DUAL SPACES AND β-ORTHOGEODESIC MODELS.

CARSTEN WIUF

Department of Statistics
Institute of Mathematics
University of Aarhus
Denmark

Introduction.

In recent years geometric considerations and geometry have been used in many contexts within statistics.

Statistical models admitting parameterizations fulfilling certain geometric constrains have been studied to some extend, see for instance Amari (1985) and Barndorff-Nielsen and Bl@esild (1993), and especially in context to exponential families these models have great importance.

It might therefore be of interest to study generalizations from a geometric point of view both of exponential families and of subfamilies of exponential families. This is done in for example Barndorff-Nielsen and Bl@esild (1993).

The purpose of this article is to discuss such generalizations and especially to discuss a geometric counterpart to exponential families. This geometric counterpart includes many statistical models of interest.

The paper is divided into two parts. The first part is a discussion of a geometric generalization of the well known duality between the mean value parameter and the canonical parameter of an exponential family. A similar discussion can be found in Amari (1985). However the discussion here is extended to cover manifolds from a global point of view in contrast to Amari (1985), where only local properties are considered.

Moreover it is shown that the duality property can be formulated in three different ways: existence of a special pair of parameters, a special yoke, or a pair dual torsion-free flat connections.

The last part extend and discuss a concept called 'orthogeodesic model' introduced by Barndorff-Nielsen and Bl@esild (1993). The extended concept is called 'β-orthogeodesic' and its relation to the above mentioned generalization of exponential families is elucidated through a structure theorem for β-orthogeodesic models.

Moreover an incorrect proposition in Barndorff-Nielsen and Bl@esild (1993) postulating the equivalence of two different definitions of the othogeodesic property is corrected and a simple counterexample to the incorrect proposition is given.

Preliminaries.

Only little differential geometrical background and notation will be given here. Special notation in context to dual spaces will be introduced and explained when necessary. For a more detailed discussion see for instance Amari (1985) and Lauritzen (1987).

A Riemannian manifold is a couple $(\mathcal{M}, i)$, where $\mathcal{M}$ is a differential manifold and i a positive definite (0,2)-tensor. A statistical manifold is a triple $(\mathcal{M}, i, T)$, where $(\mathcal{M}, i)$ is a Riemannian manifold and T a symmetric (0,3)-tensor on $\mathcal{M}$.

Let $\{\Omega_\varepsilon\}_{\varepsilon \in E}$ be an atlas of $\mathcal{M}$, and let ω_ε be a point in Ω_ε. Generic components of $\omega_\varepsilon = (\omega_\varepsilon^1, ..., \omega_\varepsilon^d)$ will be denoted by r, s, t etc, and the coordinate frame $\partial/\partial\omega_\varepsilon^r$ will be denoted ∂_r. We consider ω_ε as a function from $\omega_\varepsilon^{-1}(\Omega_\varepsilon) \subseteq \mathcal{M}$ into $\Omega_\varepsilon \subseteq R^d$, i.e. $\omega_\varepsilon : \omega_\varepsilon^{-1}(\Omega_\varepsilon) \subseteq \mathcal{M} \to \Omega_\varepsilon \subseteq \mathcal{R}^\lceil$, d denoting the dimension of $\mathcal{M}$.

If $(\mathcal{M}, i, T)$ is a statistical manifold, then for all $\alpha \in R$ a torsion-free connection $\overset{\alpha}{\nabla}$ on $\mathcal{M}$ is given by the Christoffel symbols

$$\overset{\alpha}{\Gamma}_{rst} = \overset{0}{\Gamma}_{rst} - \frac{\alpha}{2} T_{rst}. \tag{1}$$

Here $\overset{0}{\Gamma}_{rst}$ denotes the Christoffel symbols in the Riemannian connection $\overset{0}{\nabla}$ on $\mathcal{M}$. The family of connections $\overset{\alpha}{\nabla}$ for $\alpha \in R$ is called the α-connections. We say that the family of α-connections is derived from T, and the geometry on $\mathcal{M}$ specified by i and $\overset{\alpha}{\nabla}$ will be called the α-geometry derived from T.

The curvature tensor $\overset{\alpha}{R}$ of the connection $\overset{\alpha}{\nabla}$ may be expressed in terms of the Christoffel symbols of $\overset{\alpha}{\nabla}$ as

$$\overset{\alpha}{R}_{rstu} = \{\partial_r \overset{\alpha}{\Gamma}{}^v_{st} - \partial_s \overset{\alpha}{\Gamma}{}^v_{rt}\} i_{vu} + \overset{\alpha}{\Gamma}_{rwu} \overset{\alpha}{\Gamma}{}^w_{st} - \overset{\alpha}{\Gamma}_{swu} \overset{\alpha}{\Gamma}{}^w_{rt} . \tag{2}$$

The connection $\overset{\alpha}{\nabla}$ is said to be α-flat if $\overset{\alpha}{R}= 0$ and this is the case if and only if there exists coordinates such that $\overset{\alpha}{\Gamma}_{rst}$ expressed in these coordinates vanish.

Moreover we note that a yoke $\gamma : \mathcal{M} \times \mathcal{M} \to R$ on $\mathcal{M}$ (see Bl@esild (1991) for a definition) given in local coordinates $\gamma(\omega_\varepsilon; \hat{\omega}_\varepsilon)$ defines a family of connections $\{\overset{\alpha}{\Gamma}\}_{\alpha \in R}$ by

$$\overset{1}{\Gamma}_{rst}= \partial_r \partial_s \hat{\partial}_t \gamma(\omega_\varepsilon; \omega_\varepsilon), \;\; \overset{-1}{\Gamma}_{rst}= \partial_t \hat{\partial}_r \hat{\partial}_s \gamma(\omega_\varepsilon; \omega_\varepsilon) \text{ and } T_{rst} = \overset{1}{\Gamma}_{rst} - \overset{-1}{\Gamma}_{rst} \tag{3}$$

and (1). Here $\hat{\partial}_r$ means differentiation w.r.t. $\hat{\omega}_\varepsilon$.

Dual systems and geometry.

Let $(\mathcal{M}, i)$ denote a Riemannian manifold and let $(\mathcal{M}, i, T)$ denote a statistical manifold.

Amari (1985) discusses Riemannian manifolds, where two special coordinate systems exist globally such that the condition (5) in the below definition is fulfilled. Here we extend the discussion to cover the general case, where more than one chart is necessary to cover $\mathcal{M}$, and in such a way that $\mathcal{M}$ becomes a ± 1-flat statistical manifold.

Definition 1: $(\mathcal{M}, i)$ is said to be a dual space if there exist two atlas $\{\Theta_\varepsilon\}_{\varepsilon \in E}$ and $\{H^\varepsilon\}_{\varepsilon \in E}$ such that

$$\theta_\varepsilon^{-1}(\Theta_\varepsilon) = (\eta^\varepsilon)^{-1}(H^\varepsilon) \text{ for all } \varepsilon \in E, \tag{4}$$

$$< \partial_\rho, \partial^\sigma >= \delta_\rho^\sigma \tag{5}$$

for all $\theta_\varepsilon(p) \in \Theta_\varepsilon$ and $\eta^\varepsilon(p) \in H^\varepsilon$, and such that

$$\theta_{\varepsilon_1}^\rho(p) = K_\sigma^\rho(\varepsilon_1, \varepsilon_2)\theta_{\varepsilon_2}^\sigma(p) + K^\rho(\varepsilon_1, \varepsilon_2) \tag{6}$$

and

$$\eta_\rho^{\varepsilon_1}(p) = L_\rho^\sigma(\varepsilon_1, \varepsilon_2)\eta_\sigma^{\varepsilon_2}(p) + K_\rho(\varepsilon_1, \varepsilon_2)$$

for $p \in U \subseteq \theta_{\varepsilon_1}^{-1}(\Theta_{\varepsilon_1}) \cap \theta_{\varepsilon_2}^{-1}(\Theta_{\varepsilon_2})$, U open and connected, and some matrices $K_\sigma^\rho(\varepsilon_1, \varepsilon_2)$ and $L_\rho^\sigma(\varepsilon_1, \varepsilon_2)$ and vectors $K^\rho(\varepsilon_1, \varepsilon_2)$ and $K_\rho(\varepsilon_1, \varepsilon_2)$. □

In definition 1 ρ, σ, τ etc denotes generic components of θ and η. They will however occur as upper indices on θ and lower indices on η. ∂_ρ means differentiation w.r.t. θ^ρ_ε and ∂^ρ means differentiation w.r.t. η^ε_ρ, i.e. $\partial_\rho = \partial/\partial\theta^\rho_\varepsilon$ and $\partial^\rho = \partial/\partial\eta^\varepsilon_\rho$ are the coordinate frames expressed in θ^ρ_ε and η^ε_ρ-coordinates. There will be no ε in general on ∂_ρ and ∂^ρ since it causes no confuse.

$< \cdot, \cdot >$ is the inner product product given by i.

Condition (4) says that Θ_ε and H^ε parameterize the same subset of $\mathcal{M}$ for any $\varepsilon \in E$. Condition (5) says that the coordinate frames corresponding to θ_ε and η^ε are i-orthogonal to each other, and finally condition (6) says that if some subset of $\mathcal{M}$ is parameterized by more than one chart, then the parameters will be affinely dependent in both θ and η.

The last condition (6) turns out to be important. We can think of (6) as a homogeneity condition on the parameters θ and η.

An atlas $\{\Omega_\varepsilon\}_{\varepsilon\in E}$ of a manifold $\mathcal{M}$ fulfilling condition (6) will be called an affine atlas, since all coordinate transformations $\omega^{-1}_{\varepsilon_1} \circ \omega_{\varepsilon_2}$ are affine functions from an open subset of R^d into R^d.

The definition is exemplified below.

Note from (5) that if (6) is fulfilled for the θ-coordinate then its fulfilled for the η-coordinate as well. From (6) we have

$$\partial^\rho_{\varepsilon_2} = K^\rho_\sigma(\varepsilon_1, \varepsilon_2)\partial^\sigma_{\varepsilon_1},$$

and hence from (5)

$$\delta^\sigma_\rho =< \partial^{\varepsilon_2}_\rho, \partial^\sigma_{\varepsilon_2} >=< K^\rho_\tau(\varepsilon_1, \varepsilon_2)\partial^{\varepsilon_1}_\tau, \eta^{\varepsilon_1/\upsilon}_\sigma \partial^\upsilon_{\varepsilon_1} >= K^\rho_\tau(\varepsilon_1, \varepsilon_2)\eta^{\varepsilon_1/\tau}_\sigma.$$

It means $\eta^{\varepsilon_1/\tau}_\sigma$ must be constant and the proof is completed.

Again we will drop the ε-index when there is no ambiguity involved, and we say that $(\theta_\varepsilon, \eta^\varepsilon)_{\varepsilon\in E}$ or just (θ, η) is a dual system on $(\mathcal{M}, i)$.

Note that if (θ, η) is a dual system then so is (η, θ).

The term dual can be explained in either of two ways: If we take ∂_ρ as the coordinate frame of the tangent bundle $T\mathcal{M}$ of $\mathcal{M}$ then it is seen from (5) that $< \partial^\sigma, \cdot >$ is the canonical frame in the dual bundle $T\mathcal{M}^*$ to $T\mathcal{M}$, and hence the duality is explained through the connection to dual vector spaces.

The other explanation will become apparent later but let us reveal that on each dual space there exists a pair of flat dual connections, see e.g. Lauritzen (1983) for detailed discussion of dual connections.

Corollary 1: If $(\mathcal{M}, \rangle)$ is a dual space then the constants depending on ε in condition (6) are related in the following way:

$$K_\sigma^\rho(\varepsilon_1, \varepsilon_2) L_\rho^\tau(\varepsilon_1, \varepsilon_2) = \delta_\sigma^\tau \quad \text{and} \quad K_\rho^\sigma(\varepsilon_2, \varepsilon_1) = L_\rho^\sigma(\varepsilon_1, \varepsilon_2).$$

Moreover

$$K_\sigma^\rho(\varepsilon_1, \varepsilon_2) K^\sigma(\varepsilon_2, \varepsilon_1) + K^\rho(\varepsilon_1, \varepsilon_2) = 0,$$

and

$$L_\rho^\sigma(\varepsilon_1, \varepsilon_2) L_\sigma(\varepsilon_2, \varepsilon_1) + L_\rho(\varepsilon_1, \varepsilon_2) = 0.$$

Proof: The first statement follows from (5) since $\theta^\rho_{\varepsilon_1/\sigma} = K_\sigma^\rho(\varepsilon_1, \varepsilon_2)$ and $\eta_\rho^{\varepsilon_1/\sigma} = L_\rho^\sigma(\varepsilon_1, \varepsilon_2)$. The other statements follows from $\theta_{\varepsilon_2}(\theta_{\varepsilon_1}(\theta_{\varepsilon_2})) = \theta_{\varepsilon_2}$ and $\eta^{\varepsilon_2}(\eta^{\varepsilon_1}(\eta^{\varepsilon_2})) = \eta^{\varepsilon_2}$ and (5). □

Example 1: Let $\mathcal{M}$ be $S^1 = \{\ (\cos\xi, \sin\xi) \mid \xi \in [0, 2\pi]\ \}$, the unit circle in R^2, and let i be the Euclidean metric induced from R^2. The following charts define an atlas of S^1:

$$\xi_1 : \{\ (\cos\xi^1, \sin\xi^1) \mid \xi^1 \in]0, 2\pi[\ \} \to]0, 2\pi[, \qquad \xi_1(\cos\xi^1, \sin\xi^1) = \xi^1,$$

$$\xi_2 : \{\ (\cos\xi^2, \sin\xi^2) \mid \xi^2 \in]-\pi, \pi[\ \} \to]-\pi, \pi[, \qquad \xi_2(\cos\xi^2, \sin\xi^2) = \xi^2,$$

It is easily seen that $(\xi^i, \xi^i)_{i\in\{1,2\}}$ fulfills (4)-(6) and hence is a dual system on S^1.

The family of von-Mises distributions equipped with the Fisher information metric and fixed dispersion parameter κ is diffeomorfic to S^1 and our example then shows us that we can not in general restrict considerations to models covered by a single chart. □

Example 2: Let $(\mathcal{M}, \rangle)$ be a 1-dimensional Riemannian manifold parametrized by an open interval I of R such that $i(t) = 1$ for all $t \in I$ (which is always possible). Let $\theta = \theta(t)$ be a diffeomorfism of I onto $\theta(I)$ and let $\eta = \eta(t)$ be given by

$$\frac{d\eta}{dt}(t) = \frac{dt}{d\theta}(\theta(t)). \tag{7}$$

Since $\theta = \theta(t)$ is a diffeomorfism then $\frac{dt}{d\theta} \neq 0$ and (7) defines a diffeomorfism from I onto $\eta(I)$. Note that $\partial_\theta = \frac{dt}{d\theta}\partial_t$ and $\partial_\eta = \frac{d\theta}{dt}\partial_t$ and hence

$$< \partial_\theta, \partial_\eta >=< \frac{dt}{d\theta}\partial_t, \frac{d\theta}{dt}\partial_t >= 1.$$

Condition (6) is trivially fulfilled since $\mathcal{M}$ is covered by a single chart. This means that (θ, η) is a dual system on $(\mathcal{M}, \rangle)$ and especially (t, t) is a dual system (put $\theta = t$). $\square$

Example 3: Let $(\mathcal{M}, \rangle)$ an exponential family equipped with the Fisher information metric and with open canonical parameter space Θ. Put $H = \eta(\Theta)$, η being the mean value map. Then (θ, η), $\theta \in \Theta$ and $\eta \in H$, is a dual system on $(\mathcal{M}, \rangle)$. It follows from

$$< \partial_\rho, \partial^\sigma >=< \partial_\rho, \theta^{\tau/\sigma}\partial_\tau >= \delta_\rho^\sigma,$$

since $< \partial_\rho, \partial_\sigma >= i_{\rho\sigma}$ and $\theta^{\tau/\sigma} = i^{\tau\sigma}$. Again condition (6) is trivially fulfilled since $\mathcal{M}$ is covered by a single chart. $\square$

The theorem stated below is from Amari (1985), and it is essential in the study of dual spaces.

Theorem 1: (Amari (1985) p.80) Let $\theta = (\theta^\rho)_\rho$ and $\eta = (\eta_\rho)_\rho$ be two parametrizations of an open subset of $(\mathcal{M}, i)$. Then

$$< \partial_\rho, \partial^\sigma >= \delta_\sigma^\rho \tag{8}$$

if and only if there exist functions κ of θ and λ of η such that

$$\theta^\rho = \partial^\rho \lambda(\eta) = \lambda^{/\rho}(\eta) \qquad \eta_\rho = \partial_\rho \kappa(\theta) = \kappa_{/\rho}(\theta) \tag{9}$$

$$i_{\rho\sigma} = \partial_\rho \partial_\sigma \kappa(\theta) = \eta_{\rho/\sigma} \qquad i^{\rho\sigma} = \partial^\rho \partial^\sigma \lambda(\eta) = \theta^{\rho/\sigma} \tag{10}$$

$$\kappa(\theta) + \lambda(\eta) - \theta^\rho \eta_\rho = 0. \tag{11}$$

If (8) is fulfilled then κ and λ are uniquely given modulo the same constant, i.e.

$$\tilde{\kappa}(\theta) = \kappa(\theta) + k \quad \text{and} \quad \tilde{\lambda}(\eta) = \lambda(\eta) - k$$

since differentiating (23) w.r.t. θ^ρ must give η_ρ, and (11) then gives us $\tilde{\lambda}$. Moreover λ is the Legendre transform of κ, and κ the Legendre transform of λ, i.e. $\lambda = \overset{\vee}{\kappa}$ and $\kappa = \overset{\vee}{\lambda}$.

It is however sufficient that $i_{\rho\sigma} = \partial_\rho \partial_\sigma \kappa(\theta)$ for some function κ in order to have (8), and in that case all functions $\tilde{\kappa}$ fulfilling $i_{\rho\sigma} = \partial_\rho \partial_\sigma \tilde{\kappa}(\theta)$ are given by

$$\tilde{\kappa}(\theta) = \kappa(\theta) + k_\rho \theta^\rho + k \tag{12}$$

for some constant vector k_ρ and constant scalar k. All choices of η are then determined uniquely modulo a translation as seen from (9) and (12). $\square$

Corollary 2: If $(\theta_\varepsilon, \eta^\varepsilon)$, $\varepsilon \in E$ is a dual system on $(\mathcal{M}, i)$ then there exists functions κ^ε and λ_ε such that (9)-(11) are satisfied for all $\varepsilon \in E$, and such that with the notation i definition 1 one has

$$\kappa^{\varepsilon_1}(\theta_{\varepsilon_1}) = \kappa^{\varepsilon_2}(\theta_{\varepsilon_2}) + K_\rho(\varepsilon_1, \varepsilon_2)\theta^\rho_{\varepsilon_1} \tag{13}$$

and

$$\lambda_{\varepsilon_1}(\eta^{\varepsilon_1}) = \lambda_{\varepsilon_2}(\eta^{\varepsilon_2}) + K^\rho(\varepsilon_1, \varepsilon_2)\eta^{\varepsilon_1}_\rho - K^\rho(\varepsilon_1, \varepsilon_2)K_\rho(\varepsilon_1, \varepsilon_2) \tag{14}$$

modulo a constant for all $\varepsilon_1, \varepsilon_2 \in E$ whenever $(\theta_{\varepsilon_1}, \eta^{\varepsilon_1})$ and $(\theta_{\varepsilon_2}, \eta^{\varepsilon_2})$ parametrize the same open connected subset of $\mathcal{M}$. The constant will have different sign for κ^ε and λ_ε but the same absolute value.

Proof: Assume $(\theta_\varepsilon, \eta^\varepsilon)$ is a dual system on $(\mathcal{M}, i)$. Since (5) per definition is satisfied, then according to theorem 1 there exists functions κ^ε and λ_ε such that (9)-(11) are satisfied for all $\varepsilon \in E$. Differentiating the right hand side of (13) once w.r.t $\theta^\rho_{\varepsilon_1}$ results in

$$L^\sigma_\rho(\varepsilon_1, \varepsilon_2)\eta^{\varepsilon_2}_\sigma(\theta_{\varepsilon_2}) + K_\rho(\varepsilon_1, \varepsilon_2) = L^\sigma_\rho(\varepsilon_1, \varepsilon_2)\eta^{\varepsilon_2}_\sigma + K_\rho(\varepsilon_1, \varepsilon_2) = \eta^{\varepsilon_1}$$

and differentiating twice w.r.t. $\theta^\rho_{\varepsilon_1}$ and $\theta^\sigma_{\varepsilon_1}$ results in

$$L^\tau_\rho(\varepsilon_1, \varepsilon_2)L^\upsilon_\sigma(\varepsilon_1, \varepsilon_2)i_{\tau\upsilon}(\theta_{\varepsilon_2}) = i_{\rho\sigma}(\theta_{\varepsilon_1})$$

where we have used (9), (10) and $\theta^\sigma_{\varepsilon_2/\rho} = L^\sigma_\rho(\varepsilon_1, \varepsilon_2)$ ((6) and corollary 1). Similarly for the right hand side of (14), i.e. the right hand sides define functions fulfilling (9) and (10). Note that (using (6))

$$\begin{aligned}&\kappa^{\varepsilon_2}(\theta_{\varepsilon_2}) + K_\rho(\varepsilon_1, \varepsilon_2)\theta^\rho_{\varepsilon_1} + \\ &\qquad \lambda_{\varepsilon_2}(\eta^{\varepsilon_2}) + K^\rho(\varepsilon_1, \varepsilon_2)\eta^{\varepsilon_1}_\rho - K^\rho(\varepsilon_1, \varepsilon_2)K_\rho(\varepsilon_1, \varepsilon_2) - \theta^\rho_{\varepsilon_1}\eta^{\varepsilon_1}_\rho = 0,\end{aligned}$$

which means that (11) is fulfilled too. According to theorem 1 κ^{ε_1} and λ_{ε_1} are uniquely determined modulo a constant and the proof is completed. $\square$

Proposition 1: Assume $(\mathcal{M}, i)$ is a dual space with dual system $(\theta_\varepsilon, \eta^\varepsilon)$, $\varepsilon \in E$. The function γ defined on an open subset of $\mathcal{M} \times \mathcal{M}$ containing the diagonal $\{(p,p)|p \in \mathcal{M}\}$ into R given in local coordinates θ_ε by

$$\begin{aligned}\gamma(\theta_\varepsilon; \hat{\theta}_\varepsilon) &= \theta_\varepsilon^\rho \eta_\rho^\varepsilon(\hat{\theta}_\varepsilon) - \kappa^\varepsilon(\theta_\varepsilon) - \lambda_\varepsilon(\eta^\varepsilon(\hat{\theta}_\varepsilon)) \\ &= \theta_\varepsilon^\rho \hat{\eta}_\rho^\varepsilon - \kappa^\varepsilon(\theta_\varepsilon) - \lambda_\varepsilon(\hat{\eta}^\varepsilon) \\ &= (\theta_\varepsilon^\rho - \hat{\theta}_\varepsilon^\rho)\hat{\eta}_\rho^\varepsilon - \kappa^\varepsilon(\theta_\varepsilon) + \kappa^\varepsilon(\hat{\theta}_\varepsilon)\end{aligned} \tag{15}$$

fulfills

(a) $\gamma(\theta_\varepsilon; \hat{\theta}_\varepsilon) \leq 0$ and equality if and only if $\theta_\varepsilon = \hat{\theta}_\varepsilon$

(b) $\partial_\rho \gamma(\theta_\varepsilon; \theta_\varepsilon) = \tilde{\partial}_\rho \gamma(\theta_\varepsilon; \theta_\varepsilon) = 0$

(c) $\partial_\sigma \partial_\rho \gamma(\theta_\varepsilon; \theta_\varepsilon) = -i_{\rho\sigma}(\theta_\varepsilon)$,

where $\hat{\partial}_\rho$ means differentiation w.r.t. $\hat{\theta}_\varepsilon^\rho$. Especially γ is a normed yoke on its domain in $\mathcal{M} \times \mathcal{M}$.

Note that γ is globally defined on $\mathcal{M} \times \mathcal{M}$ only if for all points $p, q \in \mathcal{M}$ there exists a chart θ_ε such that $p, q \in \theta_\varepsilon^{-1}(\Theta_\varepsilon)$.

Proof: We must first of all check that γ is welldefined. From (11) it follows that the two different expressions defining γ are identical. Let θ_{ε_1} and θ_{ε_2} be two local parametrizations of some open subset of $\mathcal{M}$. From (6) we know the relation between the dual coordinates, and from corollary 2 we know the relation between κ and λ functions. Expressing $\gamma(\theta_{\varepsilon_1}; \hat{\theta}_{\varepsilon_1})$ in terms of θ_{ε_2} yields (using (6) and corollary 2, and dropping the dependence of the constants on ε_1 and ε_2)

$$\begin{aligned}\gamma(\theta_{\varepsilon_1}; \hat{\theta}_{\varepsilon_1}) &= (K_\sigma^\rho \theta_{\varepsilon_2}^\sigma + K^\rho)(L_\rho^\tau \hat{\eta}_\tau^{\varepsilon_2} + K_\rho) - \kappa^{\varepsilon_2}(\theta_{\varepsilon_2}) - K_\rho \theta_{\varepsilon_1}^\rho \\ &\quad -\lambda_{\varepsilon_2}(\hat{\eta}^{\varepsilon_2}) - K^\rho \hat{\eta}_\rho^{\varepsilon_1} - K^\rho K_\rho \\ &= \theta_{\varepsilon_2}^\rho \hat{\eta}_\rho^{\varepsilon_2} - \kappa^{\varepsilon_2}(\theta_{\varepsilon_2}) - \lambda_{\varepsilon_2}(\hat{\eta}^{\varepsilon_2}) \\ &= \gamma(\theta_{\varepsilon_2}; \hat{\theta}_{\varepsilon_2}),\end{aligned}$$

which means that γ is well-defined.

(a), (b), and (c) follows from Amari (1985) p.84.

Since (c) is valid especially on the diagonal $\theta = \hat{\theta}$ it means together with (a) and (b), that γ is a normed yoke on its domain in $\mathcal{M} \times \mathcal{M}$. $\square$

Quite general in statistical contexts a dual system of a dual space $(\mathcal{M}, \rangle)$ consists of single charts θ and η and hence γ is a yoke on $\mathcal{M}$.

In the following we will call γ a normed yoke on $\mathcal{M}$ derived from the dual system $(\theta_\varepsilon, \eta^\varepsilon)$ (neglecting that γ in general only is defined on a subset of $\mathcal{M} \times \mathcal{M}$).

We will hereafter drop the ε-index unless there can be doubt about the meaning. It makes the notation simpler to read.

Corollary 3: Assume $(\mathcal{M}, \rangle)$ is a dual space. The function γ^c (c for complementary) defined on the domain of γ in $\mathcal{M} \times \mathcal{M}$ by

$$\gamma^c(\theta; \hat{\theta}) = \gamma(\hat{\theta}; \theta)$$

is the normed yoke derived from the dual system (η, θ).

Proof: From Bl@esild (1991) p.102 it follows that γ^c is a yoke, and the corollary is now a consequence of theorem 1. $\square$

Assume (θ, η) is a dual system and let κ and λ be given such that (9)-(11) are fulfilled. We know from the theorem 1, that all other versions of $\tilde{\kappa}$ and $\tilde{\lambda}$ are given by $\tilde{\kappa} = \kappa + k$ and $\tilde{\lambda} = \lambda - k$ for some constant k . It means $\tilde{\kappa}(\theta) + \tilde{\lambda}(\hat{\eta}) = \kappa(\theta) + \lambda(\hat{\eta})$, and therefore

Corollary 4: Assume $(\mathcal{M}, \rangle)$ is a dual space. The normed yoke γ derived from (θ, η) given in proposition 1 are uniquely determined from (θ, η) alone. $\square$

Example 4: Let $\mathcal{M} = (\mathfrak{X}, \mathcal{P})$ be an exponential family with open parameter space and densities w.r.t. some measure μ on $t(\mathfrak{X})$ (t denoting the minimal canonical statistic) given by

$$\frac{dP_\theta}{d\mu}(t) = \exp\{\theta^\rho t_\rho - \kappa(\theta)\} \quad t \in t(\mathfrak{X}).$$

Then the normed log-likelihood function restricted to $t \in \eta(\Theta)$, η denoting the mean value map, is a version of the yoke γ derived from the dual system (θ, η) on $(\mathcal{M}, \rangle)$ where i is the Fisher information metric (see example 3). It follows from the fact that κ differentiated twice w.r.t. θ yields the Fisher information metric and theorem 1. $\square$

Assume $(\mathcal{M}, \rangle)$ is a dual space. Let $\{\overset{\alpha}{\Gamma}\}_{\alpha \in R}$ be the family of α-connections derived from the yoke γ.

In order to use the Einstein's summation convention we need to adopt a notation a little unlike the usual notation: when the α-connection symbols are considered as functions of θ they will be denoted $\overset{\alpha}{\Gamma}_{\rho\sigma\tau}$ and $\overset{\alpha}{\Gamma}{}^{\rho}_{\sigma\tau}$ and when they are considered as functions of η they will be denoted $\overset{\alpha}{\Gamma}{}^{\sigma\tau}_{\rho}$ and $\overset{\alpha}{\Gamma}{}^{\rho\sigma\tau}$. It is in line with the convention that the metric expressed in η-coordinates is denoted $i^{\rho\sigma}$. Furthermore, recall that $\hat{\partial}_\rho$ means differentiation w.r.t. $\hat{\theta}^\rho$ and $\hat{\partial}^\rho$ means differentiation w.r.t. $\hat{\eta}_\rho$.

Proposition 2: The 1-connection derived from γ is 1-flat and θ is a 1-flat parameter. The -1-connection derived from γ is -1-flat and η is a -1-flat parameter.

Proof: Follows easily from (3) in 'Preliminaries'. $\square$

Definition 2: $(\mathcal{M}, i, T)$ is called an α-space if the curvature tensor $\overset{\alpha}{R}$ derived from $\overset{\alpha}{\nabla}$ satisfies $\overset{\alpha}{R} = 0$. From Lauritzen (1983) we know that an α-space is a $-\alpha$-space too. $\square$

Remark 1: Assume $(\mathcal{M}, \rangle)$ is a dual space. Put $D_{\rho\sigma\tau} = 2(\overset{0}{\Gamma}_{\rho\sigma\tau} - \overset{1}{\Gamma}_{\rho\sigma\tau})$, where $\overset{0}{\Gamma}$ and $\overset{1}{\Gamma}$ are the Christoffel symbols of the connections $\overset{0}{\nabla}$ and $\overset{1}{\nabla}$ derived from the yoke γ. Then $(\mathcal{M}, i, D)$ is an ± 1-space (proposition 2). D will hereafter refer to the tensor derived from γ, and $\overset{\alpha}{\Gamma}$ to the α-connection derived from γ or equivalently from D. $\square$

Proposition 3: If $(\mathcal{M}, i, T)$ is an α-space then one can choose an α-flat parameter θ and a -α-flat parameter η such that $(\mathcal{M}, i)$ is a dual space with dual system (θ, η).

Proof: According Amari (1985) th.3.5 p.81 the proposition is true for small neighborhoods around any $p \in \mathcal{M}$, i.e. we need only check condition (6). Let θ and ω be α-flat parameters parameterizing the same connected subset of $\mathcal{M}$. Since θ and ω

are both α-flat, it follows from the transformation law for connection symbols

$$\overset{\alpha}{G}{}^{\rho}_{\sigma\tau} = \{\overset{\alpha}{G}{}^{t}_{rs}\ \theta^{r}_{/\rho}\theta^{s}_{/\sigma} + \theta^{t}_{/\rho\sigma}\}\omega^{\tau}_{/t},$$

where $\overset{\beta}{G}$, $\beta \in R$, refers to the β-connection derived from T, that

$$\theta^{t}_{/\rho\sigma}\omega^{\tau}_{/t} = 0,$$

and hence it is seen that they are affinely connected. Similar for $-\alpha$-flat parameters. □

Proposition 4: Let $(\mathcal{M}, i, T)$ be an α-space such that $(\mathcal{M}, \rangle)$ is a dual space with the dual system (θ, η) chosen in such a way that $\overset{\alpha}{G}_{\rho\sigma\tau} = 0$ (see proposition 3), where $\overset{\beta}{G}$ refers to the β-connection derived from T. Then $\overset{0}{G} = \overset{0}{\Gamma}$, $\overset{\alpha}{G} = \overset{1}{\Gamma}$ and $\alpha T = D$.

Proof: $\overset{0}{G} = \overset{0}{\Gamma}$ since the 0-connection depend on i only. Since θ is α-flat one has $\overset{\alpha}{G}_{\rho\sigma\tau} = 0$ and from proposition 2 one obtains that $\overset{1}{\Gamma}_{\rho\sigma\tau} = 0$, i.e. $\overset{\alpha}{G} = \overset{1}{\Gamma}$. From $\overset{0}{\Gamma}_{\rho\sigma\tau}$ $-\frac{1}{2}D_{\rho\sigma\tau} = \overset{1}{\Gamma}_{\rho\sigma\tau} = \overset{\alpha}{G}_{\rho\sigma\tau} = \overset{0}{G}_{\rho\sigma\tau}$ $-\frac{\alpha}{2}T_{\rho\sigma\tau}$ it follows that $\alpha T = D$. □

Note that the derived geometry of a 0-space $(\mathcal{M}, \rangle, \mathcal{T})$ is particular simple since $\overset{0}{\Gamma} = \overset{0}{G}$, $D = \alpha T = 0T = 0$ and hence $\overset{\beta}{\Gamma} = 0$ for all $\beta \in R$.

Collecting the results in proposition 3 and remark 1 we have the following characterization of dual spaces:

Proposition 5: If $(\mathcal{M}, i, T)$ is an α-space then $(\mathcal{M}, i)$ is a dual space. If $(\mathcal{M}, i)$ is a dual space then $(\mathcal{M}, i, D)$ is an 1-space. □

Proposition 5 can be considered as a reformulation of a problem stated in Amari (1985) p.106[6]: 'If $(\mathcal{M}, \rangle)$ is a Riemannian manifold is it then possible to associate a tensor T such that $(\mathcal{M}, \rangle, \mathcal{T})$ is α-flat for some α, and if not what is the condition imposed on i to guarantee this.' Proposition 5 tells us that this is possible if and only if $(\mathcal{M}, \rangle)$ is a dual space. The reformulation is poor in the sense that we do not have (at the moment at least) any effective way to control if $(\mathcal{M},\rangle)$ is dual.

Example 5: (example 2 continued) Put $I =]0,1[$ and $\theta = \theta(t) = t^2$. Then $\frac{dt}{d\theta} = \frac{1}{2\sqrt{\theta}}$ and $\frac{dt}{d\theta}(\theta(t)) = \frac{1}{2t}$, and if we put $\eta(t) = \frac{1}{2}\log t = \frac{1}{4}\log\theta$ then (θ,η) is a dual system on I. We can choose $\kappa_\theta(\theta)$ to be

$$\kappa(\theta) = \int \eta(\theta)d\theta = \frac{1}{4}\theta\log\theta - \frac{1}{4}\theta$$

and hence γ_θ given by

$$\gamma_\theta(\theta;\tilde{\theta}) = \frac{1}{4}\theta\log\tilde{\theta} - \frac{1}{4}\theta\log\theta + \frac{1}{4}\theta - \frac{1}{4}\tilde{\theta}$$

or in t-coordinate

$$\gamma_\theta(t;\tilde{t}) = \frac{1}{2}t^2\log\tilde{t} - \frac{1}{2}t^2\log t + \frac{1}{4}t^2 - \frac{1}{4}\tilde{t}^2$$

is the yoke derived from the dual system (θ,η). Similarly we obtain the following expression for the yoke γ_t derived from the dual system (t,t) $(\kappa_t(t) = \frac{1}{2}t^2)$

$$\gamma_t(t;\tilde{t}) = t\tilde{t} - \frac{1}{2}t^2 - \frac{1}{2}\tilde{t}^2.$$

It is easy to see that $\gamma_\theta \neq \gamma_t$ and hence different dual systems might derive different yokes. From 'Preliminaries' we see that the ± 1-connections derived from γ_θ are given by

$$_\theta\overset{1}{\Gamma}_{ttt} = 1/t \qquad _\theta\overset{-1}{\Gamma}_{ttt} = -1/t$$

whereas all α-connections derived from γ_t expressed in t-coordinate are 0 since it is so in the ± 1- and 0-connections. This means that the family of connections derived from different dual systems not in general are identical. $\square$

The above example leads to the following definition. We first note that an expression like $\gamma = \tilde{\gamma}$ means that the two functions γ and $\tilde{\gamma}$ are identical on their common domain in $\mathcal{M}\times\mathcal{M}$, and not that they necessarily are defined on the same subset of $\mathcal{M}\times\mathcal{M}$.

Definition 3: Let $(\mathcal{M},i)$ be a dual space with dual coordinates (θ,η). The yoke γ

derived from (θ, η) is said to be D-invariant (D for dual) if the yoke $\tilde{\gamma}$ derived from any other choice of dual coordinates $(\tilde{\theta}, \tilde{\eta})$ is equal to γ, i.e. $\tilde{\gamma} = \gamma$. $\square$

Proposition 6: Let $(\mathcal{M}, i)$ be a dual space with dual coordinates (θ, η). The yoke γ derived from (θ, η) is D-invariant if and only if all other dual systems $(\tilde{\theta}, \tilde{\eta})$ on $(\mathcal{M}, i)$ is given by

$$\tilde{\theta}^\alpha = C^\alpha_\rho \theta^\rho + D^\alpha \qquad \tilde{\eta}_\alpha = A^\rho_\alpha \eta_\rho + B_\alpha \tag{16}$$

and

$$C^\alpha_\rho A^\rho_\beta = \delta^\alpha_\beta.$$

where α, β etc are generic components of $\tilde{\theta}$ and $\tilde{\eta}$.

Proof: Assume γ is D-invariant and let $(\tilde{\theta}, \tilde{\eta})$ be an other dual system. Then

$$\gamma(\theta; \theta') = \gamma(\theta(\tilde{\theta}); \theta(\tilde{\theta}')) = \tilde{\gamma}(\tilde{\theta}, \tilde{\theta}') = \tilde{\theta}^\alpha \tilde{\eta}'_\alpha - \tilde{\kappa}(\tilde{\theta}) - \tilde{\lambda}(\tilde{\eta})$$

where $\tilde{\kappa}$ and $\tilde{\lambda}$ refer to functions fulfilling (9)-(11) in the dual system $(\tilde{\theta}, \tilde{\eta})$. From proposition 2 it follows that $\overset{1}{\Gamma}_{\alpha\beta\gamma} = 0$ and $\overset{1}{\Gamma}_{\rho\sigma\tau} = 0$ and hence $\tilde{\theta}$ and θ are affinely connected (consequence of the transformation law for connection symbols). Similarly using the -1-connection one obtains that $\tilde{\eta}$ and η are affinely connected. Finally since $(\tilde{\theta}, \tilde{\eta})$ is a dual system then

$$\delta^\alpha_\beta = < \partial^\alpha, \partial_\beta > = C^\alpha_\rho A^\sigma_\beta < \partial^\rho, \partial_\sigma > = C^\alpha_\rho A^\rho_\beta.$$

It proves the first implication. Oppositely, if $\tilde{\theta}$ and $\tilde{\eta}$ are given by (16) then

$$\tilde{\kappa}(\tilde{\theta}) = \kappa(A_\alpha(\tilde{\theta}^\alpha - D^\alpha)) + \tilde{\theta}^\alpha B_\alpha$$

and

$$\tilde{\lambda}(\tilde{\eta}) = \lambda(C^\alpha(\tilde{\eta}_\alpha - B_\alpha)) + \tilde{\eta}_\alpha D^\alpha - B_\alpha D^\alpha$$

fulfill condition (9)-(11) as is seen by calculation and using the fact that $\kappa(\theta) + \lambda(\eta) - \theta^\rho \eta_\rho = 0$. The yoke $\tilde{\gamma}$ derived from $(\tilde{\theta}, \tilde{\eta})$ is then given by

$$\begin{aligned}
\tilde{\gamma}(\tilde{\theta}; \tilde{\theta}') &= \tilde{\theta}^\alpha \tilde{\eta}'_\alpha - \tilde{\kappa}(\tilde{\theta}) - \tilde{\lambda}(\tilde{\eta}') \\
&= (C^\alpha_\rho \theta^\rho + D^\alpha)(A^\sigma_\alpha \eta'_\sigma + B_\alpha) - \kappa(A_\alpha(\tilde{\theta}^\alpha - D^\alpha)) - \tilde{\theta}^\alpha B_\alpha \\
&\quad -\lambda(C^\alpha(\tilde{\eta}'_\alpha - B_\alpha)) - \tilde{\eta}'_\alpha D^\alpha + B_\alpha D^\alpha \\
&= \theta^\rho \eta'_\rho - \kappa(\theta) - \lambda(\eta') \\
&= \gamma(\theta; \theta'),
\end{aligned}$$

i.e. $\gamma = \tilde{\gamma}$ and γ is D-invariant. It proves the second implication, and the proposition is proved. $\square$

Proposition 6 is analogous to the structure theorem for minimal representations of exponential families. The proof given here has the advantage that it first of all is more general and secondly that it is purely geometric, i.e. that there is no reference to density functions etc.

Definition 4: Proposition 6 defines an equivalence relation on the class of dual systems on $(\mathcal{M}, i)$ in the following way ($\sim$ means equivalent)

$$(\theta, \eta) \sim (\tilde{\theta}, \tilde{\eta}) \quad \Leftrightarrow \quad \gamma = \tilde{\gamma}.$$

We say that $(\tilde{\theta}, \tilde{\eta})$ is D-invariant in relation to (θ, η) if $(\theta, \eta) \sim (\tilde{\theta}, \tilde{\eta})$. $\square$

From definition 4 we see that if $(\theta, \eta) \sim (\tilde{\theta}, \tilde{\eta})$ then the α-connections derived from the yokes γ and $\tilde{\gamma}$ are identical. We can prove an opposite statement to that: let $\{\overset{\alpha}{\Gamma}\}_{\alpha \in R}$ be a family of connections derived from a dual s system (θ, η) on $(\mathcal{M}, i)$ via the yoke γ. We call $\{\overset{\alpha}{\Gamma}\}_{\alpha \in R}$ a D-family on $(\mathcal{M}, i)$. Then we have:

Proposition 7: Let $(\mathcal{M}, i)$ be a dual space, (θ_1, η^1) and (θ_2, η^2) dual systems on $(\mathcal{M}, i)$, and $\{{}_1\overset{\alpha}{\Gamma}\}_{\alpha \in R}$ and $\{{}_2\overset{\alpha}{\Gamma}\}_{\alpha \in R}$ D-families on $(\mathcal{M}, i)$ derived from (θ_1, η^1) and (θ_2, η^2). Then

$$(\theta_1, \eta^1) \sim (\theta_2, \eta^2) \quad \Leftrightarrow \quad {}_1\overset{\alpha}{\Gamma} = {}_2\overset{\alpha}{\Gamma} \;\text{ for all }\; \alpha \in R.$$

Proof: The implication $\Rightarrow$ is trivial (definition 4). Assume ${}_1\overset{\alpha}{\Gamma} = {}_2\overset{\alpha}{\Gamma}$ for all $\alpha \in R$. Since ${}_1\overset{1}{\Gamma} = 0$ in θ_1-coordinates (proposition 2) then ${}_2\overset{1}{\Gamma} = 0$ in θ_1-coordinates per assumption. Similar ${}_1\overset{-1}{\Gamma} = {}_2\overset{-1}{\Gamma} = 0$ in η^1-coordinates. But we too have that ${}_2\overset{1}{\Gamma} = 0$ in θ_2-coordinates and ${}_2\overset{-1}{\Gamma} = 0$ in η^2-coordinates. It means using the transformation law for connection symbols that θ_1 and θ_2 are affinely connected and similar for η^1 and η^2. Putting

$$\theta_1^\rho = C^\rho_\sigma \theta_2^\sigma + D^\rho \quad \text{and} \quad \eta^1_\rho = A^\sigma_\rho \eta^2_\sigma + B_\rho$$

then

$$< \partial^1_\rho, \partial^\sigma_1 >= \delta^\sigma_\rho =< \partial^2_\tau, \partial^\upsilon_2 > C^\tau_\rho A^\sigma_\upsilon = C^\tau_\rho A^\sigma_\tau$$

since both systems are dual (∂^1_ρ means of cause differentiation w.r.t. θ^ρ_1 etc). The result follows now from proposition 6 and definition 4. □

A note of caution. From proposition 2 we know that θ is a flat parameter in the 1-connection $\overset{1}{\Gamma}$ derived from γ and that θ is a flat parameter in the -1-connection ${}_c\overset{-1}{\Gamma}$ derived from γ^c. It means that $\overset{1}{\Gamma}= {}_c\overset{-1}{\Gamma}$. Similarly we obtain $\overset{-1}{\Gamma}= {}_c\overset{1}{\Gamma}$ and hence will the family of connections derived from γ equal the family of connections derived from γ^c but not in the sense of proposition 7.

β-Orthogeodesic Models.

In the last section we have seen that a dual space behave from a geometric point of view similarly to an exponential family. Hence we can introduce models analogous to those studied in Barndorff-Nielsen and Bl@esild (1983,1993) with the dual parameters θ and η playing the role as the canonical parameter and the mean value parameter in exponential family theory.

In this section we will be concerned with orthogeodesic models and we will relate then to dual spaces through a structure theorem analogous to the structure theorem given in Barndorff-Nielsen and Bl@esild (1993). Before discussing the structure theorem we will take a look at the definitions of the orthogeodesic property as given in Barndorff-Nielsen and Bl@esild (1993), generalize these definitions slightly and show that the definitions not are equivalent as postulated in Barndorff-Nielsen and Bl@esild (1993).

From now on we will assume that $\mathcal{M}$ can be covered by single coordinate chart although our results can be stated similarly without this assumption, but in a more cumbersome way.

Here follows the two definitions from Barndorff-Nielsen and Bl@esild (1993) slightly altered: Definition 3.1 and Definition 3.1' numbered A and A' respectively. For $\beta = 1$ the definitions given below are identical to the definitions given in Barndorff-Nielsen and Bl@esild (1993).

Definition A: $\mathcal{M}$ is said to be β-orthogeodesic relative to the parameterization

$\omega = (\chi, \psi)$ if the following conditions are satisfied:

(o) χ and ψ are variation independent.
(i) χ and ψ are i-orthogonal, i.e. $i_{aj} = 0$
(ii) The ψ-part of i depends on ψ only, i.e. $i_{jk}(\chi, \psi) = i_{jk}(\psi)$.
(iii) For every value of χ and ψ: $\overset{\beta}{\Gamma}{}^a_{jk} = 0$.
(iv) For every value of χ and ψ: $\overset{\beta}{\Gamma}{}^k_{ij} = 0$.

Let $\mathcal{M}_\chi$ be the submanifold $\{ \ p \in \mathcal{M} \mid \exists\psi : p = p(\chi, \psi) \ \}$ and let ${}_\chi\overset{\alpha}{\Gamma}$ denote the α-connections induced on $\mathcal{M}_\chi$ from $\mathcal{M}$. Note if (i) is fulfilled, then one has

$$ {}_\chi\overset{\alpha}{\Gamma}{}^i_{jk} = \overset{\alpha}{\Gamma}{}^i_{jk}, \quad {}_\chi\overset{\alpha}{\Gamma}_{ijk} = \overset{\alpha}{\Gamma}_{ijk}, \quad \text{and} \quad {}_\chi T_{ijk} = T_{ijk} = 2({}_\chi\overset{0}{\Gamma}_{ijk} - {}_\chi\overset{1}{\Gamma}_{ijk}). \tag{17} $$

Note that ${}_\chi T_{ijk} = T_{ijk}$ always is fulfilled per definition of the induced tensor (similar to ${}_\chi i_{jk} = i_{jk}$).
For the α-curvature tensor ${}_\chi\overset{\alpha}{R}$ of the induced α-connection on $\mathcal{M}_\chi$ we have

$$ {}_\chi\overset{\alpha}{R}_{jklm} = \overset{\alpha}{R}_{jklm} + (\overset{-\alpha}{H}_{jma}\overset{\alpha}{H}_{klb} - \overset{-\alpha}{H}_{jla}\overset{\alpha}{H}_{kmb}) i^{ab}, \tag{18} $$

where $\overset{\alpha}{H}_{jka}$ denotes components of the α-shape tensor ${}_\chi\overset{\alpha}{H}$ (see Barndorff-Nielsen and Bl@esild (1993) (2.13) and (2.27)).
And if (ii) is fulfilled, then

$$ \overset{0}{\Gamma}_{ijk}(\chi, \psi) = \overset{0}{\Gamma}_{ijk}(\psi), \tag{19} $$

since the Riemannian connection symbols $\overset{0}{\Gamma}_{ijk}$ is derived from i_{jk} (see e.g. Barndorff-Nielsen and Bl@esild (1993) (2.4)).
Note, that (i) and (iii) implies that $\mathcal{M}_\chi$ is β-geodesic, and (i) and (iv) implies that $\mathcal{M}_\chi$ is β-flat.
Finally note if definition A is fulfilled with $\beta = 0$ then (iii) is a consequence of (i) and (ii) (see e.g. Barndorff-Nielsen and Bl@esild (1993) (2.4))

Definition A': $\mathcal{M}$ is said to be β-orthogeodesic if the following conditions are satisfied:

(o)' $\mathcal{M}$ is a product manifold of the form $\mathcal{M} = X \times \Psi$,
where X and Ψ are differentiable manifolds.

(i)' The factorization of $\mathcal{M}$ is i-orthogonal.

(ii)' For every value of χ the restriction of i to the submanifold $\mathcal{M}_\chi$ does not depend on χ.

(iii)' For every value of χ the submanifold $\mathcal{M}_\chi$ is β-geodesic,
i.e. the β-shape tensor ${}_\chi\overset{\beta}{H}$ vanishes identically.

(iv)' For every value of χ the submanifold $\mathcal{M}_\chi$ is β-flat,
i.e. the curvature tensor ${}_\chi\overset{\beta}{R}$ vanishes identically.

If $\beta = 1$ the condition (iii)' is different in formulation to the similar condition (iii)' in Barndorff-Nielsen and Bl@esild (1993). Condition (iii)' in Barndorff-Nielsen and Bl@esild (1993) says:

For every value of χ and some value of $\alpha \neq 0$ the submanifold $\mathcal{M}_\chi$ is α-geodesic, i.e. the α-shape tensor ${}_\chi\overset{\alpha}{H}$ vanishes identically.

According to theorem 4.2 in Barndorff-Nielsen and Bl@esild (1993) the two conditions are equivalent under (i)' and (ii)', and the only reason for changing it here is, that it makes (iii)' more similar to the other conditions in definition 1'.

If $\beta = 0$ then (iii)' is a consequence of (i)' and (ii)' since $\overset{0}{\Gamma}{}^a_{jk} = \overset{0}{H}{}^a_{jk} = 0$ (see e.g. Barndorff-Nielsen and Bl@esild (1993) (2.4) p.5 and (2.19)).

We will now discuss to what extent the two definitions A and A' of the orthogeodesic property are equivalent. In order to do that, we will first prove two lemmas connecting different statements on geometric quantities to each other.

Lemma 1: If (i)-(iii) in definition A are satisfied, then the following statements are equivalent:

(a) $\overset{\alpha}{\Gamma}{}^i_{jk} = \overset{\alpha}{\Gamma}{}^i_{jk}(\psi)$ for some $\alpha \neq 0$ (a') $\overset{\alpha}{\Gamma}{}^i_{jk} = \overset{\alpha}{\Gamma}{}^i_{jk}(\psi)$ for all α

(b) $\overset{\alpha}{\Gamma}_{ijk} = \overset{\alpha}{\Gamma}_{ijk}(\psi)$ for some $\alpha \neq 0$ (b') $\overset{\alpha}{\Gamma}_{ijk} = \overset{\alpha}{\Gamma}_{ijk}(\psi)$ for all α

(c) $T_{ijk} = T_{ijk}(\psi)$

If $\beta \neq 0$ the following conditions are equivalent to the above for all $\alpha \in R$:

(d) $\overset{\alpha}{R}_{aijk} = 0$ for some $\alpha \neq 0$ (d') $\overset{\alpha}{R}_{aijk} = 0$ for all α.

and in that case one has

(A) $\overset{\alpha}{\Gamma}_{ajk}=\overset{\alpha}{\Gamma}_{jak}=\overset{\alpha}{\Gamma}_{jka}=0$ for all α, and (B) $\overset{\alpha}{\Gamma}{}^{a}_{jk}=\overset{\alpha}{\Gamma}{}^{j}_{ak}=\overset{\alpha}{\Gamma}{}^{j}_{ka}=0$ for all α.

Moreover in analogy with (17) one has (C) ${}_{\chi}\overset{\alpha}{R}=\overset{\alpha}{R}$ for all α.

If $\beta=0$ then the equivalences hold only in general if $\alpha=\beta=0$.

Proof: Follows easily using (1) and (2), and will be omitted here. □

Lemma 2: Assume (i) in definition A is satisfied.

If ${}_{\tilde{\chi}}\overset{\alpha}{R}=0$ for some $\tilde{\chi}$, then

(a) $\overset{\alpha}{\Gamma}{}^{i}_{jk}=\overset{\alpha}{\Gamma}{}^{i}_{jk}(\psi)$,

implies

(e) $\exists$ reparameterization ϕ of ψ: $\overset{\alpha}{\Gamma}{}^{\iota}_{\kappa\lambda}(\chi,\phi)=0$ for all χ and ϕ.

Opposite to that one has

(e) $\exists$ reparameterization ϕ of ψ: $\overset{\alpha}{\Gamma}{}^{\iota}_{\kappa\lambda}(\chi,\phi)=0$ for all χ and ϕ,

implies

(a) $\overset{\alpha}{\Gamma}{}^{i}_{jk}=\overset{\alpha}{\Gamma}{}^{i}_{jk}(\psi)$.

Proof: Assume (i) in definition A. Transforming ψ: $\phi=\phi(\psi)$ it follows from (i) and the transformation law for connection symbols, that

$$\overset{\alpha}{\Gamma}{}^{\iota}_{\lambda\kappa}=\{\overset{\alpha}{\Gamma}{}^{i}_{jk}\,\psi^{j}_{/\lambda}\psi^{k}_{/\kappa}+\psi^{i}_{/\kappa\lambda}\}\phi^{\iota}_{/i}\ , \tag{20}$$

where ι,κ etc denote indices with respect to ϕ. Assume (a). If ${}_{\tilde{\chi}}\overset{\alpha}{R}=0$ for some $\tilde{\chi}$, then there exists a parameterization $\phi=\phi(\psi)$ such that: ${}_{\tilde{\chi}}\overset{\alpha}{\Gamma}{}^{\iota}_{\lambda\kappa}=0$ (Amari (1985) p.152), and from (i) and (17) $\overset{\alpha}{\Gamma}{}^{\iota}_{\lambda\kappa}={}_{\tilde{\chi}}\overset{\alpha}{\Gamma}{}^{\iota}_{\lambda\kappa}=0$. From (20) and (a) one has $\overset{\alpha}{\Gamma}{}^{\iota}_{\lambda\kappa}=\overset{\alpha}{\Gamma}{}^{\iota}_{\lambda\kappa}(\psi)$, i.e. independent of χ. For $\tilde{\chi}$ its 0 and hence it is 0 for all χ. This proofs (e). Assume (e). If there exists a parameterization $\phi=\phi(\psi)$ such that $\overset{\alpha}{\Gamma}{}^{\iota}_{\lambda\kappa}=0$ for all χ and ϕ, then $\overset{\alpha}{\Gamma}{}^{i}_{jk}=\{\overset{\alpha}{\Gamma}{}^{\iota}_{\lambda\kappa}\,\phi^{\lambda}_{/j}\phi^{\kappa}_{/k}+\phi^{\iota}_{/jk}\}\psi^{i}_{/\iota}=\phi^{\iota}_{/jk}\psi^{i}_{/\iota}$ is a function only of ψ. This proofs (a). □

Note from (e) that ${}_{\tilde{\chi}}\overset{\alpha}{R}=0$, (a) and (i) imply that there exists a reparameterization ϕ

of ψ such that $\overset{\alpha}{\Gamma}{}^{\iota}_{\kappa\lambda}(\chi,\phi)=0$ for all χ and ϕ and hence ${}_{\chi}\overset{\alpha}{R}=0$ for all χ (see (17) and (2)).

As indicated by lemma 1 there seems to be two different cases: $\beta \neq 0$ and $\beta = 0$. Well, so it also turns out. Using lemma 1 and 2 we are now able to prove the equivalence between the definition A and A' as generally as possible.

Proof: definition A implies definition A'.

The proof is given in Barndorff-Nielsen and Bl@esild (1993) p.1033 with 1 replaced by $\beta \in R$, and will not be given here. $\square$

Proof of the existence of a parameterization, such that definition A' implies definition A if one (and hence all) of the conditions (a),(a'),(b),(b'),(c),(d), and (d') is fulfilled for some $\beta \neq 0$.

From Barndorff-Nielsen and Bl@esild (1993) p.1033 it follows, that definition A' implies (o)-(iii) in definition A with $\beta = 1$ replaced by $\beta \neq 0$. It follows from lemma 1 that all conditions are equivalent, since (i)-(iii) are fulfilled. If one of these conditions is satisfied, it follows from lemma 2, that there exists a parameterization, such that (iv) is satisfied. $\square$

Proof of the existence of a parameterization, such that definition A' implies definition A if $\beta = 0$.

Similar to the proof above we need only to prove (iv), and it is a consequence of lemma 2 with $\alpha = 0$ and $\overset{0}{\Gamma}_{ijk}=\overset{0}{\Gamma}_{ijk}(\psi)$ (see (19)). $\square$

Example 6: If $\mathcal{M}$ is a conjugate symmetric space ($\overset{\alpha}{R}=\overset{-\alpha}{R}$ for all $\alpha \in R$, see Lauritzen (1987) p.186) and (i)'-(iii)' in definition A' are satisfied with $\beta \neq 0$, then (d) is satisfied too. It follows from the below argument: calculating $\overset{\alpha}{R}_{ijka}$, $\alpha \in R$ using that (i)'-(iii)' implies (i)-(iii) and lemma 1 we find that (2)

$$\overset{\alpha}{R}_{ijka}=(\partial_i \overset{\alpha}{\Gamma}{}^{r}_{jk} - \partial_j \overset{\alpha}{\Gamma}{}^{r}_{ik})i_{ra} + \overset{\alpha}{\Gamma}_{ira}\overset{\alpha}{\Gamma}{}^{r}_{jk} - \overset{\alpha}{\Gamma}_{jra}\overset{\alpha}{\Gamma}{}^{r}_{ik} = 0$$

since there will be at least one 0 in all terms. Finally since $\mathcal{M}$ is conjugate symmetric $\overset{\alpha}{R}_{kaij}=\overset{\alpha}{R}_{ijka}=0$ (Lauritzen (1987) p.187), and hence (d). Especially an exponential family is conjugate symmetric. □

Example 7: (Counterexample) Let $\mathcal{M}$ be a 2-dimensional Riemannian manifold parameterized by $\omega = (\chi, \psi)$ and such that (o)-(ii) in definition A is satisfied. Let a symmetric (0,3)-tensor T on $\mathcal{M}$ fulfill

$$T_{\psi\psi\chi} = 0, \qquad T_{\psi\psi\psi} = f(\chi, \psi),$$

where f for ψ fixed is a non-constant function of χ. Hence it is easily checked that $(\mathcal{M}, \rangle, \mathcal{T})$ is a statistical manifold fulfilling (o)-(iii) with $\beta = 1$, but not $\overset{1}{\Gamma}{}^i_{jk}=\overset{1}{\Gamma}{}^i_{jk}(\psi)$, and hence (iv) is not satisfied with $\beta = 1$. It is therefore not possible to reparameterize ψ, such that definition A is satisfied. However since $d_\psi = 1$ one has ${}_\chi\overset{1}{R}=0$ for all χ, which means that definition A' is fulfilled. □

As the example shows the class of models fulfilling definition A' is bigger than the class of models fulfilling definition A. It is however convenient to have a coordinate free characterization of the conditions in definition A, and we therefore want to add an extra condition to definition A' in such a way, that definition A and definition A' with this extra condition are equivalent for all $\beta \in R$. From the discussion above it follows that this extra condition must be such that is trivially fulfilled for $\beta = 0$ under (i)'-(iv)'. I therefore propose the following coordinate free characterization (from now on definition A' is suspended and the meaning of the word β-orthogeodesic will be the one given in definition A):

Definition B': $\mathcal{M}$ is said to be β-orthogeodesic if the following conditions are satisfied:

- (o)' $\mathcal{M}$ is a product manifold of the form $\mathcal{M} = X \times \Psi$, where X and Ψ are differentiable manifolds.
- (i)' The factorization of $\mathcal{M}$ is i-orthogonal.
- (ii)' For every value of χ the restriction of i to the submanifold $\mathcal{M}_\chi$ does not depend on χ.
- (iii)' For every value of χ the submanifold $\mathcal{M}_\chi$ is β-geodesic, i.e. the β-shape tensor ${}_\chi\overset{\beta}{H}$ vanishes identically.

(iv)' For every value of χ the submanifold $\mathcal{M}_\chi$ is β-flat, i.e. the curvature tensor ${}_\chi\overset{\beta}{R}$ vanishes identically.

(v)' For every value of χ the restriction of βT to the submanifold $\mathcal{M}_\chi$ does not depend on χ.

Here βT means the (0,3)-tensor with components βT_{rst}. Note that the first four conditions are equivalent to the conditions i definition A'. Note too that (v)' is trivially fulfilled if $\beta = 0$.

If definition B' is satisfied with $\beta \neq 0$ then there exists a reparameterization ϕ of ψ, such that definition A is satisfied (follows from lemma 1 and 2). The other way round goes as well. If definition A is fulfilled then by lemma 1 and (2), definition B' is satisfied too. We have proved the following statement:

Proposition 8: There exists a reparameterization ϕ of ψ, such that definition A and definition B' are equivalent. $\square$

We note that if $\mathcal{M}$ is a conjugate symmetric space fulfilling definition A' then (v) is satisfied too, since definition A' implies that there exists a reparameterization such that definition A is fulfilled (example 6) and definition A implies definition B'. It means especially that (v) is satisfied.

Finally I will like to point out that if $\beta \neq 0$ then $(\mathcal{M}, i, T) = (\mathcal{M}, i, \beta T)$ since T is just multiplied by a constant factor (see (1)). Put $\tilde{\alpha} = \alpha/\beta$ and let $\overset{\tilde{\alpha}}{G}$ be the connections derived from $(\mathcal{M}, i, \beta T)$ in the sense of (1). Then

$$\overset{\tilde{\alpha}}{G}_{rst} = \overset{0}{G}_{rst} - \frac{\tilde{\alpha}}{2}(\beta T_{rst}) = \overset{0}{G}_{rst} - \frac{\alpha/\beta}{2}(\beta T_{rst}) = \overset{0}{G}_{rst} - \frac{\alpha}{2} T_{rst} = \overset{\alpha}{\Gamma}_{rst}$$

and especially $\overset{1}{G} = \overset{\beta}{\Gamma}$. We can therefore without loss of generality assume $\beta = 1$ if $\beta \neq 0$ in definition A and B'. In this sense we are left with only two cases: $\beta = 1$ and $\beta = 0$. We will however not limit ourself to this simplification from reasons given in the next example.

Example 8: Often β-orthogeodesic statistical models can be considered as submanifolds of a bigger family: e.g. the family of all location-scale t-distributions with ν degrees of freedom $\mathcal{P}^\nu = \{t(\mu, \sigma; \nu) | \mu \in R, \sigma > 0\}$, where μ is the location parameter and σ the scale parameter, is a submanifold of the family of all t-distributions $\mathcal{P} = \{t(\mu, \sigma; \nu) | \mu \in R, \sigma > 0, \nu > 0\}$. It is shown in the example part below that $\mathcal{P}^\nu$ is β-orthogeodesic for some β dependent of ν in the expected induced geometry. It means that we can not rescale the skewness tensor T on $\mathcal{P}$ in any way to make $\mathcal{P}^\nu$ 1-orthogeodesic for all ν in the induced geometries. □

We can also prove a statement that gives a correspondence between β- and $-\beta$-orthogeodesic models:

Proposition 9: Assume $\mathcal{M}$ is β-orthogeodesic in the sense of definition B'. Then $\mathcal{M}$ is also $-\beta$-orthogeodesic.

Proof: If $\beta = 0$ then the statement is trivial. Assume $\beta \neq 0$. Conditions (o)'-(ii)' are independent of β, and are therefore fulfilled. Clearly the condition (v)' is satisfied as well. From (i)', (ii)', and the definition of the Riemannian connection it follows that the submanifold $\mathcal{M}_\chi$ is 0-geodesic, and from Lauritzen (1987) p.188 it follows that $\mathcal{M}_\chi$ is $-\beta$-geodesic, since it is both 0- and β-geodesic. This proofs (iii)'. Finally note from (17) that the induced α-geometries, $\alpha \in R$ makes $\mathcal{M}_\chi$ a statistical manifold, since ${}_\chi\overset{\alpha}{\Gamma} = \overset{\alpha}{\Gamma}(\chi, \cdot)$ and hence

$$
{}_\chi\overset{\alpha}{\Gamma}_{ijk} = \overset{\alpha}{\Gamma}_{ijk}(\chi, \cdot) = \overset{0}{\Gamma}_{ijk}(\chi, \cdot) - \frac{\alpha}{2} T_{ijk}(\chi, \cdot) = {}_\chi\overset{\alpha}{\Gamma}_{ijk} - \frac{\alpha}{2} T_{ijk}(\chi, \cdot).
$$

Since ${}_\chi\overset{\beta}{R} = 0$ it follows from Lauritzen (1987) p.186 that $\mathcal{M}_\chi$ is conjugate symmetric, and therefore ${}_\chi\overset{-\beta}{R} = 0$. It proofs (iv)', and in conclusion $\mathcal{M}$ is $-\beta$-orthogeodesic. □

The Structure Theorem for β-spaces.

In this part we will prove a structure theorem for a subclass of the family of statistical manifolds similar in geometric nature to exponential families. The structure theorem is a generalization of a similar structure theorem given in Barndorff-Nielsen and Bl@esild (1993), but it has the advantage that there is no reference to underlying probability measures.

Let $\mathcal{M}$ be a β-space (see definition 2), and let (θ, η) be a dual system chosen according to proposition 3.

Indices ρ, σ etc will refer to θ in the β-connection and to η in the $-\beta$-connection. If $\mathcal{M}$ is β-orthogeodesic and hence also $-\beta$-orthogeodesic (see proposition 9), then i, j etc refer to the β-flat parameter ψ for $\mathcal{M}_\chi$, and to the $-\beta$-flat parameter π for $\mathcal{M}_\chi$. Moreover indices occur as upper indices in θ and ψ, whereas they occur as lower indices in η and π. Put $\upsilon = (\chi, \pi)$. Generic components of ν are denoted υ_r, υ_s etc, and it means that generic components of χ have lower indices when χ is considered a subparameter of υ and upper indices when considered a subparameter of $\omega = (\chi, \psi)$. Finally when the α-connection symbols are considered as functions of υ they will be denoted $\overset{\alpha}{\Gamma}{}_r^{st}$ and $\overset{\alpha}{\Gamma}{}^{rst}$ (similarly to η).

We are now ready to prove the following structure theorem for β-spaces. Since ψ always can be reparameterized such that definition A is fulfilled if $\mathcal{M}$ is β-orthogeodesic relative to (χ, ψ), we will characterize β-orthogeodesic models in the sense of definition A.

Structure theorem for β-spaces:

Assume $\mathcal{M}$ is a β-space.

Then $\mathcal{M}$ is β-orthogeodesic in the sense of definition A relative to the parameterization $\omega = (\chi, \psi)$ if and only if χ and ψ are variation independent and there exists scalars $\alpha(\psi)$ and $\gamma(\chi)$, vectors $B_\rho(\chi)$ and $D^\rho(\chi)$, matrices $A_\rho^i(\chi)$ and $C_i^\rho(\chi)$, such that the following conditions are satisfied:

(a) $\theta^\rho(\chi, \psi) = \psi^i C_i^\rho(\chi) + D^\rho(\chi)$
(b) $\eta_\rho(\chi, \psi) = \alpha_{/j}(\psi) A_\rho^j(\chi) + B_\rho(\chi)$
(c) $\kappa(\chi, \psi) = \alpha(\psi) + \gamma(\chi) + \psi^i C_i^\rho(\chi) B_\rho(\chi)$
(d) $A_\rho^j(\chi) C_i^\rho(\chi) = \delta_i^j$
(e) $A_\rho^j(\chi) C_{i/a}^\rho(\chi) = 0$
(f) $A_\rho^j(\chi) D_{/a}^\rho(\chi) = 0$
(g) $B_{\rho/a}(\chi) C_i^\rho(\chi) = 0$
(h) $\gamma_{/a}(\chi) = B_\rho(\chi) D_{/a}^\rho(\chi)$.

The vectors $B_\rho(\chi)$ and $D^\rho(\chi)$, matrices $A_\rho^i(\chi)$ and $C_i^\rho(\chi)$ and scalars $\alpha(\psi)$ and $\gamma(\chi)$ are called a β-orthogeodesic representation of $\mathcal{M}$.

Proof: Assume $\mathcal{M}$ is β-orthogeodesic in the sense of definition A relative to the parameterization $\omega = (\chi, \psi)$. Proposition 9 tells us that $\mathcal{M}$ is $-\beta$-orthogeodesic, and let $\upsilon = (\chi, \pi)$ be a corresponding parameterization fulfilling definition A. (iii)-(iv) in definition A implies that

$$0 = \overset{\beta}{\Gamma}{}^{r}_{jk} = \{\overset{\beta}{\Gamma}{}^{\rho}_{\sigma\tau}\, \theta^{\sigma}_{/j}\theta^{\tau}_{/k} + \theta^{\rho}_{/jk}\}\omega^{r}_{/\rho},$$

and (indices are moved)

$$0 = \overset{-\beta}{\Gamma}{}^{jk}_{r} = \{\overset{-\beta}{\Gamma}{}^{\sigma\tau}_{\rho}\, \eta^{/j}_{\sigma}\eta^{/k}_{\tau} + \eta^{/jk}_{\rho}\}\upsilon^{/\rho}_{r}.$$

Since $0 = \overset{\beta}{\Gamma}{}^{\rho}_{\sigma\tau} = \overset{-\beta}{\Gamma}{}^{\sigma\tau}_{\rho}$ (θ and η are $\pm\beta$-flat) one has $0 = \theta^{\rho}_{/jk}\omega^{r}_{/\rho} = \eta^{/jk}_{\rho}\upsilon^{/\rho}_{r}$, or by multiplication with $\theta^{\sigma}_{/r}$ and $\eta^{/r}_{\sigma}$: $0 = \theta^{\rho}_{/jk} = \eta^{/jk}_{\rho}$. We can therefore find matrices and vectors, such that

$$\theta^{\rho}(\chi, \psi) = \psi^{i} C^{\rho}_{i}(\chi) + D^{\rho}(\chi), \tag{21}$$

i.e. (a) is satisfied, and

$$\eta_{\rho}(\chi, \pi) = \pi_{j}\tilde{A}^{j}_{\rho}(\chi) + B_{\rho}(\chi). \tag{22}$$

Inserting $\theta^{\rho}_{/jk} = 0$ in

$$\kappa_{/jk} = \kappa_{/\rho\sigma}\theta^{\rho}_{/j}\theta^{\sigma}_{/k} + \kappa_{/\rho}\theta^{\rho}_{/jk},$$

it follows from (10) and (ii) that

$$\kappa_{/jk} = \kappa_{/\rho\sigma}\theta^{\rho}_{/j}\theta^{\sigma}_{/k} = i_{\rho\sigma}\theta^{\rho}_{/j}\theta^{\sigma}_{/k} = i_{jk}(\psi),$$

i.e. there exists scalars $\alpha(\psi)$ and $\gamma(\chi)$ and a vector $\beta_{i}(\chi)$, such that

$$\kappa(\chi, \psi) = \alpha(\psi) + \gamma(\chi) + \psi^{i}\beta_{i}(\chi). \tag{23}$$

From (9) $\eta_{\rho} = \partial_{\rho}\kappa$ one has $\eta_{\rho} = \partial_{\rho}\kappa = \omega^{r}_{/\rho}\partial_{r}\kappa$, and by multiplication with $\theta^{\rho}_{/s}$

$$\eta_{\rho}\theta^{\rho}_{/s} = \partial_{s}\kappa. \tag{24}$$

Putting $s = i$ it follows ((21), (22), and (23))

$$\{\pi_{j}\tilde{A}^{j}_{\rho}(\chi) + B_{\rho}(\chi)\}C^{\rho}_{i}(\chi) = \pi_{j}\tilde{A}^{j}_{\rho}(\chi)C^{\rho}_{i}(\chi) + B_{\rho}(\chi)C^{\rho}_{i}(\chi) = \alpha_{/i}(\psi) + \beta_{i}(\chi).$$

Differentiating this expression w.r.t. π_{j} one obtains

$$\partial^{j}\alpha_{/i}(\psi(\pi)) = \tilde{A}^{j}_{\rho}(\chi)C^{\rho}_{i}(\chi),$$

and since the left side does not depend on χ it follows that $\tilde{A}^j_\rho(\chi)C^\rho_i(\chi)$ is constant as a function of χ, and therefore $B_\rho(\chi)C^\rho_i(\chi) = \beta_i(\chi) + \alpha_{/i}(\psi) - \pi_j K^j_i$ where $K^j_i = \tilde{A}^j_\rho(\chi)C^\rho_i(\chi)$. Again the left side is independent of ψ and hence $B_\rho(\chi)C^\rho_i(\chi) = \beta_i(\chi)+ K_i$ for some constant K_i. We can assume K_i equals 0, since we can put '$\alpha(\psi) := \alpha(\psi) - \psi^i K_i$' in (23). In that case we obtain from above

$$\alpha_{/i}(\psi) = \pi_j \tilde{A}^j_\rho(\chi)C^\rho_i(\chi).$$

Since $i_{jk} = \alpha_{/jk}(\psi)$ is positive definite (and therefore $det\ \alpha_{/jk}(\psi) \neq 0$) $\phi = \alpha_{/*}(\psi)$ and ψ are in one-to-one correspondence (at least locally), and hence in one-to-one correspondence with π as well (locally). From $\alpha_{/i}(\psi) = \pi_j \tilde{A}^j_\rho(\chi)C^\rho_i(\chi)$ we then conclude, that $det\ \tilde{A}^j_\rho(\chi)C^\rho_i(\chi) \neq 0$, since otherwise the correspondence between ϕ and π will not be one-to-one. Moreover we can find an invertible matrix M^k_j, such that $M^k_j \tilde{A}^j_\rho(\chi)C^\rho_i(\chi) = \delta^k_i$ (as shown above $\tilde{A}^j_\rho(\chi)C^\rho_i(\chi)$ is independent of χ). Put $A^k_\rho(\chi) = M^k_j \tilde{A}^j_\rho(\chi)$ and $\tilde{\pi}_k = \pi_i N^i_k$ with $N = M^{-1}$. It then follows from the above, (22) and (23), that

$$A^j_\rho(\chi)C^\rho_i(\chi) = M^j_k \tilde{A}^k_\rho(\chi)C^\rho_i(\chi) = \delta^j_i, \tag{25}$$

$$\alpha_{/i}(\psi) = \pi_j \tilde{A}^j_\rho(\chi)C^\rho_i(\chi) = \tilde{\pi}_j A^j_\rho(\chi)C^\rho_i(\chi) = \tilde{\pi}_i\ , \tag{26}$$

$$\eta_\rho(\chi,\psi) = \pi_j \tilde{A}^j_\rho(\chi) + B_\rho(\chi) = \alpha_{/j}(\psi)A^j_\rho(\chi) + B_\rho(\chi), \tag{27}$$

and

$$\kappa(\chi,\psi) = \alpha(\psi) + \gamma(\chi) + \psi^i C^\rho_i(\chi)B_\rho(\chi). \tag{28}$$

This proofs (b), (c), and (d).

From (i) in definition A and (10) one has

$$\kappa_{/ia} = \kappa_{/\rho\sigma}\theta^\rho_{/i}\theta^\sigma_{/a} + \kappa_{/\rho}\theta^\rho_{/ia} = i_{ai} + \kappa_{/\rho}\theta^\rho_{/ia} = \kappa_{/\rho}\theta^\rho_{/ia},$$

and using (28), (9), (27), and $\theta^\rho_{/ia} = C^\rho_{i/a}(\chi)$ (see (21)) it follows that

$$C^\rho_{i/a}(\chi)B_\rho(\chi) + C^\rho_i(\chi)B_{\rho/a}(\chi) = \{\alpha_{/j}(\psi)A^j_\rho(\chi) + B_\rho(\chi)\}C^\rho_{i/a}(\chi),$$

or after reduction

$$C^\rho_i(\chi)B_{\rho/a}(\chi) = \alpha_{/j}(\psi)A^j_\rho(\chi)C^\rho_{i/a}(\chi).$$

Assume $A^j_\rho(\chi)C^\rho_{i/a}(\chi) \neq 0$. Since ψ and $\alpha_{/*}(\psi)$ vary in open subsets of R^{d_ψ}, the image of $\alpha_{/*}(\Psi)$ under $A^j_\rho(\chi)C^\rho_{i/a}(\chi)$ will be different from the zero-space. Hence we can find ψ_1 and ψ_2, such that

$$\alpha_{/j}(\psi_1)A^j_\rho(\chi)C^\rho_{i/a}(\chi) \neq \alpha_{/j}(\psi_2)A^j_\rho(\chi)C^\rho_{i/a}(\chi),$$

in contradiction to the fact that $\alpha_{/j}(\psi)A^j_\rho(\chi)C^\rho_{i/a}(\chi) = C^\rho_i(\chi)B_{\rho/a}(\chi)$ is independent of χ. Consequently $A^j_\rho(\chi)C^\rho_{i/a}(\chi) = 0$ and $C^\rho_i(\chi)B_{\rho/a}(\chi) = 0$. It proofs (e) and (g).

Putting $s = a$ in (24) one has, using (21), (27), (28), and (g), that

$$\{\alpha_{/j}(\psi)A^j_\rho(\chi) + B_\rho(\chi)\}\{\psi^i C^\rho_{i/a}(\chi) + D^\rho_{/a}(\chi)\} = \gamma_{/a}(\chi) + \psi^i C^\rho_i(\chi)B_{\rho/a}(\chi),$$

and after a reduction and using (e), that

$$\alpha_{/j}(\psi)A^j_\rho(\chi)D^\rho_{/a}(\chi) = \gamma_{/a}(\chi) - B_\rho(\chi)D^\rho_{/a}(\chi).$$

Using the same argument as above, we conclude that both $A^j_\rho(\chi)D^\rho_{/a}(\chi)$ and $\gamma_{/a}(\chi) - B_\rho(\chi)D^\rho_{/a}(\chi)$ are 0. It proofs (f) and (h), and the proof of the 'only if' part of the theorem is complete.

Assume there exists scalars, vectors, and matrices such that (a)-(h) are fulfilled. We have to prove (o)-(iv) in definition A. (o) is assumed to be valid. From (10) one has

$$i_{rs} = i_{\rho\sigma}\theta^\rho_{/r}\theta^\sigma_{/s} = \kappa_{/\rho\sigma}\theta^\rho_{/r}\theta^\sigma_{/s} = \eta_{\rho/\sigma}\theta^\rho_{/r}\theta^\sigma_{/s} = \eta_{\rho/t}\omega^t_{/\sigma}\theta^\rho_{/r}\theta^\sigma_{/s} = \eta_{\rho/s}\theta^\rho_{/r}. \tag{29}$$

If $r = j$ and $s = a$ (29) becomes ((a), (b), (e), and (f))

$$i_{ja} = \eta_{\rho/j}\theta^\rho_{/a} = \alpha_{/ij}(\psi)A^i_\rho(\chi)\{\psi^k C^\rho_{k/a}(\chi) + D^\rho_{/a}(\chi)\} = 0,$$

and (i) is satisfied. Similarly, if $r = j$ and $s = k$ (29) becomes ((a), (b), and (d))

$$i_{jk} = \eta_{\rho/j}\theta^\rho_{/k} = \alpha_{/ij}(\psi)A^i_\rho(\chi)C^\rho_k(\chi) = \alpha_{/jk}(\psi),$$

and (ii) is satisfied. From (a) one has

$$\overset{\beta}{\Gamma}{}^r_{jk} = \{\overset{\beta}{\Gamma}{}^\rho_{\sigma\tau}\,\theta^\sigma_{/j}\theta^\tau_{/k} + \theta^\rho_{/jk}\}\omega^r_{/\rho} = 0.$$

It proves (iii) and (iv), and the proof of the theorem is completed. □

From the second part of the proof we note, that only (a), (b), (d), (e), and (f) are used.

From (d) and (e) it follows that (e') $A^j_{\rho/a}(\chi)C^\rho_i(\chi) = 0$, and from (29), (a), (b), (e'), and (g) one has

$$i_{aj} = \eta_{\rho/a}\theta^\rho_{/j} = \{\alpha_{/i}(\psi)A^i_{\rho/a}(\chi) + B_{\rho/a}(\chi)\}C^\rho_j(\chi) = 0.$$

It gives us that (a), (b), (d), (e'), and (g) are sufficient to ensure, that $\mathcal{M}$ is β-orthogeodesic in the sense of definition A relative to $\omega = (\chi, \psi)$.

Corollary 1 to the structure theorem:

Assume $\mathcal{M}$ is a β-space.

Then $\mathcal{M}$ is β-orthogeodesic in the sense of definition A relative to the parameterization $\omega = (\chi, \psi)$ if and only if χ and ψ are variation independent and there exists a scalar $\alpha(\psi)$, vectors $B_\rho(\chi)$ and $D^\rho(\chi)$, and matrices $A^i_\rho(\chi)$ and $C^\rho_i(\chi)$, such that either

the conditions (a), (b), (d), (e), and (f) or the conditions (a), (b), (d), (e'), and (g) are satisfied.

Furthermore, if one set of the conditions above are satisfied then so are the rest of the conditions (a)-(g). Moreover λ, the Legendre transform of κ, considered as a function of $\omega = (\chi, \psi)$ is given by

(c') $\quad \lambda(\chi, \psi) = \beta(\psi) + \delta(\chi) + \alpha_{/j}(\psi)A^j_\rho(\chi)D^\rho(\chi),$

where $\beta(\psi) = \psi^j\alpha_{/j}(\psi) - \alpha(\psi) = \overset{\vee}{\alpha}(\psi)$ is the Legendre transform of α, and $\delta(\chi) = D^\rho(\chi)B_\rho(\chi) - \gamma(\chi)$.

Proof: See the remark before the corollary. The last assertion is seen to be valid by insertion in (11). □

Corollary 2 to the structure theorem:

Assume $\mathcal{M}$ is a β-space.

Then $\mathcal{M}$ is β-orthogeodesic in the sense of definition A relative to the parameterization $\omega = (\chi, \psi)$ if and only if $\mathcal{M}$ is $-\beta$-orthogeodesic in the sense of definition A relative to the parameterization $\nu = (\chi, \pi)$ with $\pi = \alpha_{/*}(\psi)$. Moreover the scalars, vectors and matrices can be chosen to be: $\alpha(\pi) := \beta(\psi(\pi))$, $\gamma(\chi) := \delta(\chi)$; $B_\rho(\chi) := D^\rho(\chi)$, $D^\rho(\chi) := B_\rho(\chi)$ and $C^i_\rho(\chi) := A^\rho_i(\chi)$, $A^\rho_j(\chi) := C^j_\rho(\chi)$ (with notation from 'Corollary 1 to the structure theorem').

Proof: Since $\mathcal{M}$ is a β-space, it is a $-\beta$-space too and (η, θ) are $-\beta$-dual coordinates too. From the 'Structure theorem for β-spaces' we have the existence of matrices etc, such that

$$\eta_\rho(\chi, \pi) = \pi_j A^j_\rho(\chi) + B_\rho(\chi),$$

and

$$\theta^\rho(\chi, \pi) = \psi^i(\pi) C^\rho_i(\chi) + D^\rho(\chi),$$

i.e. (a) is satisfied and (b) is satisfied if $\psi^i(\pi) = f^{/i}(\pi)$ for some function f. Since (d)-(g) is satisfied too, it follows from 'Corollary 1 to the structure theorem' that $\mathcal{M}$ is $-\beta$-orthogeodesic relative to the parameterization $\nu = (\chi, \pi)$ if f exists. Put $f(\pi) = \beta(\psi(\pi)) = \beta(\pi) = \psi^i(\pi)\pi_i - \alpha(\psi(\pi))$. Then

$$\beta^{/j}(\pi) = \psi^{i/j}(\pi)\pi_i + \psi^j(\pi) - \alpha_{/i}(\psi)\psi^{i/j}(\pi)$$

$$= \psi^{i/j}(\pi)\pi_i + \psi^j(\pi) - \pi_i\psi^{i/j}(\pi) = \psi^j(\pi),$$

so f fits the demand, and $\mathcal{M}$ is $-\beta$-orthogeodesic relative to the parameterization $\upsilon = (\chi, \pi)$. The proof is now easily completed using symmetry arguments. □

We might now introduce more specific models similarly to those discussed in Barndorff-Nielsen and Bl@esild (1983). It seems however not possibly to keep the distinction between τ-parallel and θ-parallel models in the setting discussed here, since the distinction made in Barndorff-Nielsen and Bl@esild (1983) arises from probabilistic properties of the model, and not from properties originating in geometry. From a geometric point of view there is no difference between θ and η.

Moreover it might be pointed out that the statistical importance of the models discussed in this article do not seem to have the same importance as in the case of exponential families.

Statistical Examples.

This part contains examples of statistical β-orthogeodesic models. The models will be equipped with the expected geometry, since calculations are easily performed in this geometry in contrast to e.g. the observed geometry. Some of these models turn out to be β-spaces as well, but many 'nice' β-orthogeodesic models do not share this property. It will become clear after reading this.

Location-Scale Models: We will focus on location-scale models on the real axes. In Barndorff-Nielsen and Bl@esild (1993) example 3.1 it is shown that all these (under regularity conditions) are 1-orthogeodesic models, and it is easy to see that they are β-orthogeodesic for all $\beta \in R$. But will they be β-spaces too for some $\beta \in R$?

Let $f : R \to R$ be a symmetric density around 0 w.r.t. to some measure λ on R fulfilling: $\lambda(A) = \lambda(-A)$ for all Borel sets $A \subseteq R_+$. The Lebesgue measure on R fulfills this constrain, and in most examples λ will be the Lebesgue measure.

Put $s(x) = \frac{x-\mu}{\sigma}$, $\mu \in R$ and $\sigma > 0$, and let

$$\mathcal{G} = \left\{ \frac{1}{\sigma} f(s(\cdot)) | \mu \in R, \sigma > 0 \right\}$$

be the position-scale model generated from f.

Note that if the support of f is a compact interval (which is symmetric around 0, since f is symmetric), then the densities in $\mathcal{G}$ will have different supports and they will not be equivalent.

Assume f is twice differentiable except perhaps in a finite number of points, and that differentiation w.r.t. σ and μ commutes with integration w.r.t. λ.

Put $g = \log f$ and define $I_{n;k} = I_{n;k}(f)$ by

$$I_{n;k} = \int_J \{1 + yg^{'}(y)\}^n g^{'}(y)^k f(y) \ d\lambda(y),$$

where J is the support of f, and $g^{'}$ means differentiation w.r.t. y. Assume $I_{1;0} = I_{0;1} = 0$ and $|I_{n;k}| < +\infty$ for $(n;k) = (2;0)$, $(n;k) = (0;2)$, $(n;k) = (3;0)$, $(n;k) = (1;2)$.

With the above assumptions we have the following expressions for i, $\overset{\alpha}{\Gamma}$ and $\overset{\alpha}{R}$.

$$i_{\sigma\sigma} = \frac{1}{\sigma^2} I_{2;0} \quad i_{\sigma\mu} = 0 \quad i_{\mu\mu} = \frac{1}{\sigma^2} I_{0;2} \tag{30}$$

$$\overset{\alpha}{\Gamma}_{\sigma\sigma\sigma} = -\frac{1}{\sigma^3} \left\{ I_{2;0} - \frac{\alpha}{2} I_{3;0} \right\} \quad \overset{\alpha}{\Gamma}_{\sigma\sigma\mu} = \overset{\alpha}{\Gamma}_{\mu\sigma\sigma} = \overset{\alpha}{\Gamma}_{\sigma\mu\sigma} = \overset{\alpha}{\Gamma}_{\mu\mu\mu} = 0 \tag{31}$$

$$\overset{\alpha}{\Gamma}_{\sigma\mu\mu} = \overset{\alpha}{\Gamma}_{\mu\sigma\mu} = -\frac{1}{\sigma^3} \left\{ I_{0;2} - \frac{\alpha}{2} I_{1;2} \right\} \quad \overset{\alpha}{\Gamma}_{\mu\mu\sigma} = \frac{1}{\sigma^3} \left\{ I_{0;2} + \frac{\alpha}{2} I_{1;2} \right\}$$

$$\overset{\alpha}{\Gamma}{}^{\sigma}_{\sigma\sigma} = -\frac{1}{\sigma} \left\{ 1 - \frac{\alpha}{2} \frac{I_{3;0}}{I_{2;0}} \right\} \quad \overset{\alpha}{\Gamma}{}^{\mu}_{\sigma\sigma} = \overset{\alpha}{\Gamma}{}^{\sigma}_{\mu\sigma} = \overset{\alpha}{\Gamma}{}^{\sigma}_{\sigma\mu} = \overset{\alpha}{\Gamma}{}^{\mu}_{\mu\mu} = 0 \tag{32}$$

$$\overset{\alpha}{\Gamma}{}^{\mu}_{\sigma\mu}=\overset{\alpha}{\Gamma}{}^{\mu}_{\mu\sigma}=-\frac{1}{\sigma}\left\{1-\frac{\alpha}{2}\frac{I_{1;2}}{I_{0;2}}\right\} \qquad \overset{\alpha}{\Gamma}{}^{\sigma}_{\mu\mu}=\frac{1}{\sigma}\left\{\frac{I_{0;2}}{I_{2;0}}+\frac{\alpha}{2}\frac{I_{1;2}}{I_{2;0}}\right\}$$

$$T_{\sigma\sigma\sigma}=-\frac{1}{\sigma^3}I_{3;0} \quad T_{\sigma\mu\mu}=-\frac{1}{\sigma^3}I_{1;2} \quad T_{\sigma\sigma\mu}=T_{\mu\mu\mu}=0 \tag{33}$$

$$\overset{\alpha}{R}_{\mu\sigma\sigma\mu}=-\frac{1}{\sigma^4}\left\{I_{0;2}+\left(\frac{I_{3;0}}{I_{2;0}}I_{0;2}-2I_{1;2}\right)\frac{\alpha}{2}+\left(\frac{I_{1;2}^2}{I_{0;2}}-\frac{I_{3;0}}{I_{2;0}}I_{1;2}\right)\frac{\alpha^2}{4}\right\} \tag{34}$$

$$\overset{\alpha}{R}_{\mu\sigma\mu\sigma}=\frac{1}{\sigma^4}\left\{I_{0;2}-\left(\frac{I_{3;0}}{I_{2;0}}I_{0;2}-2I_{1;2}\right)\frac{\alpha}{2}+\left(\frac{I_{1;2}^2}{I_{0;2}}-\frac{I_{3;0}}{I_{2;0}}I_{1;2}\right)\frac{\alpha^2}{4}\right\}$$

$$\overset{\alpha}{R}_{\sigma\mu\sigma\mu}=-\overset{\alpha}{R}_{\mu\sigma\sigma\mu} \quad \overset{\alpha}{R}_{\sigma\mu\mu\sigma}=-\overset{\alpha}{R}_{\mu\sigma\mu\sigma} \quad \overset{\alpha}{R}_{rstu}=0 \text{ otherwise.}$$

We note the following lemmas:

Lemma 3: $\mathcal{G}$ is a conjugate symmetric space if and only if

$$\frac{I_{3;0}}{I_{2;0}}I_{0;2}-2I_{1;2}=0. \tag{35}$$

Proof: A conjugate symmetric space fulfills per definition $\overset{\alpha}{R}=\overset{-\alpha}{R}$ for all $\alpha\in R$. The lemma then follows from (34). □

Lemma 4: If there exists $\beta\in R$: $\overset{\beta}{R}=0$ then

$$\frac{I_{1;2}^2}{I_{0;2}}-\frac{I_{3;0}}{I_{2;0}}I_{1;2}<0. \tag{36}$$

Assume (35) and (36). Then $\overset{\beta}{R}=0$ is satisfied with

$$\beta=\pm 2\frac{I_{0;2}}{I_{1;2}}. \tag{37}$$

Proof: Note that $\beta\neq 0$ since $I_{0;2}>0$ (see (34)). If $\overset{\beta}{R}=0$ then $\mathcal{G}$ is conjugate symmetric and (35) is valid. Since $I_{2;0}>0$ (the Fisher information is assumed positive definite) then the first part of the lemma is a consequence of (34). From (34) using (35) we see that β satisfy

$$0 = I_{0;2} + \left(\frac{I_{1;2}^2}{I_{0;2}} - \frac{I_{3;0}}{I_{2;0}} I_{1;2}\right) \frac{\beta^2}{4} = I_{0;2} + \frac{I_{1;2}}{I_{0;2}} \left(I_{1;2} - \frac{I_{3;0}}{I_{2;0}} I_{0;2}\right) \frac{\beta^2}{4}.$$

Adding and subtracting $I_{1;2}$ inside the brackets and using (35) yields

$$0 = I_{0;2} - \frac{I_{1;2}^s}{I_{0;2}} \frac{\beta^2}{4}$$

and the result follows easily. $\square$

Lemma 4 can be reformulated in a more informative way. Note first since $I_{0;2} > 0$ that condition (36) can be rewritten as

$$\frac{I_{3;0}}{I_{2;0}} I_{0;2} I_{1;2} - I_{1;2}^2 > 0.$$

Since $I_{1;2}$ occurs in both terms the condition can only be fulfilled if $I_{1;2} \neq 0$. Opposite to that (see (35))

$$\text{if } \; I_{1;2} > 0 \; \text{ then } \; \frac{I_{3;0}}{I_{2;0}} I_{0;2} - I_{1;2} > 0,$$

and

$$\text{if } \; I_{1;2} < 0 \; \text{ then } \; \frac{I_{3;0}}{I_{2;0}} I_{0;2} - I_{1;2} < 0,$$

or if we combine these two statements:

$$\text{if } \; I_{1;2} \neq 0 \; \text{ then } \; \frac{I_{3;0}}{I_{2;0}} I_{0;2} I_{1;2} - I_{1;2}^2 > 0.$$

I.e. we have shown:

Lemma 5: If there exists $\beta \in R$: $\overset{\beta}{R} = 0$ then $I_{1;2} \neq 0$. And if (35) is satisfied then the opposite implication is valid too. $\square$

Let us remark that the class of location-scale models fulfilling the assumptions made in the beginning is now divided into three subclasses, three types:

(1) Models which do not satisfy (35).
(2) Models which satisfy (36) but not (35).
(3) Models which satisfy (36) and (35).

None of these types is empty and concrete examples of each type will be given below.

Assume now $\overset{\beta}{R}=0$. Note from (35) that $\frac{I_{3;0}I_{0;2}}{I_{2;0}I_{1;2}}=2$, and therefore with

$$\beta=-2\frac{I_{0;2}}{I_{1;2}}$$

we have (32)

$$\overset{\beta}{\Gamma}{}^{\sigma}_{\sigma\sigma}=-\frac{3}{\sigma} \quad \text{and} \quad \overset{\beta}{\Gamma}{}^{\mu}_{\mu\sigma}=\overset{\beta}{\Gamma}{}^{\mu}_{\sigma\mu}=-\frac{2}{\sigma}. \tag{38}$$

The other symbols β-connection symbols are 0. Putting $\sigma=\psi^{-1/2}$ one obtains using the transformation law for connection symbols

$$\overset{\beta}{\Gamma}{}^{\psi}_{\psi\psi}=0 \quad \text{and} \quad \overset{\beta}{\Gamma}{}^{\mu}_{\mu\psi}=\overset{\beta}{\Gamma}{}^{\mu}_{\psi\mu}=\frac{1}{\psi}. \tag{39}$$

The remaining symbols are still 0. Similarly, the $-\beta$-connection symbols are in (μ,ψ)-coordinates given by

$$\overset{-\beta}{\Gamma}{}^{\psi}_{\psi\psi}=-\frac{2}{\psi} \quad \text{and} \quad \overset{-\beta}{\Gamma}{}^{\psi}_{\mu\mu}=-\frac{4}{\psi^2}\frac{I_{0;2}}{I_{2;0}}, \tag{40}$$

and the remaining are 0.

When $\mathcal{G}$ is a β-space we will find β-dual coordinates and express these in terms of the β-orthogeodesic coordinates $\omega=(\mu,\psi)$ as described in the 'Structure theorem for β-spaces'. Generally we shall solve the equations

$$\overset{\beta}{\Gamma}{}^{r}_{st}=\{\overset{\beta}{\Gamma}{}^{\rho}_{\sigma\tau}\ \theta^{\sigma}_{/s}\theta^{\tau}_{/t}+\theta^{\rho}_{st}\}\omega^{r}_{/\rho}=\theta^{\rho}_{/st}\omega^{r}_{/\rho},$$

since $\overset{\beta}{\Gamma}{}^{\rho}_{\sigma\tau}=0$ in β-flat coordinates θ. We have similar equations in η- and ω-coordinates using the $-\beta$-connection. These are however often hard to solve and in concrete examples it can be of great help to have a qualified guess. There is no uniqueness since affine transformations of θ still will be a 1-flat parameter. Here it obvious that we can choose $\theta=\psi(\mu,-1/2)$ since as seen from (39) upper β-connection symbols are independent of the model considered, and we know that the chosen θ works for the normal distribution. Here we can choose

$$\alpha(\psi)=-\frac{1}{4}I_{2;0}\log\psi \qquad \gamma(\mu)=0 \tag{41}$$

$$A(\mu) = (0, -2) \quad C(\mu) = (\mu, -1/2) \quad B(\mu) = I_{0;2}(\mu, \mu^2) \quad D(\mu) = (0, 0)$$

$$\theta(\mu, \psi) = \psi(\mu, -1/2) \quad \eta(\mu, \psi) = -\frac{1}{4\psi} I_{2;0}(0, -2) + I_{0;2}(\mu, \mu^2) \tag{42}$$

$$\kappa(\mu, \psi) = -\frac{1}{4} I_{2;0} \log \psi + \frac{1}{2} I_{0;2} \psi \mu^2 \tag{43}$$

and

$$\lambda(\mu, \psi) = \frac{1}{4} I_{2;0} \log \psi - \frac{1}{4} I_{2;0}.$$

All the models discussed here are however dual spaces. To see this first transform μ into $\mu = \sqrt{2I_1/I_2}\nu$. Then we have the following expression for i in (ν, σ)-coordinates:

$$i_{\sigma\sigma} = \frac{1}{\sigma^2} I_1, \quad i_{\sigma\nu} = 0, \quad i_{\nu\nu} = \frac{2}{\sigma^2} I_1,$$

i.e. a scalar multiplum of the metric expressed in (μ, σ) on the family of one dimensional normal distributions. Since the dual coordinates of $\mathcal{N}$ can be chosen to be (the canonical parameter and the mean value parameter)

$$\theta(\nu, \sigma) = \frac{1}{\sigma^2}(\mu, -1/2), \quad \text{and} \quad \eta(\nu, \sigma) = \sigma^2(0, 1) + (\mu, \mu^2),$$

then we can choose a dual system on $\mathcal{M}$ in the following way

$$\theta(\nu, \sigma) = \frac{1}{I_1\sigma^2}(\mu, -1/2), \quad \text{and} \quad \eta(\nu, \sigma) = \sigma^2(0, 1) + (\mu, \mu^2).$$

If we make a transformation of the densities similar to the transformation of μ, $x = \sqrt{2I_1/I_2}y$, then (ν, σ) still have interpretion as a location and a scale parameter.

As mentioned above the class of location-scale models fulfilling the assumptions made in the beginning is divided into three types:

(1) Models which do not satisfy (35).
(2) Models which satisfy (36) but not (35).
(3) Models which satisfy (36) and (35).

I will no give some concrete examples showing that none of these classes are empty, and that models which in some sense are quite similar to the class of normal distributions may have quite poor geometrical properties.

The spherical distribution: Let f_ν be the density w.r.t. Lebesque measure on $(-1,1)$ of the spherical distribution $Sph(\nu)$ with parameter $\nu > -1$, i.e.

$$f_\nu(y) = \frac{\Gamma(1+\nu/2)}{\sqrt{\pi}\Gamma(1/2+\nu/2)}(1-y^2)^{(\nu-1)/2}, \quad y \in (-1,1)$$

The assumptions concerning the density function f_ν and the integrals $I_{n;k}$ mentioned in the beginning of this section are satisfied for $\nu > 5$, and from formula 3.251.1 in Gradshteyn and Ryzhik (1980),

$$\frac{1}{\lambda}B(\frac{\mu}{\lambda}) = \int_0^1 x^{\mu-1}(1-x^\lambda)^{\nu-1}\,dx \quad \mu, \nu\lambda > 0,$$

we have

$$I_{1;0} = I_{0;1} = 0 \quad I_{2;0} = \frac{2\nu}{\nu-3} \quad I_{0;2} = \frac{\nu(\nu-1)}{\nu-3}$$

$$I_{3;0} = -\frac{8\nu(\nu+1)}{(\nu-3)(\nu-5)} \quad I_{1;2} = -\frac{2\nu(\nu^2-1)}{(\nu-3)(\nu-5)}.$$

If $\nu > 5$ then $I_{1;2} \neq 0$ (< 0), and

$$\frac{I_{3;0}}{I_{2;0}}I_{0;2} - 2I_{1;2} = 0.$$

It means both (35) and (36) are satisfied, and the families $\mathcal{G}_\nu$ are

$$\pm\frac{\nu-5}{\nu+1} - \text{spaces},$$

and thus examples of type (3) models. □

A class of error distributions: Let f_ν be the density $E(\nu)$, $\nu > 0$ (E stands for error) w.r.t. the Lebesgue measure on R given by

$$f_\nu(y) = \nu\frac{1}{\Gamma(1/(2\nu))}\exp\{-(y^2)^\nu\}, \quad y \in R.$$

Note if $\nu = 1$ then f_1 is the normal density with mean 0 and variance 1/2. The assumptions concerning the density function f_ν and the integrals $I_{n;k}$ mentioned in the beginning of this section are satisfied for $\nu > 1/2$, and from formula 3.478.1 in Gradshteyn and Ryzhik (1980),

$$\frac{1}{p}\Gamma(\frac{q}{p}) = \int_0^\infty x^{q-1}\exp(-x^p)\,dx \quad q,p > 0,$$

we have

$$I_{1;0} = I_{0;1} = 0 \quad I_{0;2} = 2\nu \quad I_{3;0} = -8\nu^2$$
$$I_{2;0} = 4\nu^2\frac{\Gamma(2-1/(2\nu))}{\Gamma(1/(2\nu))} \quad I_{1;2} = -8\nu^2(2\nu-1)\frac{\Gamma(2-1/(2\nu))}{\Gamma(1/(2\nu))}.$$

Moreover it is seen that

$$\frac{I_{3;0}}{I_{2;0}}I_{0;2} - 2I_{1;2} = 16\nu^2(\nu-1)\frac{\Gamma(2-1/(2\nu))}{\Gamma(1/(2\nu))}.$$

Consequently, lemma 1 implies that the family $\mathcal{G}_\nu$ is only a conjugate symmetric space if $\nu = 1$ (the family generated by $N(0,1/2)$), and hence also a β-space only if $\nu = \beta = 1$. Thus the family $E(\nu)$ ($\nu \neq 1$) is an example of models of type (1). $\square$

The t-distribution: Let f_ν be density of the t-distribution with $\nu > 0$ degrees of freedom, i.e.

$$f(y) = \frac{1}{\sqrt{\nu}B(1/2,1/2\nu)}\left\{1+\frac{y^2}{\nu}\right\}^{-1/2(\nu+1)},$$

The assumptions concerning the density function f_ν and the integrals $I_{n;k}$ mentioned in the beginning of this section are satisfied for all $\nu > 0$, and from formula 8.380.3 in Gradshteyn and Ryzhik (1980),

$$B(x,y) = 2\int_0^\infty \frac{t^{2x-1}}{(1+t^2)^{x+y}}\,dt \quad x,y > 0,$$

we have

$$I_{1;0} = I_{0;1} = 0 \quad I_{2;0} = \frac{2\nu}{\nu+3} \quad I_{0;2} = \frac{\nu+1}{\nu+3}$$
$$I_{3;0} = -\frac{8\nu(\nu-1)}{(\nu+3)(\nu+5)} \quad I_{1;2} = -\frac{2(\nu^2-1)}{(\nu+3)(\nu+5)}.$$

For all $\nu > 0$ we have

$$\frac{I_{3;0}}{I_{2;0}} I_{0;2} - 2I_{1;2} = 0,$$

and all families $\mathcal{G}_\nu$ are therefore conjugate symmetric. It is seen too, that $I_{1;2} \neq 0$ if and only if $\nu \neq 1$, i.e. if $\nu \neq 1$ then the families $\mathcal{G}_\nu$ are

$$\pm \frac{\nu+5}{\nu-1} - \text{spaces},$$

and of type (3). If $\nu = 1$ (the Cauchy-distribution) then the generated family is an example of type (2). Note too if ν tends to infinity then $\pm\frac{\nu+5}{\nu-1}$ tends to ± 1, which is consistent with the fact that the family of normal distributions is ± 1-flat. $\square$

The t-distribution revisited: Let $\mathcal{P}$ be the location-scale family corresponding to the family of all t-distributions, i.e.

$$\mathcal{P} = \{t(\mu, \sigma; \nu) | \mu \in R, \sigma > 0, \nu > 0\},$$

where $t(\mu, \sigma; \nu)$ is the density given by

$$t(\mu, \sigma; \nu)(x) = \frac{1}{\sigma\sqrt{\nu} B(1/2, 1/2\nu)} \left\{ 1 + \frac{(x-\mu)^2}{\nu\sigma^2} \right\}^{-1/2(\nu+1)}.$$

From the previous example, calculations and formula 8.380.3 in Gradshteyn and Ryzhik (1980) we have the following expressions for the Fisher information metric on $\mathcal{P}$:

$$i_{\sigma\sigma} = \frac{2\nu}{\sigma^2(\nu+3)} \quad i_{\mu\sigma} = 0 \quad i_{\mu\mu} = \frac{\nu+1}{\sigma^2(\nu+3)} \tag{44}$$

$$i_{\sigma\nu} = -\frac{2}{\sigma(\nu+1)(\nu+3)} \quad i_{\mu\nu} = 0 \quad i_{\nu\nu} = i_{\nu\nu}(\nu),$$

where $i_{\nu\nu}(\nu)$ is some function only dependent of ν involving derivatives of $B(1/2, 1/2\nu)$. The specific form of $i_{\nu\nu}(\nu)$ is not important. Let $\tilde{\tau} : R \times R_+ \times R_+ \to R \times R_+ \times R_+$ be the reparametrization of $\mathcal{P}$ defined by

$$\tilde{\tau}(\mu, \sigma, \nu) = (\mu, \tau(\sigma, \nu), \nu)$$

and

$$\tau(\sigma, \nu) = \sigma \frac{\nu+1}{\nu}.$$

We then have that

$$\sigma(\tau,\nu) = \tau\frac{\nu}{\nu+1}$$

and

$$i_{\nu\tau} = i_{\nu\sigma}\sigma_{/\tau} + i_{\sigma\sigma}\sigma_{/\nu}\sigma_{/\tau} = -\frac{2}{\sigma(\nu+1)(\nu+3)}\sigma_{/\tau} + \frac{2\nu}{\sigma^2(\nu+3)}\sigma_{/\nu}\sigma_{/\tau} = 0.$$

In the (μ,τ,ν)-parameter i is given by

$$i_{\tau\tau} = \frac{2\nu}{\tau^2(\nu+3)} \quad i_{\mu\tau} = 0 \quad i_{\mu\mu} = \frac{(\nu+1)^3}{\tau^2\nu^2(\nu+3)} \tag{45}$$

$$i_{\tau\nu} = 0 \quad i_{\mu\nu} = 0 \quad i_{\nu\nu} = i_{\nu\nu}(\nu),$$

i.e. i is diagonal in $\tilde{\tau}$-coordinates. Furthermore since τ for ν fixed just is a rescalation of σ then (μ,τ) parametrize $\mathcal{P}_\nu$ for all ν. This means that the expected geometry on $\mathcal{P}_\nu$ in (μ,τ)-coordinates is given by

$${}_{\nu}\overset{\alpha}{\Gamma}_{rst} = \overset{\alpha}{\Gamma}_{rst} \tag{46}$$

for $r,s,t \in \{\mu,\tau\}$. Again since the transformation $(\mu,\tau) \mapsto (\mu,\sigma)$ is a rescalation of σ then (μ,τ) is a β-orthogeodesic parametrization of $\mathcal{P}_\nu$ for some β dependent of ν, because (μ,σ) is a β-orthogeodesic parametrization (see previous example and (46)). Moreover $\mathcal{P}_\nu$ is β-flat $(\nu \neq 1)$ in the induced geometry (see previous example).

This means that $\mathcal{P}$ is a foliation of β-flat β-orthogeodesic leaves $\mathcal{P}_\nu$ (except for $\nu \neq 1$). As seen from (45) $\mathcal{P}$ is not β-orthogeodesic for any β relative to the parameter $(\chi,\psi) = (\nu,(\mu,\tau))$ since the (μ,τ)-part of the metric depends on ν.

In (μ,τ)-coordinates the densities are given by

$$t(\mu,\tau;\nu)(x) = \frac{\nu+1}{\tau\nu^{3/2}B(1/2,1/2\nu)}\left\{1+\frac{(\nu+1)^2(x-\mu)^2}{\nu^3\tau^2}\right\}^{-1/2(\nu+1)}.$$

Note that the families $\mathcal{P}_\nu$ in (μ,τ)-coordinates are location-scale models too. □

References.

Amari, S.-I. (1985): Differential-Geometric Methods in Statistics. Lecture Notes in

Statistics 28. Heidelberg: Springer-Verlag.

Barndorff-Nielsen, O.E. and P.Bl@esild (1983): Exponential Models with Affine Dual Foliations. Ann. Statist., 11 753-769.

Barndorff-Nielsen, O.E. and P.Bl@esild (1993): Orthogeodesic Models. Ann. Statist., 21 1018-1039.

Bl@esild, P. (1991): Yokes and Tensors Derived from Yokes. Ann. Inst. Statist. Math., Vol 43 no.1.

Lauritzen, S.L. (1983): Statistical Manifolds. In Differential Geometry in Statistical Inference. Lecture Notes-Monograph Series. Hayward: Institute of Mathematical Statistics.

Gradshteyn, I.S. and I.M.Ryzhik (1980): Tables of Integrals, Series and Products. New York: Academic Press.