Geometry of quantum states: dual connections and divergence functions

Anna Jenčová
Mathematical Institute, Slovak Academy of Sciences
Štefánikova 49, SK–814 73 Slovakia

Abstract. In a finite quantum state space with a monotone metric, a family of torsion-free affine connections is introduced, in analogy with the classical α-connections defined by Amari. The dual connections with respect to the metric are found and it is shown that these are, in general, not torsion-free. The torsion and the Riemannian curvature are computed and the existence of efficient estimators is treated. Finally, geodesics are used to define a divergence function.

1 The classical case

Let $\mathcal{S} = \{p(x,\theta) \mid \theta \in \Theta \subseteq R^m\}$ be a smooth family of classical probability distributions on a sample space $\mathcal{X}$. Then $\mathcal{S}$ can naturally be viewed as a differentiable manifold. The differential-geometrical aspects of a statistical manifold and their statistical implications were studied by many authors. The Riemannian structure is given by the Fisher information metric tensor

$$g_{ij}(\theta) = E_\theta[\partial_i \log p(x,\theta)\partial_j \log p(x,\theta)]$$

where ∂_i denotes $\frac{\partial}{\partial\theta_i}$. In 1972, Chentsov in [6] introduced a family of affine connections in $\mathcal{S}$ and proved that the Fisher information and these connections were unique (up to a constant factor) in the manifold of distributions on a finite number of atoms, in the sense that these are invariant with respect to transformations of the sample space. In [1], Amari defined a one-parameter family of α-connections in $\mathcal{S}$, which turned out to be the same as defined by

Chentsov. These may be introduced using the following α-representations of the tangent space:

Let g_α be a one-parameter family of functions, given by

$$g_\alpha(x) = \begin{cases} \frac{2}{1-\alpha} x^{\frac{1-\alpha}{2}} & \alpha \neq 1 \\ \log x & \alpha = 1 \end{cases}$$

Let $l_\alpha(x, \theta) = g_\alpha(p(x, \theta))$. The vector space spanned by the functions $\partial_i l_\alpha(x, \theta)$, $i = 1, \ldots, p$ is called the α-representation of the tangent space. The metric tensor g_{ij} is then

$$g_{ij}(\theta) = \int \partial_i l_\alpha(x, \theta) \partial_j l_{-\alpha}(x, \theta) dP$$

The α-connections are defined by

$$\Gamma^\alpha_{ijk}(\theta) = \int \partial_i \partial_j l_\alpha(x, \theta) \partial_k l_{-\alpha}(x, \theta) dP$$

From this, it is clear that these connections are torsion-free, i. e. $S^\alpha_{ijk} = \Gamma^\alpha_{ijk} - \Gamma^\alpha_{jik} = 0$, $\forall i, j, k$, $\forall \theta$. Let now ∇ and ∇^* be two covariant derivatives on $\mathcal{S}$, then we say that the covariant derivatives (the affine connections) are dual with respect to the metric if

$$Xg(Y, Z) = g(\nabla_X Y, Z) + g(Y, \nabla^*_X Z)$$

or, in coefficients,

$$\partial_i g_{jk} = \Gamma_{ijk} + \Gamma^*_{ikj}$$

It is easy to see that the α and $-\alpha$ connections are mutually dual.

Further, let η be another coordinate system in $\mathcal{S}$. The natural basis of the tangent space T_P at $P \in \mathcal{S}$ is $\{\partial_i\}$, $\partial_i = \frac{\partial}{\partial \theta_i}$ for the coordinate system θ and $\{\partial^i\}$, $\partial^i = \frac{\partial}{\partial \eta_i}$ for η. We say that θ and η are mutually dual if their natural bases are biorthogonal, i.e. if

$$g(\partial_i, \partial^j) = \delta_i^j$$

The metric tensor in the basis $\{\partial^i\}$ is given by

$$g(\partial^i, \partial^j) = g^{ij}, \qquad (g^{ij}) = (g_{ij})^{-1}$$

The next Theorem (the Theorem 3.4 in [1]) gives the necessary and sufficient condition for existence of such pair of coordinate systems.

Theorem 1.1 *If a Riemannian manifold $\mathcal{S}$ has a pair of dual coordinate systems (θ, η), then there exist potential functions $\psi(\theta)$ and $\phi(\eta)$ such that*

$$g_{ij} = \partial_i \partial_j \psi(\theta), \qquad g^{ij} = \partial^i \partial^j \phi(\eta) \tag{1}$$

Conversely, if either potential function ψ or ϕ exist such that (1) holds, there exists a pair of dual coordinate systems. The dual coordinate systems are related by the Legendre transformations

$$\theta_i = \partial^i \phi(\eta) \qquad \eta_i = \partial_i \psi(\theta)$$

and the two potential functions satisfy the identity

$$\psi(\theta) + \phi(\eta) - \sum_i \theta_i \eta_j = 0 \tag{2}$$

The most interesting results of [1] concern the case that the manifold $\mathcal{S}$ is α-flat, i.e. the Riemannian curvature tensor of the α-connection vanishes. Then $\mathcal{S}$ is also $-\alpha$-flat. It is known that for flat manifolds, an affine coordinate system exists, i.e. the coefficients of the connection vanish. The next Theorem (Theorem 3.5 in [1]) reveals the dualistic structure of α-flat manifolds.

Theorem 1.2 *When a Riemannian manifold S is flat with respect to a pair of torsion-free dual affine connections ∇ and ∇^*, there exists a pair (θ, η) of dual coordinate systems such that θ is a ∇-affine and η is a ∇^*-affine coordinate system.*

This result can be directly applied to the exponential and mixture families

$$p(x, \theta) = \exp\{\sum_i \theta_i c_i(x) - \psi(\theta)\}$$

and

$$p(x, \theta) = \sum_i \theta_i c_i(x) + (1 - \sum_i \theta_i) c_{n+1}(x)$$

which are ± 1-flat, and the extended α-families

$$l_\alpha(x, \theta) = \sum_i^{n+1} \theta_i c_i(x)$$

which are $\pm\alpha$-flat (note that the extended α-families are not normed to 1).

Let us consider an α-flat family $\mathcal{S}$ with the dual coordinate systems (θ, η) and let $\psi(\theta)$ and $\phi(\eta)$ be the potential functions. In [1], a divergence is introduced in $\mathcal{S}$. It is called the α-divergence and it is given by

$$D_\alpha(\theta, \theta') = \psi(\theta) + \phi(\eta') - \sum_i \theta_i \eta'_i$$

The divergence is not a usual distance, but it has some important properties:

(i) it is strictly positive, $D_\alpha(\theta, \theta') \geq 0$ with $D_\alpha(\theta, \theta') = 0$ iff $\theta = \theta'$

(ii) it is jointly convex in θ and θ'

(iii) $D_\alpha(\theta, \theta') = D_{-\alpha}(\theta', \theta)$

(iv) it satisfies the relation

$$D_\alpha(\theta, \theta + d\theta) = D_\alpha(\theta + d\theta, \theta) = \frac{1}{2} \sum_{ij} g_{ij}(\theta) d\theta_i d\theta_j$$

hence it induces the Riemannian distance, given by the Fisher information.

Moreover, the divergences are shown to satisfy a generalized Pythagorean relation.

There were some attempts to introduce a similar family of affine structures in the differentiable manifold of states of an n-level quantum system with a monotone metric. The main difficulty here is that, as was first shown in [7], there is no unique quantum analogue of the Fisher information metric and, moreover, the connections are in general not torsion-free. Hasegawa in [9] studied the case of quantum exponential and mixture families with the Kubo-Mori metric. In [13], the exponential and mixture connections (i.e. the case $\alpha = \pm 1$) were defined also for arbitrary monotone metric and a divergence function was introduced. The aim of this paper is to use a similar method to define the α-connections and divergence functions for each α and to investigate the dualistic properties of the manifold.

2 The state space

Let $\mathcal{M}$ denote the differentiable manifold of all n-dimensional complex hermitean matrices and let $\mathcal{M}^+ = \{M \in \mathcal{M} \mid M > 0\}$. Let $\tilde{T}_M$ be the tangent space at M, then $\tilde{T}_M$ can be identified with $\mathcal{M}$ considered as a vector space. We introduce a Riemannian structure in $\mathcal{M}^+$, defining an inner product in $\tilde{T}_M$ by

$$\lambda_M(X,Y) = \operatorname{Tr} X J_M(Y)$$

where J_M is a suitable superoperator on matrices. The state space of an n-level quantum system can be identified with the submanifold

$$\mathcal{D} = \{D \in \mathcal{M}^+ \mid \operatorname{Tr} D = 1\}$$

The tangent space $T_D \subset \tilde{T}_D$ is the real vector space of all self-adjoint traceless matrices. If we consider the restriction of the Riemannian structure λ onto $\mathcal{D}$, it is natural to require that λ is monotone, in the sense that if T is a stochastic map, then

$$\lambda_{T(M)}(T(X), T(X)) \leq \lambda_M(X,X), \ \forall M \in \mathcal{M}^+, X \in \mathcal{M}$$

As it was proved in [14], this is true iff J_M is of the form

$$J_M = (R_M^{\frac{1}{2}} f(L_M R_M^{-1}) R_M^{\frac{1}{2}})^{-1} \tag{3}$$

where $f : R^+ \to R$ is an operator monotone function such that $f(t) = tf(t^{-1})$ for every $t > 0$ and $L_M(A) = MA$ and $R_M(A) = AM$, for each matrix A. We also adopt a normalization condition $f(1) = 1$. Notice that we have $J_M(X) = M^{-1}X$ whenever X and M commute, in particular, $J_M(M) = I$.

Let T_D^* be the cotangent space of $\mathcal{D}$ at D, then T_D^* is the vector space of all observables A with zero mean at D (i.e. $\operatorname{Tr} DA = 0$). It is easy to see that

$$T_D^* = \{J_D(H) \mid H \in T_D\} \tag{4}$$

The metric λ induces an inner product in T_D^*, namely,

$$\varphi(A_1, A_2) = \lambda(J_D^{-1}(A_1), J_D^{-1}(A_2)) = \operatorname{Tr} A_1 J_D^{-1}(A_2)$$

It can be interpreted as a generalized covariance of the observables A_1 and A_2.

Example 2.1 Let the metric be determined by $J_D(H) = G$, where $GD + DG = 2H$, then it is called the metric of the symmetric logarithmic derivative. This metric is monotone, with the corresponding operator monotone function $f(x) = \frac{1+x}{2}$, see [5, 11, 15].

Example 2.2 Another important example of a monotone metric is the well-known Kubo-Mori metric determined by $J_D(H) = \frac{d}{dt}\log(D + tH)|_{t=0}$. The monotone metric is given by

$$\lambda_D(H, K) = \mathrm{Tr}\, H J_D(K) = \frac{\partial^2}{\partial t \partial s}\mathrm{Tr}\,(D + tH)\log(D + sK)|_{t,s=0}$$

Note that this metric is induced by the relative entropy $D(\rho, \sigma) = \mathrm{Tr}\,\rho(\log\rho - \log\sigma)$ when the density matrices ρ and σ are infinitesimally distant from each other.

Example 2.3 [11, 16] Let $J_D(H) = \frac{1}{2}(D^{-1}H + HD^{-1})$. The corresponding metric is monotone, with $f(x) = \frac{2x}{x+1}$, and it is called the metric of the right logarithmic derivative.

For more about monotone metrics and their use see [15, 16, 10].

3 The g-representation

Let $g : R \to R$ be a smooth (strictly) monotone function. We define an operator $L_g[M] : \tilde{T}_M \to \tilde{T}_{g(M)}$ by

$$L_g[M](H) = \frac{d}{ds}g(M + sH)|_{s=0}$$

The following Lemma was proved in [17]. As it is frequently used in the sequel, we repeat the proof here.

Lemma 3.1 *(i) $L_g[M]$ is a linear map.*

(ii) $L_{f\circ g}[M] = L_f[g(M)]L_g[M]$. In particular, if g is invertible then $L_g[M]$ is invertible and $L_g[M]^{-1} = L_{g^{-1}}[g(M)]$.

(iii) $L_g[M]$ is self-adjoint, with respect to the inner product $\langle A, B\rangle = \mathrm{Tr}\, AB$ in $\mathcal{M}$.

(iv) $L_g[M](M) = g'(M)M$*, where* $g'(x) = \frac{d}{dx}g(x)$*.*

(v) $L_g[M]^{-1}(I) = (g'(M))^{-1}$*,* I *is the identity matrix.*

Proof. It is convenient to use the orthogonal (with respect to the inner product $\langle A, B\rangle$ above) decomposition of the tangent space introduced in [9], $\tilde{T}_M = C(M) \oplus C(M)^\perp$, where $C(M) = \{X \in \mathcal{M} \ : \ XM = MX\}$ and $C(M)^\perp = \{i[M, X] \ : \ X \in \mathcal{M}\}$.

Let $H \in \tilde{T}_M$ be decomposed as $H = H^c + i[M, X]$, then, according to [4] (p. 124), $\frac{d}{ds}g(M + sH)|_{s=0} = g'(M)H^c + i[g(M), X]$. The statements (i) and (ii) follow easily from this equality. Let $K \in \tilde{T}_M$, $K = K^c + i[M, Y]$, then

$$\begin{aligned}\operatorname{Tr} K L_g[M](H) &= \operatorname{Tr} K^c g'(M)H^c - \operatorname{Tr}[M, Y][g(M), X] = \\ &= \operatorname{Tr} g'(M)K^c H^c - \operatorname{Tr}[g(M), Y][M, X] = \operatorname{Tr} L_g[M](K)H\end{aligned}$$

which proves (iii). Further,

(iv) $L_g[M](M) = \frac{d}{ds}g((1 + s)M)|_{s=0} = g'(M)M$

(v) From (ii),

$$L_g[M]^{-1}(I) = L_{g^{-1}}[g(M)](I) = \frac{d}{ds}g^{-1}(g(M) + sI)|_{s=0} = (g'(M))^{-1}$$

In what follows, we omit the indication of the point in square brackets if no confusion is possible. The vector space

$$T_D^g = \{L_g(H) \mid H \in T_D\}$$

will be called the g-representation of the tangent space T_D. The corresponding inner product in T_D^g is

$$\lambda_D^g(G_1, G_2) = \lambda_D(L_g^{-1}(G_1), L_g^{-1}(G_2)) = \operatorname{Tr} G_1 K_g(G_2)$$

where $K_g = L_g^{-1} J_D L_g^{-1}$. Similarly as before, the g-representation of the cotangent space is the space of all linear functionals on T_D^g and it is given by

$$T_D^{g*} = \{K_g(G) \mid G \in T_D^g\} = \{L_{g^{-1}}(A) \mid A \in T_D^*\}$$

The inner product

$$\varphi_D^g(B_1, B_2) = \lambda_D^g(K_g^{-1}(B_1), K_g^{-1}(B_2)) = \operatorname{Tr} B_1 K_g^{-1}(B_2)$$

will be called the generalized g-covariance of B_1 and B_2.

Clearly, if g is the identity function, we obtain the usual tangent and cotangent spaces T_D and T_D^*.

Lemma 3.2

$$\begin{aligned} G \in T_D^g &\iff \operatorname{Tr}(g'(D))^{-1}G = 0 \\ B \in T_D^{g*} &\iff \operatorname{Tr} g'(D)DB = 0 \end{aligned}$$

Proof. Both statements follow easily from $\operatorname{Tr} H = 0$, $H \in T_D$ and Lemma 3.1 (v) and (iv), respectively.

Example 3.1 The quantum analogue of Amari's α-representations is obtained if we put

$$g(x) = g_\alpha(x).$$

In the sequel, we use the letter α to indicate the function g_α, e.g. α-representation, T_D^α, etc. For $\alpha \neq \pm 1$, $(g'_\alpha(D))^{-1} = \frac{1+\alpha}{2} g_{-\alpha}(D)$ and $g'_\alpha(D)D = \frac{1-\alpha}{2} g_\alpha(D)$ and thus

$$\begin{aligned} G \in T_D^\alpha &\iff \operatorname{Tr} g_{-\alpha}(D)G = 0 \\ B \in T_D^{\alpha*} &\iff \operatorname{Tr} g_\alpha(D)B = 0, \end{aligned}$$

An application of Lemma 3.2 also for $\alpha = \pm 1$ shows that $T_D^{\alpha*} = T_D^{-\alpha}$ for each α, so that $K_\alpha = L_\alpha^{-1} J_D L_\alpha^{-1}$ is an isomorphism $K_\alpha : T_D^\alpha \to T_D^{-\alpha}$. This shows that the α- and $-\alpha$-representations are in some sense dual, as in the classical case. In particular, if we put $J_D = J_\alpha = L_{-\alpha} L_\alpha$, then $K_\alpha = L_{-\alpha} L_\alpha^{-1}$ and we see that in this case $K_\alpha^{-1} = K_{-\alpha}$. The corresponding family of metrics was studied in [10] and it was shown that the metric is monotone for $\alpha \in [-3, 3]$. Moreover, for $\alpha = \pm 1$ we obtain the Kubo-Mori metric and $\alpha = \pm 3$ corresponds to the right logarithmic derivative.

4 The affine connections, torsion and curvature

Let $g : R \to R$ be a smooth strictly monotone function and let $M, M' \in \mathcal{M}^+$. Clearly, both T_M^g and T_M^{g*} for each M can be identified with $\mathcal{M}$, so that there is a natural isomorphism $\tilde{T}_M^g \to \tilde{T}_{M'}^g$ given by the identity mapping. This isomorphism induces an affine connection on $\mathcal{M}^+$. Let us denote the corresponding covariant derivative by $\tilde{\nabla}^g$.

Let $x_1, \dots, x_{N+1}$ be a coordinate system in $\mathcal{M}^+$. Let us denote $\partial_i = \frac{\partial}{\partial x_i}$ and let $\tilde{H}_i = \partial_i M(x)$, $\tilde{G}_i = L_g(\tilde{H}_i) = \partial_i g(M(x))$, $i = 1, \dots, N+1$. Then

$$L_g(\tilde{\nabla}^g_{\tilde{H}_j} \tilde{H}_i) = \partial_i \partial_j g(M(x))$$

Hence, the coefficients of the affine connection are

$$\tilde{\Gamma}^g_{ijk}(x) = \lambda_x(\tilde{\nabla}^g_{\tilde{H}_i} \tilde{H}_j, \tilde{H}_k) = \operatorname{Tr} \partial_i \partial_j g(M(x)) K_g(\tilde{G}_k)$$

$i, j, k = 1, \dots, N+1$. From this, it follows that this connection is torsion-free.

If we use the functions g_α, we obtain a one-parameter family of torsion-free connections $\tilde{\nabla}^\alpha$, analogical to Amari's family of α-connections. It is easy to see that, unlike the classical case, the connections $\tilde{\nabla}^\alpha$ and $\tilde{\nabla}^{-\alpha}$ are not dual in general.

To obtain the dual connection, consider the affine connection on $\mathcal{M}^+$ induced by a similar identification $\tilde{T}^{g*}_M \to \tilde{T}^{g*}_{M'}$. The covariant derivative will be denoted by $\tilde{\nabla}^{g*}$, this notation will be justified below. We have

$$L_g^{-1} J_M(\tilde{\nabla}^{g*}_{\tilde{H}_i} \tilde{H}_j) = \partial_i K_g(\tilde{G}_j(x))$$

The coefficients of this connection are

$$\tilde{\Gamma}^{g*}_{ijk} = \lambda(\tilde{\nabla}^{g*}_{\tilde{H}_i} \tilde{H}_j, \tilde{H}_k) = \varphi^g(L_g^{-1} J_M(\tilde{\nabla}^{g*}_{\tilde{H}_i} \tilde{H}_j), K_g(\tilde{G}_k)) = \operatorname{Tr} \partial_i K_g(\tilde{G}_j) \tilde{G}_k \quad (5)$$

Proposition 4.1 *$\tilde{\nabla}^g$ and $\tilde{\nabla}^{g*}$ are dual.*

Proof. For $i, j, k = 1, \dots, N+1$, we have

$$\begin{aligned} \partial_i \lambda(\tilde{H}_j, \tilde{H}_k) &= \partial_i \lambda^g(\tilde{G}_j, \tilde{G}_k) = \partial_i \operatorname{Tr} \tilde{G}_j K_g(\tilde{G}_k) = \\ &= \operatorname{Tr} \partial_i \tilde{G}_j K_g(\tilde{G}_k) + \operatorname{Tr} \tilde{G}_j \partial_i K_g(\tilde{G}_k) = \tilde{\Gamma}^g_{ijk} + \tilde{\Gamma}^{g*}_{ikj}. \end{aligned}$$

Remark 4.1 Notice that for $g = id$, we obtain the mixture and exponential $\nabla^{(m)}$ and $\nabla^{(e)}$ connections defined in [13].

The components of the torsion tensor are

$$\tilde{S}^{g*}_{ijk} = \tilde{\Gamma}^{g*}_{ijk} - \tilde{\Gamma}^{g*}_{jik} = \operatorname{Tr}\{\partial_i K_g(\tilde{G}_j(x)) - \partial_j K_g(\tilde{G}_i(x))\} \tilde{G}_k$$

so that this connection is torsion-free iff

$$\partial_i K_g(\tilde{G}_j) = \partial_j K_g(\tilde{G}_i), \quad \forall i, j = 1, \dots, N+1$$

Obviously, this is not always the case.

Example 4.1 Let us consider the connection $\tilde{\nabla}^{\alpha *}$, $\alpha \in [-3, 3]$ and let the metric tensor be determined by J_α. Then we have $K_\alpha(\tilde{G}_j(x)) = L_{-\alpha}(\tilde{H}_j(x)) = \partial_j g_{-\alpha}(M(x))$, so that the connection is torsion-free. Moreover, $L_\alpha^{-1} J_\alpha = L_{-\alpha}$ and thus

$$L_{-\alpha}(\tilde{\nabla}^{\alpha *}_{\tilde{H}_i} \tilde{H}_j) = \partial_i \partial_j g_{-\alpha}(M(x))$$

From this, it follows that with this choice of the metric tensor, $\tilde{\nabla}^{\alpha *} = \tilde{\nabla}^{-\alpha}$ as in the classical case. In particular, the exponential connection $\tilde{\nabla}^{-1*}$ is torsion-free and $\tilde{\nabla}^{\pm 1*} = \tilde{\nabla}^{\mp 1}$ with the Kubo-Mori metric. Similarly, $\tilde{\nabla}^{\pm 3*} = \tilde{\nabla}^{\mp 3}$ and the connections are torsion-free, with the metric of the right logarithmic derivative.

Consider now $\mathcal{D}$ as an N-dimensional submanifold in $\mathcal{M}$ and let $t_1, \dots, t_N$ be a coordinate system in $\mathcal{D}$. As there is no danger of confusion, we use the symbol ∂_i also for $\frac{\partial}{\partial t_i}$. Let $H_i = \partial_i D(t)$, $G_i = \partial_i g(D(t))$, $i = 1, \dots, N$. The affine structure in $\mathcal{D}$ is obtained by projecting the above affine connections orthogonaly onto $\mathcal{D}$. Clearly, each density matrix D is orthogonal to the tangent space T_D in $\tilde{T}_D$. Indeed, if $H \in T_D$,

$$\lambda_D(H, D) = \operatorname{Tr} H J_D(D) = \operatorname{Tr} H = 0$$

Moreover, $\lambda_D(D, D) = 1$. Using (iv) and (v) of Lemma 3.1, it follows that the covariant derivative is given by

$$\begin{aligned}
L_g(\nabla^g_{H_j} H_i) &= \partial_i \partial_j g(D_t) - g'(D_t) D_t \operatorname{Tr} (g'(D_t))^{-1} \partial_i \partial_j g(D_t) \\
L_g^{-1} J_D(\nabla^{g*}_{H_i} H_j) &= \partial_i K_g(G_j(t)) - (g'(D_t))^{-1} \operatorname{Tr} g'(D_t) D_t \partial_i K_g(G_j(t))
\end{aligned}$$

and the coefficients are

$$\begin{aligned}
\Gamma^g_{ijk}(t) &= \operatorname{Tr} K_g(G_k) \partial_i \partial_j g(D_t) \\
\Gamma^{g*}_{ijk}(t) &= \operatorname{Tr} G_k \partial_i K_g(G_j(t))
\end{aligned}$$

And now, some differential geometry [1, 2]. Let $\tilde{R}$ be the Riemannian curvature tensor of an affine connection $\tilde{\nabla}$ on an m-dimensional manifold $\mathcal{M}$ and let $\tilde{R}^*$ be the curvature tensor of the dual connection. Let X, Y, Z, W be vector fields. Then we have (see Lauritzen in [2])

$$\tilde{R}(X, Y, Z, W) = -\tilde{R}^*(X, Y, W, Z) \tag{6}$$

In particular, $\tilde{R} = 0$ iff $\tilde{R}^* = 0$. Further, let $\mathcal{N}$ be a p-dimensional submanifold in $\mathcal{M}$ with a coordinate system x. Let $X_1, \dots, X_p$ be the natural basis of the tangent space, associated with x and let $Y_1, \ldots, Y_{m-p}$ be orthonormal vector fields on $\mathcal{M}$ normal to $\mathcal{N}$. Recall that the Euler-Shouten imbedding curvature is given by

$$H_{ijl} = \lambda(\tilde{\nabla}_{X_i} X_j, Y_l), \quad i, j = 1, \ldots, p, \; l = 1, \ldots, m-p$$

Let us denote

$$H^*_{ijl} = \lambda(\tilde{\nabla}^*_{X_i} X_j, Y_l), \quad i, j = 1, \ldots, p, \; l = 1, \ldots, m-p$$

The submanifold $\mathcal{N}$ is called autoparallel if its imbedding curvature vanishes, i.e. the parallel shift of a vector in $T_x(\mathcal{N})$ along a curve $\rho(s)$ in $\mathcal{N}$ stays in $T_{\rho(s)}(\mathcal{N})$. Let ∇ be the orthogonal projection of $\tilde{\nabla}$ onto $\mathcal{N}$.

Proposition 4.2 *Let R be the Riemannian curvature tensor of ∇ and let $\tilde{\nabla}$ be torsion-free. Then*

$$\tilde{R}_{ijkl} = R_{ijkl} + \sum_{\nu} (H_{ik\nu} H^*_{jl\nu} - H_{jk\nu} H^*_{il\nu}) \tag{7}$$

for $i, j, k, l = 1, \ldots, p$.

Proof. The proof of this statement for the case of a metric connection, i.e. $\nabla = \nabla^*$, can be found in [12], the general case is obtained by an easy modification of this proof.

Let us now return to the submanifold $\mathcal{D}$ in $\mathcal{M}^+$. The Euler-Shouten imbedding curvature is

$$H^g_{ij1} = \mathrm{Tr}\,(g'(D))^{-1} \partial_i \partial_j g(D), \quad i, j = 1, \ldots, N$$

and

$$H^{g*}_{ij1} = \mathrm{Tr}\, g'(D) D \partial_i K_g(G_j), \quad i, j = 1, \ldots, N$$

Proposition 4.3 *Let $g = g_\alpha$. Then*

$$\begin{aligned} H^\alpha_{ij1} &= -\frac{1+\alpha}{2} \mathrm{Tr}\, H_i J_\alpha(H_j) \\ H^{\alpha*}_{ij1} &= -\frac{1-\alpha}{2} \mathrm{Tr}\, H_i J_D(H_j) \end{aligned}$$

Proof. Let $\alpha \neq \pm 1$. Compute, using Lemma 3.2,

$$\begin{aligned}
H_{ij1}^{\alpha} &= \frac{1+\alpha}{2} \mathrm{Tr}\, g_{-\alpha}(D) \partial_i \partial_j g_\alpha(D) \\
&= \frac{1+\alpha}{2} \{\partial_j \mathrm{Tr}\, g_{-\alpha}(D) \partial_i g_\alpha(D) - \mathrm{Tr}\, \partial_j g_{-\alpha}(D) \partial_i g_\alpha(D)\} \\
&= -\frac{1+\alpha}{2} \mathrm{Tr}\, L_{-\alpha}(H_j) L_\alpha(H_i)
\end{aligned}$$

and

$$\begin{aligned}
H_{ij1}^{\alpha*} &= \frac{1-\alpha}{2} \mathrm{Tr}\, g_\alpha(D) \partial_i K_\alpha(G_j) \\
&= \frac{1-\alpha}{2} \{\partial_i \mathrm{Tr}\, g_\alpha(D) L_\alpha^{-1} J_D(H_j) - \mathrm{Tr}\, \partial_i g_\alpha(D) L_\alpha^{-1} J_D(H_j)\} \\
&= -\frac{1-\alpha}{2} \mathrm{Tr}\, H_j J_D(H_i)
\end{aligned}$$

The proof in case $\alpha = \pm 1$ is nearly the same.

Let $\tilde{R}^g$ be the Riemannian curvature tensor of $\tilde{\nabla}^g$ in $\mathcal{M}^+$. Clearly, $\mathcal{M}^+$ can be parametrized in such a way that

$$g(M(x)) = \sum_{i=1}^{N+1} x_i G_i \tag{8}$$

In this case, we will say that $\mathcal{M}^+$ is parametrized as an extended g-family. We have $\partial_i \partial_j g(M(x)) = 0$, hence $\tilde{\Gamma}_{ijk}^g(x) = 0$, for $i, j, k = 1, \ldots, N+1$ and for each x. It means that this parametrization is affine and therefore $\tilde{R}^g = 0$. Thus also $\tilde{R}^{g*} = 0$. From the identity (7), the curvature tensor R^g in $\mathcal{D}$ is equal to

$$R_{ijkl}^g = H_{jk1}^g H_{il1}^{g*} - H_{ik1}^g H_{jl1}^{g*}$$

If $g = g_\alpha$,

$$R_{ijkl}^\alpha(D) = \frac{1-\alpha^2}{4} \{\mathrm{Tr}\, H_j J_\alpha(H_k) \mathrm{Tr}\, H_i J_D(H_l) - \mathrm{Tr}\, H_i J_\alpha(H_k) \mathrm{Tr}\, H_j J_D(H_l)\}$$

We see that $R^\alpha = 0$ if $\alpha = \pm 1$. It is also clear that $R^\alpha = R^{-\alpha}$. Statistical manifolds with this property are called conjugate symmetric and were studied by Lauritzen (see [2]). Here we have to be aware that $R^{-\alpha} \neq R^{\alpha*}$. The curvature tensor $R^{\alpha*}$ can be computed using the identity (6).

Let $\mathcal{M}'$ be a submanifold in $\mathcal{M}^+$ such that the Riemannian curvature tensor of the projection $\nabla^{\alpha'}$ of $\tilde{\nabla}^{\alpha}$ onto $\mathcal{M}'$ vanishes. Then $\mathcal{M}'$ is flat so that there exists an affine coordinate system θ. The dual connection is curvature-free and if it is also torsion-free then it follows from Theorem 1.2 that there exists a dual $\nabla^{\alpha'*}$-affine coordinate system η. This is the case if, for example, $J = J_\alpha$ or if $\mathcal{M}'$ is one dimensional, see Section 6. However, if the dual connection is not torsion-free, there is no coordinate system dual to θ. Indeed, if η is a dual coordinate system, then it follows from Amari's proof of Theorem 1.2 that η is $\nabla^{\alpha'*}$-affine, i.e. the components of the connection vanish, $\Gamma^{\alpha*}_{ijk}(\eta) = 0 \ \forall\eta,\ \forall i,j,k$. But then we have for the components of the torsion tensor $S^{\alpha*}_{ijk} = \Gamma^{\alpha*}_{ijk} - \Gamma^{\alpha*}_{jik} = 0$.

5 A statistical interpretation

Throughout this paragraph, we will suppose that $\alpha \neq 1$. We will investigate a statistical interpretation of the α-representations of the tangent and cotangent space and the α-connections.

Let $\mathcal{D}' \subseteq \mathcal{D}$ be a smooth p-dimensional submanifold and let $\theta_1, \dots, \theta_p$ be the coordinate system in $\mathcal{D}'$. Let T'_θ be the tangent space of $\mathcal{D}'$ at θ, $H'_i = \frac{\partial}{\partial\theta_i} D(\theta)$, $G'_i = L_g(H'_i)$ and let ∇'^{g} and ∇'^{g*} denote the orthogonal projections of the affine connections onto $\mathcal{D}'$. In [17], locally unbiased estimators were defined and a generalized Cramèr-Rao inequality was proved for the generalized covariance. We give an analogical definition of the α-expectation and α-unbiasedness.

Definition 5.1 *Let $B = (B_1, \dots, B_p)$ be a collection of observables. We will say that B is a locally α-unbiased estimator of θ at θ_0 if*

(i) $\operatorname{Tr} g_\alpha(D_{\theta_0}) B_i = \theta_{0i}$ *for* $i = 1, \dots, p$

(ii) $\frac{\partial}{\partial\theta_j} \operatorname{Tr} g_\alpha(D_\theta) B_i|_{\theta_0} = \operatorname{Tr} G'_j B_i = \delta_{ij},\ i,j, = 1, \dots, p$

The value $\operatorname{Tr} g_\alpha(D)A$ *will be called the α-expectation of the observable A.*

As follows from Example 3.1, the α-representation of the cotangent space $T^{\alpha*}_D$ can be interpreted as the space of all observables with zero α-expectation at D with inner product given by the generalized α-covariance φ^α. The following Lemma is obvious.

Lemma 5.1 *Let* $A = (A_1, \dots, A_p)$ *be a locally unbiased estimator of* θ *at* θ_0. *Then* $B = L_\alpha^{-1}(A) = (L_\alpha^{-1}(A_1), \dots, L_\alpha^{-1}(A_p))$ *is a locally* α*-unbiased estimator of* θ *at* θ_0. *Moreover,* $\varphi(A_i, A_j) = \varphi^\alpha(B_i, B_j)$ *and* $\operatorname{Tr} H_i' A_j = \operatorname{Tr} G_i' B_j$, $i, j = 1, \dots, p$.

The generalized Cramér-Rao inequality from [17] can now be rewritten in the following form.

Theorem 5.1 *Let* $\lambda_{ij} = \lambda(H_i', H_j') = \lambda^\alpha(G_i', G_j')$ *and let* $B = (B_1, \dots, B_p)$ *be a locally* α*-unbiased estimator of* θ *at* 0. *Then*

$$\varphi_D^\alpha(B) \geq (\lambda_{ij})^{-1}$$

in the sense of the order on positive definite matrices. Moreover, equality is attained iff B *is the biorthogonal basis of* $T_0'^{\alpha *}$.

An estimator B which is (α-)unbiased (at each point) and such that its variance attains the Cramèr-Rao bound is called (α-)efficient. Clearly, such estimator does not always exist. The following necessary and sufficient condition a generalization of a result stated (without proof) in [13].

Theorem 5.2 *The* α*-efficient estimator exists iff* $\mathcal{D}'$ *is a* $\nabla^{\alpha *}$*-autoparallel submanifold and the coordinate system* $\theta_1, \dots, \theta_p$ *is* ∇'^α*-affine.*

Proof. Let $\mathcal{D}'$ be $\nabla^{\alpha *}$-autoparallel and let the coordinate system be ∇'^α-affine, i.e. $\nabla'^\alpha_{H_i'} H_j' = 0$, $i, j = 1, \dots, p$. Let us choose a point in the parameter space ($\theta = 0$) and let $B = (B_1, \dots, B_p)$ be the biorthogonal basis of $T_0'^{\alpha *}$. We prove that B is an α-efficient estimator.

Let $X_i(\theta)$ be a $\nabla^{\alpha *}$-parallel vector field such that $L_\alpha^{-1} J_0(X_i(0)) = B_i$, $i = 1, \dots, p$, i.e. $X_i = J_\theta^{-1} L_\alpha \{B_i - g_{-\alpha}(D(\theta)) \operatorname{Tr} g_\alpha(D(\theta)) B_i\}$. As $\mathcal{D}'$ is $\nabla^{\alpha *}$-autoparallel, $X_j(\theta) \in T_\theta'$, $\forall \theta$. Compute

$$\partial_i \operatorname{Tr} g_\alpha(D(\theta)) B_j = \operatorname{Tr} G_i'(\theta) B_j = \operatorname{Tr} G_i'(\theta) L_\alpha^{-1} J_\theta(X_j(\theta)) = \lambda(H_i', X_j)$$

here we have used the identity $\operatorname{Tr} G_i'(\theta) g_{-\alpha}(D(\theta)) = 0$. Further,

$$\partial_k \lambda(H_i', X_j) = \lambda(\nabla'^\alpha_{H_k'} H_i', X_j) + \lambda(H_i', \nabla'^{\alpha *}_{H_k'} X_j) \tag{9}$$

Since the parametrization is ∇'^α-affine, we have $\nabla'^\alpha_{H_k'} H_i' = 0$, $\forall i, k$. Moreover, $X_j(\theta) \in T_\theta'$ and X_j is $\nabla^{\alpha *}$-parallel, hence X_j is $\nabla'^{\alpha *}$-parallel, so that

$\nabla'^{\alpha*}_{H'_i} X_j = 0$. It follows that $\partial_k \mathrm{Tr}\, G'_i(\theta) B_j = 0$ for all θ. Since $\mathrm{Tr}\, G'_i(0) B_j = \delta_{ij}$, we have $\mathrm{Tr}\, G'_i(\theta) B_j = \delta_{ij}$ for each θ. We see that $\mathrm{Tr}\, g_\alpha(D(0)) B_j = 0$ and $\partial_i \mathrm{Tr}\, g_\alpha(D(\theta)) B_j = \delta_{ij}$ for each θ, it follows that B is α-unbiased at each point θ. From Theorem 5.1, it now suffices to prove that $B_j - \theta_j g_{-\alpha}(D(\theta)) \in T'^{\alpha*}_\theta$. But this follows easily from the fact that $X_j(\theta) \in T'_\theta$ and $\mathrm{Tr}\, g_\alpha(D(\theta)) B_j = \theta_j$.

Conversely, let B be the α-efficient estimator, then B is α-unbiased and the matrices $B_i - \theta_i g_{-\alpha}(D(\theta))$, $i = 1, \ldots, p$ form the biorthogonal basis of $T'^{\alpha*}_\theta$ $\forall \theta$. Let $X_i = J_\theta^{-1} L_\alpha^{-1}(B_i - \theta_i g_{-\alpha}(D(\theta)))$, then X_i is a $\nabla^{\alpha*}$-parallel vector field and $X_i \in T'_\theta$ $\forall \theta$, hence X_i is $\nabla'^{\alpha*}$-parallel. Moreover,

$$\lambda(H'_i, X_j) = \mathrm{Tr}\, G'_i B_j = \delta_{ij}$$

From (9) it now follows that $\lambda(\nabla'^{\alpha}_{H'_k} H'_i, X_j) = 0$. But the matrices $X_j(\theta)$, $j = 1, \ldots, p$ form a basis of T'_θ. We may conclude that $\nabla'^{\alpha}_{H'_k} H'_i = 0$, $i, k = 1, \ldots, p$, so that the parametrization is ∇'^{α}-affine.

To see that $\mathcal{D}'$ is $\nabla^{\alpha*}$-autoparallel, it suffices to observe that the parallel vector fields $X_i(\theta)$, $i = 1, \ldots, p$ form a basis of the tangent space T'_θ for each θ.

6 Geodesics and divergence functions

Let us consider a ∇^α-autoparallel submanifold in $\mathcal{D}$ for $\alpha = \pm 1$. Then it is ∇^α-flat, hence there is an affine coordinate system θ. If $\alpha \neq \pm 1$, we will consider the autoparallel submanifolds in $\mathcal{M}^+$. As it was said at the end of Section 4, in general there is no hope for a dual coordinate system to exist, unless the dual connection is torsion-free. It means that we cannot use Amari's theory to define a divergence. However, one dimensional submanifolds are always torsion-free, so that a divergence function exists for each $\nabla^{\pm 1 *}$ and $\tilde{\nabla}^{\alpha*}$-geodesic. As suggested in [13], we use these functions to define a divergence function in $\mathcal{D}$ ($\mathcal{M}^+$).

Clearly, for each α, a $\tilde{\nabla}^{\alpha*}$-geodesic is a solution of

$$L_\alpha^{-1} J_{\rho_t}(\dot{\rho}_t) = A \tag{10}$$

where $A \in \mathcal{M}$, it means that each geodesic is determined by the observable A.

The $\nabla^{\alpha*}$-geodesic is given by

(i) $J_{\rho_t}(\dot\rho_t) = A - \operatorname{Tr}\rho_t A$ for $\alpha = -1$

(ii) $L_1^{-1}J_{\rho_t}(\dot\rho_t) = A - \rho_t \operatorname{Tr} A$ for $\alpha = 1$

(iii) $L_\alpha^{-1}J_{\rho_t}(\dot\rho_t) = A - \frac{1-\alpha^2}{4} g_{-\alpha}(\rho_t)\operatorname{Tr} g_\alpha(\rho_t)A$ for $\alpha \neq \pm 1$

Note that A can be replaced by $A+c_t$ in (i), $A+c_t\rho_t$ in (ii) and $A+c_t g_{-\alpha}(\rho_t)$ in (iii), $c_t \in R$, so that we may always suppose that $A \in T_t^{\alpha *}$.

The relation between $\nabla^{\alpha *}$- and $\tilde\nabla^{\alpha *}$-geodesics is clarified in the following proposition.

Proposition 6.1 *Let $\tilde\rho_t$ be a solution of $L_\alpha^{-1}J_{\rho_t}(\dot\rho_t) = A$. Then $\rho_t = \frac{\tilde\rho_t}{\operatorname{Tr}\tilde\rho_t}$ is a $\nabla^{\alpha *}$ geodesic.*

Proof.

From (3), $J_{\rho_t} = \operatorname{Tr}\tilde\rho_t J_{\tilde\rho_t}$. Compute

$$J_{\rho_t}(\dot\rho_t) = J_{\tilde\rho_t}(\dot{\tilde\rho}_t) - \frac{\operatorname{Tr}\dot{\tilde\rho}_t}{\operatorname{Tr}\tilde\rho_t} = J_{\tilde\rho_t}(\dot{\tilde\rho}_t) - \operatorname{Tr}\rho_t J_{\tilde\rho_t}(\dot{\tilde\rho}_t)$$

here we used the fact that $J_D(D) = I$ (twice) and that J_D is self-adjoint. Further,

$$L_\alpha^{-1}J_{\rho_t}(\dot\rho_t) = A - L_\alpha^{-1}(I)\operatorname{Tr} L_\alpha(\rho_t)A$$

We use Lemma 3.1 (iv) and (v) to complete the proof.

Remark 6.1 Let ρ_t be as in (i). For each t, the coefficient of the affine connection is equal to

$$\Gamma^{-1*}(t) = \lambda(\nabla^{-1*}_{\dot\rho_t}\dot\rho_t, \dot\rho_t) = \operatorname{Tr}\frac{d}{dt}\{J_{\rho_t}(\dot\rho_t)\}\dot\rho_t = \operatorname{Tr}\frac{d}{dt}\{A - \operatorname{Tr}\rho_t A\}\dot\rho_t = 0$$

It follows that the parameter t is ∇^{-1*}-affine. Similarly, for ρ_t as in (ii), t is ∇^{1*}-affine. In the case $\alpha \neq \pm 1$, the $\nabla^{\alpha *}$-geodesic is not flat, hence we consider (as in the classical case) geodesics in $\mathcal{M}^+$. If ρ_t is the solution of (10), t is $\tilde\nabla^{\alpha *}$-affine. Moreover, all affine coordinate systems are connected via affine transformations $t \mapsto at+b$. This coordinate transformation changes the initial point and rescales the observable A as $\frac{1}{a}A$.

Lemma 6.1 *Let $\alpha = -1$ and let ρ_t and $\tilde\rho_t$ be as above. Then $\psi(t) = \log\operatorname{Tr}\tilde\rho_t$ is a potential function for ρ_t.*

Proof. We have to prove that $\frac{d^2}{dt^2}\psi(t) = \lambda(\dot\rho_t, \dot\rho_t)$. Compute

$$\frac{d}{dt}\psi(t) = \frac{\operatorname{Tr}\dot{\tilde\rho}_t}{\operatorname{Tr}\tilde\rho_t} = \operatorname{Tr}\rho_t A$$

Moreover,

$$\lambda(\dot\rho_t, \dot\rho_t) = \operatorname{Tr}\dot\rho_t J_{\rho_t}(\dot\rho_t) = \operatorname{Tr}\dot\rho_t A = \frac{d^2}{dt^2}\psi(t)$$

Lemma 6.2 *Let $\alpha = 1$, then $\psi(t) = \operatorname{Tr}\rho_t \log\rho_t$ is a potential function for ρ_t.*

Proof. We may assume that $A \in T^{1*}_{\rho_0}$, i.e. $\operatorname{Tr} A = 0$. We have $L_1^{-1}(\log\rho_t) = \rho_t \log\rho_t$, hence

$$\psi(t) = \operatorname{Tr} L_1^{-1}(\log\rho_t) = \operatorname{Tr}\rho_t J_{\rho_t} L_1^{-1}(\log\rho_t)$$

and

$$\begin{aligned}\frac{d}{dt}\psi(t) &= \operatorname{Tr}\dot\rho_t J_{\rho_t} L_1^{-1}(\log\rho_t) + \operatorname{Tr}\rho_t \frac{d}{dt} J_{\rho_t} L_1^{-1}(\log\rho_t) = \\ &= \operatorname{Tr} A\log\rho_t + \operatorname{Tr}\rho_t L_1(\dot\rho_t) = \operatorname{Tr} A\log\rho_t\end{aligned}$$

here we have used the fact that $J_D(A) = D^{-1}A$ whenever A commutes with D, thus $J_{\rho_t}(\rho_t\log\rho_t) = \log\rho_t$, and $L_1(\rho_t) = I$. The rest of the proof is the same as above.

Lemma 6.3 *Let $\alpha \neq \pm 1$ and let $\rho_t = \tilde\rho_t$. Then $\psi(t) = \frac{2}{1-\alpha}\operatorname{Tr}\rho_t$ is the potential function .*

Proof.

$$\frac{d}{dt}\operatorname{Tr}\rho_t = \operatorname{Tr}\dot\rho_t J_{\rho_t}(\rho_t) = \operatorname{Tr}\rho_t L_\alpha(A) = \frac{1-\alpha}{2}\operatorname{Tr} g_\alpha(\rho_t)A$$

and

$$\frac{d^2}{dt^2}\operatorname{Tr}\rho_t = \frac{1-\alpha}{2}\operatorname{Tr} L_\alpha(\dot\rho_t)A = \frac{1-\alpha}{2}\lambda(\dot\rho_t, \dot\rho_t).$$

According to Theorem 1.2, ρ_t is also ∇^α ($\tilde\nabla^\alpha$)-flat and there is a dual ∇^α ($\tilde\nabla^\alpha$) - affine coordinate s. As we have seen in the proofs of the above lemmas, the dual coordinate is given by $s(t) = \frac{d}{dt}\psi(t) = \operatorname{Tr} g_\alpha(\rho_t)A$ for each

α. Moreover, there is a divergence function $D_\alpha^\rho : \rho \times \rho \to R$. A divergence measure in $\mathcal{D}$ ($\mathcal{M}^+$) can then be defined as follows. Let $\rho_0, \rho_1 \in \mathcal{D}$. Let $\tilde{\rho}_t$ be the unique $\tilde{\nabla}^{\alpha *}$-geodesic connecting these two states. If $\alpha \neq \pm 1$, the α-divergence is $D_\alpha(\rho_0, \rho_1) = D_\alpha^{\tilde{\rho}}(0, 1)$. If $\alpha = \pm 1$, we use the $\nabla^{\alpha *}$-geodesic $\rho_t = (\mathrm{Tr}\, \tilde{\rho}_t)^{-1} \tilde{\rho}_t$.

Proposition 6.2 *(i) Let $\alpha = -1$ and let $\rho_0, \rho_1 \in \mathcal{D}$. Let ρ_t be as above and let A be the unique observable determining ρ_t. Then*

$$D_\alpha(\rho_0, \rho_1) = \mathrm{Tr}\, \rho_1 A$$

(ii) If $\alpha = 1$, then

$$D_\alpha(\rho_0, \rho_1) = \mathrm{Tr}\, \rho_0 \log \rho_0 + \mathrm{Tr}\, (A - \rho_1) \log \rho_1$$

(iii) Let $\alpha \neq \pm 1$, $\rho_0, \rho_1 \in \mathcal{M}^+$. Then

$$D_\alpha(\rho_0, \rho_1) = \frac{2}{1 - \alpha} (\mathrm{Tr}\, \rho_0 - \mathrm{Tr}\, \rho_1) + \mathrm{Tr}\, g_\alpha(\rho_1) A$$

Proof. From the definition of the divergence function and the identity (2), we obtain

$$D_\alpha^\rho(t_1, t_2) = \psi(t_1) - \psi(t_2) + (t_2 - t_1) s_2$$

The rest of the proof is easy.

Let now ρ_t be a geodesic connecting two states ρ_0 and ρ_1 and let ρ_{t_1}, ρ_{t_2} be two states lying on ρ_t. Using Remark 6.1, it is easy to see that $D_\alpha(\rho_{t_1}, \rho_{t_2}) = D_\alpha^\rho(t_1, t_2)$. It also follows that for each ρ we may put $D_\alpha(\rho, \rho) = 0$. There are some properties of the divergence D_α which follow from the properties of D_α^ρ:

(i) Positivity: $D_\alpha(\rho, \sigma) \geq 0$ and $D_\alpha(\rho, \sigma) = 0$ iff $\rho = \sigma$

(ii) Let $\sigma = \rho + dtH$. Let $A = L_\alpha^{-1} J_D(H)$ and let ρ_t be the geodesic determined by A such that $\rho_{t_0} = \rho$. Then

$$D_\alpha(\rho, \rho + dtH) = D_\alpha^\rho(t_0, t_0 + dt) = dt^2 \lambda(H, H)$$

and

$$D_\alpha(\rho + dtH, \rho) = D_\alpha^\rho(t_0 + dt, t_0) = D_\alpha^\rho(t_0, t_0 + dt)$$

It means that the α-divergence induces the metric.

Example 6.1 Let λ be determined by $J_\alpha = L_{-\alpha}L_\alpha$ for some $\alpha \in (-3,3)$. As we have seen, in this case $\nabla^{\alpha *} = \nabla^{-\alpha}$ and this connection is torsion-free. Hence we have the same situation as in the classical case. Let $\alpha = \pm 1$ and let us consider the exponential family

$$\rho(\theta) = \exp\{\sum_{i=1}^m \theta_i A_i - \psi(\theta)\}$$

Then it is ± 1-flat and $\psi(\theta) = \log \mathrm{Tr}\, \exp(\theta_i A_i)$ is the potential function. Hence the coordinate systems $(\theta, \{\eta_i = \partial_i \psi(\theta)\}$ are mutually dual, θ is ∇^α-affine and η is $\nabla^{-\alpha}$-affine. The divergence is given by

$$D_1(\rho_0, \rho_1) = \mathrm{Tr}\, \rho_1 (\log \rho_1 - \log \rho_0)$$

which is the relative entropy. Similarly, for the mixture family

$$\rho(\eta) = \rho_0 + \sum_{i=1}^m \eta_i A_i, \qquad \mathrm{Tr}\, A_i = 0$$

the function $\psi(\eta) = \mathrm{Tr}\, \rho(\eta) \log \rho(\eta)$ is the potential function, so that there is a pair of dual affine coordinate systems (η, θ), see also [9]. The divergence is

$$D_{-1}(\rho_0, \rho_1) = D_1(\rho_1, \rho_0)$$

For $\alpha \neq \pm 1$, we consider the extended α-family

$$\rho(\theta) = g_\alpha^{-1}(\sum_{i+1}^m \theta_i A_i)$$

Let $\psi(\theta) = \frac{2}{1-\alpha} \mathrm{Tr}\, \rho(\theta)$. Then

$$\frac{\partial}{\partial \theta_i} \psi(\theta) = \frac{\partial}{\partial \theta_i} \frac{2}{1-\alpha} \mathrm{Tr}\, g_\alpha^{-1}(\sum_k \theta_k A_k) = \frac{2}{1-\alpha} \mathrm{Tr}\, L_\alpha^{-1} A_i = \mathrm{Tr}\, g_{-\alpha}(\rho(\theta)) A_i$$

and

$$\begin{aligned} \frac{\partial}{\partial \theta_j} \mathrm{Tr}\, g_{-\alpha}(\rho(\theta)) A_i &= \frac{\partial}{\partial \theta_j} \mathrm{Tr}\, g_{-\alpha}(g_\alpha^{-1}(\sum_k \theta_i A_k)) A_i = \\ &= \mathrm{Tr}\, L_{-\alpha} L_\alpha^{-1}(A_j) A_k = \lambda^\alpha(A_i, A_j) \end{aligned}$$

hence $\psi(\theta)$ is the potential function. Thus there is a pair of dual affine coordinate systems and a divergence

$$D_\alpha(\rho_0,\rho_1)=\mathrm{Tr}\, g_\alpha(\rho_1)(g_{-\alpha}(\rho_1)-g_{-\alpha}(\rho_0))$$

This α-divergence was defined also in [8] It is easy to see that the above divergence functions are the same as those from Proposition 6.2.

Example 6.2 ([5, 13]) Let $\alpha=-1$ and let λ be the metric of the symmetric logarithmic derivative. Then it is easy to see that

$$\rho_t=\exp\{\frac{1}{2}(tA-\psi(t))\}\rho_0\exp\{\frac{1}{2}(tA-\psi(t))\}$$

where $\psi(t)=\log\mathrm{Tr}\,\rho_0\exp tA$, is a solution of (i). Hence it is a ∇^{-1*}-geodesic and the coordinate t is $\nabla^{\alpha*}$-affine. It follows that ρ_t is also ∇^{-1}-flat and there is a ∇^{-1}-affine coordinate system $s(t)=\mathrm{Tr}\,\rho_t A$. Further, the divergence is

$$D_{-1}(\rho_0,\rho_1)=2\mathrm{Tr}\,\rho_1\log\rho_0^{-\frac{1}{2}}(\rho_0^{\frac{1}{2}}\rho_1\rho_0^{\frac{1}{2}})^{\frac{1}{2}}\rho_0^{-\frac{1}{2}}$$

Clearly, this divergence coincides with the relative entropy if ρ_0 and ρ_1 commute. Moreover, ρ_t has the Gibbs state $\exp(tA-\psi(t))$ as a special case.

Example 6.3 Let $\alpha=-1$ and let λ be the metric of the right logarithmic derivative. Then it is easy to see that

$$\tilde{\rho}_t=F^{\frac{1}{2}}\exp\{tQ_F(A)\}F^{\frac{1}{2}}$$

where Q_F is a linear operator given by $Q_F^{-1}(A)=\frac{1}{2}(F^{-\frac{1}{2}}AF^{\frac{1}{2}}+F^{\frac{1}{2}}AF^{-\frac{1}{2}})$ is a solution of (10). If ρ_0 and ρ_1 are two states,

$$\rho_t=\rho_1^{\frac{1}{2}}\exp\{(t-1)Q_{\rho_1}(A)-\psi(t)\}\rho_1^{\frac{1}{2}}$$

with $\psi(t)=\log\mathrm{Tr}\,\rho_1\exp\{(t-1)Q_{\rho_1}(A)\}$ is a ∇^{-1*}-geodesic. Choose A so that

$$\rho_1^{\frac{1}{2}}\exp\{-Q_{\rho_1}(A)\}\rho_1^{\frac{1}{2}}=\rho_0$$

ρ_t is then the ∇^{-1*}-geodesic connecting these two states. We see that the divergence is given by

$$D^{-1}(\rho_0,\rho_1)=\mathrm{Tr}\,\rho_1 A=\mathrm{Tr}\,\rho_1\log\rho_1^{\frac{1}{2}}\rho_0^{-1}\rho_1^{\frac{1}{2}}$$

This version of the relative entropy appeared also in [3].

References

[1] Amari, S. *Differential-geometrical methods in statistic*, Lecture Notes in Statistics 28(1985)

[2] S. Amari, O.E. Barndorff-Nielsen, R.E. Kass, S.L. Lauritzen, C.R. Rao *Differential geometry in statistical inference.* IMS Lecture notes-Monograph series 10(1987)

[3] Belavkin, V.P. and Staszewski, P. C^*-algebraic generalizations of relative entropy and entropy, Ann. Inst. Henri Poincaré A37 (1982) 51–58

[4] Bhatia, R. *Matrix analysis*, Springer, New York, 1997

[5] Braunstein, S.L. and Caves, C.M. Geometry of quantum states, in *Quantum Communication, Computing and Measurement*, eds. Hirota et al., Plenum Press, New York (1994)

[6] Chentsov, N.N. *Statistical decision rules and optimal inferences*, Translation of Math. Monog. 53 (Amer. Math. Society, Providence 1982)

[7] Chentsov, N.N. and Morozova, Markov invariant geometry on state manifolds (in Russian), Itogi Nauki i Tekhniki 36 (1990), 69–102

[8] Hasegawa, H. α-divergence of the non-commutative information geometry, Rep. Math. Phys. 33 (1993) 87–93

[9] Hasegawa, H. Exponential and mixture families in quantum statistics: Dual structure and unbiased parameter estimation, Rep. Math. Phys. 39 (1997) 49–68

[10] Hasegawa, H. and Petz, D. Non-commutative extension of information geometry II, Quantum Communication, Computing and Measurement, Eds. Hirota et al., Plenum Press, New York, (1997)

[11] Holevo, A.S., *Probabilistic and statistical aspects of quantum theory*, North-Holland, Amsterdam, 1982

[12] S. Kobayashi, K. Nomizu *Foundation of Differential Geometry, Volume II*, Interscience, John Wiley & Sons, New York, London, 1963

[13] Nagaoka, H. Differential geometrical aspects of quantum state estimation and relative entropy, in *Quantum Communication, Computing and Measurement*, eds. Hirota et al., Plenum Press, New York (1994)

[14] Petz, D. and Sudár, Cs., Geometries of quantum states, J. Math. Phys. **37**(1996), 2662–2673

[15] Petz, D. Information geometry of quantum states, in *Quantum Probability Communications*, vol. 10, ed. R.L. Hudson et al, World Scientific, 135–158

[16] Petz, D. Monotone metrics on matrix spaces, Linear Algebra Appl. 244(1996),81–96

[17] Petz, D. and Jenčová, A. On quantum Fisher information, to appear