α-connections, α-divergence and duality

Salem SAID - 17 June 2011

1 The manifold $P(n)$

Let $\mathbb{R}_+^{n+1}$ be the set of $p = (p_0, \ldots, p_{n+1})$ such that $p_0, \ldots, p_n > 0$. Consider the set $P(n)$

$$P(n) = \left\{ p \in \mathbb{R}_+^{n+1} \left| \sum_{i=0}^{n} p_i = 1 \right. \right\}$$

(1)

It is straightforward that $P(n)$ is a submanifold of $\mathbb{R}_+^{n+1}$, with dimension equal to $n$. For $\alpha \in \mathbb{R}$ and $i = 0, \ldots, n$, consider the functions on $P(n)$

$$x_i^\alpha(p) = \begin{cases} 
\frac{2}{1-\alpha} p_i^{\frac{1-\alpha}{2}} & \text{(for } \alpha \neq 1) \\
\log(p_i) & \text{(for } \alpha = 1) 
\end{cases}$$

(2)

Further functions to be used are

$$\pi_i(p) = p_i$$

(3)

For $p \in P(n)$ and $X \in T_p P(n)$, write

$$X_i^\alpha = X x_i^\alpha \quad X_\alpha = (X_0^\alpha, \ldots, X_n^\alpha)$$

(4)

In the following, $\langle \cdot, \cdot \rangle$ is the Euclidean scalar product on $\mathbb{R}_+^{n+1}$.

**Proposition 1** For $\alpha, \beta \in \mathbb{R}$, $p \in P(n)$ and $X \in T_p P(n)$

$$\langle X_\alpha, p^{\frac{1+\alpha}{2}} \rangle = 0 \quad X_i^\alpha = p_i^{\frac{\beta-\alpha}{2}} X_\beta^i$$

**Proof:** The first identity follows from (4), along with

$$\sum_{i=0}^{n} X \pi_i = 0$$

For example, if $\alpha \neq 1$, it is enough to use

$$\pi_i = \left( \frac{1-\alpha}{2} x_i^\alpha \right)^{\frac{2}{1-\alpha}}$$

which follows from (2). The second identity follows in a similar way. ▲
2 \(\alpha\)-connections

The Fisher metric is a Riemannian metric on \(P(n)\). At \(p \in P(n)\), it is usually introduced as

\[
(X, Y) = \langle X_0, Y_0 \rangle = \langle X_\alpha, Y_{-\alpha} \rangle
\]

for \(X, Y \in T_pP(n)\). That these expressions are equal follows from proposition 1. In particular,

\[
X^i_\alpha = p^{\frac{\alpha}{2}} X^i_0 \quad Y^i_{-\alpha} = p^{\frac{\alpha}{2}} Y^i_0
\]

For \(\alpha \in \mathbb{R}\) consider \(T^\alpha_p\), the subspace of \(\mathbb{R}^{n+1}\)

\[
T^\alpha_p = \left\{ A \in \mathbb{R}^{n+1} \mid \langle A, p^{\frac{\alpha}{2}} \rangle = 0 \right\}
\]

For \(X \in T_pP(n)\) let \(i_\alpha(X) = X_\alpha\). It can be shown using proposition 1 that \(i_\alpha\) is an isomorphism of \(T_pP(n)\) and \(T^\alpha_p\). In particular, this confirms (5) is a scalar product on \(T^\alpha_p\).

The \(\alpha\)-connection \(\nabla^\alpha\) is a connection defined on \(TP(n)\). For vector fields \(X, Y\) on \(P(n)\), it is defined for \(p \in P(n)\) by

\[
i_\alpha(\nabla^\alpha_X Y|_p) = X_p Y_\alpha - \langle X_p Y_\alpha, p^{\frac{1+\alpha}{2}} \rangle p^{\frac{1-\alpha}{2}}
\]

Here,

\[
X_p Y_\alpha = (X_p Y^0_\alpha, \ldots, X_p Y^n_\alpha)
\]

Proposition 2 states the duality of \(\nabla^\alpha\) and \(\nabla^{-\alpha}\). First, it is checked (7) is well defined. Precisely, that the right hand side belongs to \(T^\alpha_p\). This follows since

\[
\left\langle X Y_\alpha - \langle XY_\alpha, p^{\frac{1+\alpha}{2}} \rangle p^{\frac{1-\alpha}{2}}, p^{\frac{1-\alpha}{2}} \right\rangle = \langle XY_\alpha, p^{\frac{1+\alpha}{2}} \rangle - \langle XY_\alpha, p^{\frac{1+\alpha}{2}} \rangle \langle p^{\frac{1-\alpha}{2}}, p^{\frac{1-\alpha}{2}} \rangle
\]

and \(\langle p^{\frac{1-\alpha}{2}}, p^{\frac{1+\alpha}{2}} \rangle = 1\).

**Proposition 2** Let \(\alpha \in \mathbb{R}\). For vector fields \(Z, X, Y\) on \(P(n)\),

\[
Z(X, Y) = \langle \nabla^\alpha_Z X, Y \rangle + \langle X, \nabla^{-\alpha}_Z Y \rangle
\]

In particular, \(\nabla^0\) is the Levi-Civita connection of the Fisher metric.

**Proof:** From (5) and (7)

\[
(\nabla^\alpha_Z X, Y) = \langle ZX_\alpha, Y_{-\alpha} \rangle - \langle XY_\alpha, \pi^{\frac{1+\alpha}{2}} \rangle \langle \pi^{\frac{1-\alpha}{2}}, Y_{-\alpha} \rangle
\]

Here, \(\pi = (\pi_0, \ldots, \pi_n)\). Substituting proposition 1,

\[
(\nabla^\alpha_Z X, Y) = \langle ZX_\alpha, Y_{-\alpha} \rangle
\]

By an identical reasoning,

\[
(\langle X, \nabla^{-\alpha}_Z Y \rangle = \langle X_\alpha, Z Y_{-\alpha} \rangle
\]

The proposition follows. \(\blacksquare\)
3 Coordinate expression

Here, the connection $\nabla^\alpha$ is expressed in an arbitrary coordinate system. There is little loss of generality in considering only global coordinates on $P(n)$. Let $(\xi^i)$ where $i = 1, \ldots, n$ be such coordinates and $\partial_i = \partial/\partial \xi^i$.

For $\alpha \in \mathbb{R}$, let $x_\alpha = (x^0_\alpha, \ldots, x^n_\alpha)$. By (5), the metric tensor is given by

$$ (\partial_i, \partial_j) = (\partial x_\alpha, \partial x_{-\alpha}) $$

The connection $\nabla^\alpha$ is given by the following formula

$$ (\nabla^\alpha_{\partial_i} \partial_j, \partial_k) = (\partial_j \partial_i x_\alpha, \partial_k x_{-\alpha}) $$

To prove this, apply proposition 1 to

$$ i_\alpha (\nabla^\alpha_{\partial_i} \partial_i) = \partial_j \partial_i x_\alpha - (\partial_j \partial_i x_\alpha, p^{\frac{1+\alpha}{2}}) p^{\frac{1-\alpha}{2}} $$

This calculation has already been used for proposition 2. Direct computation from (9) gives other usual formulae. In particular, it is usual to put $x^i_1 = l_i$ and $l = (l_0, \ldots, l_n)$. Then, for $\alpha \in \mathbb{R},$

$$ (\nabla^\alpha_{\partial_i} \partial_i, \partial_k) = \left\langle \pi, \left( \partial_j \partial_i l + \frac{1-\alpha}{2} \partial_i \partial_j l \right) \partial_k l \right\rangle $$

This gives the following important formula$^1$,

$$ \nabla^\alpha = \frac{1+\alpha}{2} \nabla^1 + \frac{1-\alpha}{2} \nabla^{-1} $$

4 $\alpha$-divergence

Proposition 2 established $\nabla^\alpha$ and $\nabla^{-\alpha}$ are dual, for $\alpha \in \mathbb{R}$. Here, corresponding divergence functions $D^\alpha : P(n) \times P(n) \to \mathbb{R}_+$ are given. The divergence function for a pair of dual connections is not uniquely defined. However, the functions $D^\alpha$ are a popular choice. For $\alpha \neq \pm 1,$

$$ D^\alpha(p, q) = \frac{4}{1 - \alpha^2} \left[ 1 - \left\langle p^{\frac{1+\alpha}{2}}, q^{\frac{1+\alpha}{2}} \right\rangle \right] = D^{-\alpha}(q, p) $$

For $\alpha = \pm 1$,

$$ D^{-1}(p, q) = \langle p, \log(p) - \log(q) \rangle = D^1(q, p) $$

For $\alpha \neq 1$, it follows from (12) and (13)

$$ D^\alpha(p, q) = \langle x_\alpha(p), x_{-\alpha}(p) - x_{-\alpha}(q) \rangle $$

Note finally $(1/2)D^{0}(p, q)$ is the Hellinger distance.

$$ (1/2)D^{0}(p, q) = \langle \sqrt{p} - \sqrt{q}, \sqrt{p} - \sqrt{q} \rangle $$

The required properties of $D^\alpha$ are proved in proposition 3. For vector field $X$ on $P(n)$, $XD^\alpha(p, q)$ denotes application of $X$ to the first variable; $X'D^\alpha(p, q)$ denotes application of $X$ to the second variable.

---

$^1$Recall an affine combination of connections is again a connection.
Proposition 3  Let $D^{\alpha}(p, q)$ be given by (12) and (13). Then, $D^{\alpha}(p, q) \geq 0$ with $D^{\alpha}(p, q) = 0$ iff $p = q$. Moreover, for vector fields $X, Y, Z$ on $P(n)$

\[ XD^{\alpha}(p, p) = 0 \quad X' D^{\alpha}(p, p) = 0 \quad -Y' X D^{\alpha}(p, p) = (X_p, Y_p) \quad (15) \]

\[ -XY Z' D^{\alpha}(p, p) = (\nabla^{\alpha}_X Y|_p, Z_p) \quad (16) \]

Proof:  For $\alpha \neq \pm 1$, since \( 1 + \frac{\alpha}{2} + \frac{1 - \alpha}{2} = 1 \), it follows from Hölder inequality

\[ \sum_{i=0}^{n} p_i^{\frac{1+\alpha}{2}} q_i^{\frac{1-\alpha}{2}} \leq \left( \sum_{i=0}^{n} p_i \right)^{\frac{1+\alpha}{2}} \left( \sum_{i=0}^{n} p_i \right)^{\frac{1-\alpha}{2}} \]

where the equality only holds when $p = q$. By strict convexity of the function $-\log$,

\[ \sum_{i=0}^{n} p_i \log(p_i/q_i) \geq -\log \left( \sum_{i=1}^{n} q_i \right) = 0 \]

with equality only for $p = q$. Thus, the first part of the proposition is proven.

Relations (15) and (16) follow from (14) by direct calculation. For the first two relations in (15), note $XD^{\alpha}(p, q) = X' D^{-\alpha}(q, p)$, which follows from $D^{\alpha}(p, q) = D^{-\alpha}(q, p)$.

For (15), it is enough to write

\[ XD^{\alpha}(p, q) = \langle x_\alpha(p), x_{-\alpha}(p) - x_{-\alpha}(q) \rangle - \langle x_\alpha(p), X_{-\alpha}(q) \rangle \]

It follows from proposition 1 that $XD^{\alpha}(p, p) = 0$. Carrying on the calculation, it follows

\[ -Y' X D^{\alpha}(p, p) = \langle X_\alpha(p), Y_{-\alpha}(p) \rangle = (X_p, Y_p) \]

To obtain (16), note as for (9)

\[ (\nabla^{\alpha}_X Y|_p, Z_p) = \langle XY x_\alpha(p), Z x_{-\alpha}(p) \rangle = -XY Z' D^{\alpha}(p, p) \]

This completes the proof.\[\]

Finally, let us quote the so called Pythagorian relation, in the form of a proposition.

Proposition 4  For $p, q, r \in P(n)$ and $\alpha \in \mathbb{R}$

\[ D^{\alpha}(p, q) + D^{\alpha}(q, r) - D^{\alpha}(p, r) = \langle x_\alpha(p) - x_\alpha(q), x_{-\alpha}(r) - x_{-\alpha}(q) \rangle \]

Proof:  Assume $\alpha \neq 1$, so that (14) holds. The left hand side is readily simplified to obtain the formula. For $\alpha = 1$, it becomes possible to use $D^{1}(p, q) = D^{-1}(q, p)$.\[\]