\langle \cdot, \cdot \rangle_p$ be a continuous positive definite symmetric bilinear form assigned continuously to each $B_p \simeq T_p\mathcal{M}$. A pair of connection (∇, ∇^*) are said to be dual with respect to $\langle \cdot, \cdot \rangle_p$ if

$$\langle \tau u, \tau^* v \rangle_q = \langle u, v \rangle_p \tag{21}$$

for all $u, v \in T_p\mathcal{M}$ and all smooth curves $\gamma : [0, 1] \to \mathcal{M}$ such that $\gamma(0) = p$,$\gamma(1) = q$, where τ and τ^* denote the parallel transports associated with ∇ and ∇^*, respectively.

Equivalently, (∇, ∇^*) are dual with respect to $\langle \cdot, \cdot \rangle_p$ if

$$v\left(\langle s_1, s_2 \rangle_p\right) = \langle \nabla_v s_1, s_2 \rangle_p + \langle s_1, \nabla^*_v s_2 \rangle_p \tag{22}$$

for all $v \in T_p\mathcal{M}$ and all smooth vector fields s_1 and s_2.

The infinite dimensional generalisation of the Fisher information is given by

$$\langle u, v\rangle_p = \int_\Omega (uv) p d\mu, \quad \forall u, v \in B_p. \tag{23}$$

This is clearly bilinear, symmetric and positive definite. Moreover, continuity follows from that fact that, since $L^{\Phi_1}(p) \simeq L^{\Phi_2}(p) \subset L^{\Phi_3}(p)$, the generalised Hölder inequality gives

$$|\langle u, v\rangle_p| \leq K \|u\|_{\Phi_1,p} \|v\|_{\Phi_1,p}, \quad \forall u, v \in B_p. \tag{24}$$

If p and q are two points on the same connected component of $\mathcal{M}$, then the exponential parallel transport is given by

$$\begin{aligned} \tau_{pq}^{(1)} : T_p\mathcal{M} &\rightarrow T_q\mathcal{M} \\ u &\mapsto u - \int_\Omega uq d\mu. \end{aligned} \tag{25}$$

To obtain duality with respect to the Fisher information, we define the mixture parallel transport on $T\mathcal{M}$ as

$$\begin{aligned} \tau_{pq}^{(-1)} : T_p\mathcal{M} &\rightarrow T_q\mathcal{M} \\ u &\mapsto \frac{p}{q}u, \end{aligned} \tag{26}$$

for p and q in the same connected component of $\mathcal{M}$.

Proposition 3 *Let p and q be two points in the same connected component of $\mathcal{M}$. Then $\frac{p}{q}u \in B_q$, for all $u \in B_p$.*

Theorem 4 *The connections $\nabla^{(1)}$ and $\nabla^{(-1)}$ are dual with respect to the Fisher information.*

Proof: We have that

$$
\begin{aligned}
\langle \tau^{(1)} u, \tau^{(-1)} v \rangle_q &= \left\langle u - \int_\Omega uq d\mu, \frac{p}{q} v \right\rangle_q \\
&= \int_\Omega u \frac{p}{q} v q d\mu - \left(\int_\Omega uq d\mu \right) \int_\Omega \frac{p}{q} v q d\mu \\
&= \int_\Omega uvp d\mu \\
&= \langle u, v \rangle_p, \quad \forall u, v \in B_p,
\end{aligned}
$$

5. α–connections

We begin with Amari's α-embeddings

$$\begin{aligned} \ell_\alpha : \mathcal{M} &\rightarrow L^r(\mu) \\ p &\mapsto \frac{2}{1-\alpha} p^{\frac{1-\alpha}{2}}, \quad \alpha \in (-1,1), \end{aligned} \tag{27}$$

where $r = \frac{2}{1-\alpha}$. Observe that $\ell_\alpha(p) \in S^r(\mu)$, the sphere of radius r in $L^r(\mu)$.

Using the chain rule, the push-forward of the map ℓ_α can be implemented as

$$\begin{aligned} (\ell_\alpha)_{*(p)} : T_p\mathcal{M} = B_p &\rightarrow T_{rp^{1/r}} S^r(\mu) \\ u &\mapsto p^{\frac{1-\alpha}{2}} u, \end{aligned} \tag{28}$$

observing that $p^{\frac{1-\alpha}{2}} u$ is indeed an element of $T_{rp^{1/r}} S^r(\mu)$.

The tangent space to $S^r(\mu)$ at $rp^{1/r}$ is

$$T_{rp^{1/r}}S^r(\mu) = \left\{ g \in L^r(\mu) : \int_\Omega g p^{1-1/r} d\mu = 0 \right\}. \tag{29}$$

For each $f \in S^r(\mu)$, a canonical projection from the tangent space $T_{rp^{1/r}}L^r(\mu)$ onto the tangent space $T_{rp^{1/r}}S^r(\mu)$ can be uniquely defined by

$$\begin{aligned} \Pi_{rp^{1/r}} : T_{rp^{1/r}}L^r(\mu) & \longrightarrow T_{rp^{1/r}}S^r(\mu) \\ g & \longmapsto g - \left(\int_\Omega g p^{1-1/r} d\mu \right) p^{1/r}. \end{aligned} \tag{30}$$

We are now ready to define the α-connections. In what follows, $\widetilde{\nabla}$ is used to denote the trivial connection on $L^r(\mu)$.

Definition 5 *For* $\alpha \in (-1, 1)$, *let* $\gamma : (-\varepsilon, \varepsilon) \to \mathcal{M}$ *be a smooth curve such that* $p = \gamma(0)$ *and* $v = \dot{\gamma}(0)$ *and let* $s \in S(T\mathcal{M})$ *be a differentiable vector field. The* α*-connection on* $T\mathcal{M}$ *is given by*

$$\left(\nabla_v^\alpha s\right)(p) = (\ell_\alpha)^{-1}_{*(p)} \left[\Pi_{rp^{1/r}} \widetilde{\nabla}_{(\ell_\alpha)_{*(p)} v} (\ell_\alpha)_{*(\gamma(t))} s \right]. \tag{31}$$

Theorem 6 *The exponential, mixture and* α*-covariant derivatives on* $T\mathcal{M}$ *satisfy*

$$\nabla^\alpha = \frac{1+\alpha}{2} \nabla^{(1)} + \frac{1-\alpha}{2} \nabla^{(-1)}. \tag{32}$$

Corollary 7 *The connections* ∇^α *and* $\nabla^{-\alpha}$ *are dual with respect to the Fisher information* $\langle \cdot, \cdot \rangle_p$.