Duality, Monotonocity and the WYD Metrics

M. R. Grasselli
Dept. of Mathematics and Statistics
McMaster University

04/07/2002

1. Overview

- Classical information geometry: differential geometric properties of families of classical probability densities.
  - parametric and nonparametric (very complete).
  - Fisher metric (unique);
  - equivalent definitions of α-connections (including exponential and mixture);
  - divergence functions (minimization and orthogonality);

- Quantum information geometry: differential geometric properties of families of quantum probabilities.
  - parametric (density matrices) and nonparametric (very incomplete).
  - multitude of monotone metrics (BKM, WYD, Bures,...);
  - exponential and mixture connections;
  - inequivalent definitions of α-connections;
  - divergence functions and quantum entropies.

Finite Dimensional Quantum Setup

- $\mathcal{H}^N$: finite dimensional complex Hilbert space;
- $\mathcal{B}(\mathcal{H}^N)$: algebra of operators on $\mathcal{H}^N$;
- $\mathcal{A}$: N^2-dimensional real vector subspace of self-adjoint operators;
- $\mathcal{M}$: n-dimensional submanifold of all invertible density operators on $\mathcal{H}^N$, with $n = N^2 - 1$.

2. The Quantum α-connections

2.1 The α-representation

For $\alpha \in (-1, 1)$, define the α-embedding of $\mathcal{M}$ into $\mathcal{A}$ as

$$\begin{aligned} \ell_\alpha \;\; : \;\; & \mathcal{M} \to \mathcal{A} \\ & \rho \mapsto \frac{2}{1-\alpha}\rho^{\frac{1-\alpha}{2}}. \end{aligned}$$

In the next lemma, for $A \in \mathcal{B}(\mathcal{H}^N)$, let

$$\mathcal{C}(A) = \{B \in \mathcal{B}(\mathcal{H}^N) : [A, B] = 0\}$$

denote its commutant.

Lemma 1 (Hasegawa, 1996) *Let $\mathcal{S} = \rho(\theta)$ be a smooth manifold of invertible density matrices. Then there exists a anti-selfadjoint operator Δ_i such that*

$$\frac{\partial\rho}{\partial\theta^i} = \frac{\partial^c\rho}{\partial\theta^i} + [\rho, \Delta_i], \tag{2}$$

where $\frac{\partial^c\rho}{\partial\theta^i} \in \mathcal{C}(\rho)$ and $[\rho, \Delta_i] \in \mathcal{C}(\rho)^\perp$. Moreover, for any function F which is differentiable on a neighbourhood of the spectrum of ρ we have

$$\frac{\partial F(\rho)}{\partial\theta^i} = \frac{\partial^c F(\rho)}{\partial\theta^i} + [F(\rho), \Delta_i], \tag{3}$$

where $\frac{\partial^c F(\rho)}{\partial\theta^i} \in \mathcal{C}(\rho)$ and $[F(\rho), \Delta_i] \in \mathcal{C}(\rho)^\perp$.

At each point $\rho \in \mathcal{M}$, consider the subspace of $\mathcal{A}$ defined by

$$\mathcal{A}_\rho^{(\alpha)} = \left\{ A \in \mathcal{A} : \mathrm{Tr}\left(\rho^{\frac{1+\alpha}{2}} A\right) = 0 \right\},$$

and define the isomorphism

$$\begin{aligned} (\ell_\alpha)_{*(\rho)} \ : \ & T_\rho\mathcal{M} \to \mathcal{A}_\rho^{(\alpha)} \\ & v \mapsto (\ell_\alpha \circ \gamma)'(0), \end{aligned} \tag{4}$$

where $\gamma : (-\varepsilon, \varepsilon) \to \mathcal{M}$ is a curve in the equivalence class of the tangent vector v. We call this isomorphism the α-representation of the tangent space $T_\rho\mathcal{M}$. If $(\theta^1, \ldots, \theta^n)$ is a coordinate system for $\mathcal{M}$, then the α-representation of a basis tangent vector is

$$\frac{\partial}{\partial \theta^i} \mapsto \frac{\partial \ell_\alpha(\rho)}{\partial \theta^i}.$$

Using (3) with $F(\rho) = \ell_\alpha(\rho)$, we obtain

$$\frac{\partial \ell_\alpha(\rho)}{\partial \theta^i} = \rho^{\frac{1-\alpha}{2}} \frac{\partial^c \log \rho}{\partial \theta^i} + \frac{2}{1-\alpha}[\rho^{\frac{1-\alpha}{2}}, \Delta_i]. \tag{5}$$

Therefore, it follows from the normalisation condition $\mathrm{Tr}\rho = 1$ and the cyclicity of the trace that

$$\mathrm{Tr}\left(\rho^{\frac{1+\alpha}{2}} \frac{\partial \ell_\alpha(\rho)}{\partial \theta^i}\right) = \mathrm{Tr}\left(\frac{\partial^c \rho}{\partial \theta^i} + \frac{2}{1-\alpha}[\rho, \Delta_i]\right) = 0,$$

so that $\frac{\partial \ell_\alpha(\rho)}{\partial \theta^i} \in \mathcal{A}_\rho^{(\alpha)}$.

2.2 The Covariant Derivative $\nabla^{(\alpha)}$

Let $r = \frac{2}{1-\alpha}$. If we equip $\mathcal{A}$ with the the r-norm

$$\|A\|_r := (\mathrm{Tr}|A|^r)^{1/r}\,,$$

then the α-embedding can be vied as a mapping from $\mathcal{M}$ to the positive part of the sphere of radius r, since for any $\rho \in \mathcal{M}$ we have

$$\|\ell_\alpha(\rho)\|_r = \left(\mathrm{Tr}\left|r\rho^{1/r}\right|^r\right)^{1/r} = r,$$

so that $\ell_\alpha(\rho) \in S^r$.

The tangent space at a point $0 \leq \sigma \in S^r$ is

$$T_\sigma S^r = \left\{A \in \mathcal{A} : \mathrm{Tr}(A\sigma^{r-1}) = 0\right\}.$$

If we put $\sigma = \ell_\alpha(\rho) = r\rho^{1/r}$, we find that

$$T_{r\rho^{1/r}} S^r = \left\{ A \in \mathcal{A} : \mathsf{Tr}(A\rho^{1-1/r}) = 0 \right\} = \mathcal{A}_\rho^{(\alpha)},$$

so that the α-representation (4) is indeed an isomorphism between tangent spaces, as the push-forward notation suggests.

For each $0 \le \sigma \in S^r$, the canonical projection from the tangent space $T_\sigma \mathcal{A}$ onto the tangent space $T_\sigma S^r$ is uniquely given by

$$\begin{aligned} \Pi_\sigma \quad : \quad & T_\sigma \mathcal{A} \to T_\sigma S^r \\ & A \mapsto A - \left(r^{-r} \mathsf{Tr} \left[A \sigma^{r-1} \right] \right) \sigma. \end{aligned}$$

For $\sigma = \ell_\alpha(\rho) = r\rho^{1/r}$, this gives

$$\begin{aligned} \Pi_{r\rho^{1/r}} \quad : \quad & T_{r\rho^{1/r}} \mathcal{A} \to T_{r\rho^{1/r}} S^r \\ & A \mapsto A - \left(\mathsf{Tr} \left[\rho^{\frac{1+\alpha}{2}} A \right] \right) \rho^{\frac{1-\alpha}{2}}. \end{aligned}$$

Definition 6 *For $\alpha \in (-1, 1)$, let $\gamma : (-\varepsilon, \varepsilon) \to \mathcal{M}$ be a smooth curve such that $\rho = \gamma(0)$ and $v = \dot{\gamma}(0)$ and let $s \in S(T\mathcal{M})$ be a differentiable vector field. The α-connection on $T\mathcal{M}$ is given by*

$$\left(\nabla_v^{(\alpha)} s\right)(\rho) = (\ell_\alpha)_{*(\rho)}^{-1} \left[\Pi_{r\rho^{1/r}} \widetilde{\nabla}_{(\ell_\alpha)_{*(\rho)} v} (\ell_\alpha)_{*(\gamma(t))} s\right]. \qquad (7)$$

Using the definition (7), we find that the α-representation of the α-covariant derivative of the vector field $\partial/\partial\theta^j$ in the direction of the tangent vector $\partial_i := \partial/\partial\theta^i$ is

$$\left(\nabla_{\partial_i}^{(\alpha)} \frac{\partial}{\partial\theta^j}\right)^{(\alpha)} = \frac{\partial^2 \ell_\alpha(\rho)}{\partial\theta^i \partial\theta^j} - \mathrm{Tr}\left(\rho^{\frac{1+\alpha}{2}} \frac{\partial^2 \ell_\alpha(\rho)}{\partial\theta^i \partial\theta^j}\right) \rho^{\frac{1-\alpha}{2}}. \qquad (8)$$

2.3 The α-parallel Transport and the Extend Manifold $\hat{\mathcal{M}}$

Consider the extended manifold $\hat{\mathcal{M}}$ of positive definite matrices. Observe first that the α-embedding in this case maps $\hat{\mathcal{M}}$ to itself. Moreover, $T\hat{\mathcal{M}} = T\mathcal{A} \simeq \mathcal{A}$. We can therefore define the α-parallel transport on $\hat{\mathcal{M}}$ simply by

$$\begin{aligned} \hat{\tau}^{(\alpha)}_{\sigma_0,\sigma_1} \quad : \quad & T_{\sigma_0}\hat{\mathcal{M}} \to T_{\sigma_1}\hat{\mathcal{M}} \\ & v \mapsto (\ell_\alpha)^{-1}_{*(\sigma_1)}\left((\ell_\alpha)_{*(\sigma_0)}v\right), \end{aligned}$$

and we find (using (7) without the projection step) that the α-representation of its covariant derivative is

$$\left(\hat{\nabla}^{(\alpha)}_{\partial_i}\frac{\partial}{\partial\theta^j}\right)^{(\alpha)} = \frac{\partial^2 \ell_\alpha(\rho)}{\partial\theta^i\partial\theta^j}, \tag{9}$$

Now let $\{X_1, \ldots, X_{n+1}\}$ be a basis for $\mathcal{A}$. For each $\sigma \in \hat{\mathcal{M}}$, we have that $\sigma^{\frac{1-\alpha}{2}} \in \mathcal{A}$, so that there exist real numbers $\xi = \{\xi^1, \ldots, \xi^{n+1}\}$ such that

$$\frac{2}{1-\alpha}\sigma^{\frac{1-\alpha}{2}} = \xi^1 X_1 + \cdots + \xi^{n+1} X_{n+1}.$$

Then $\xi = \{\xi^1, \ldots, \xi^{n+1}\}$ is a $\hat{\nabla}^{(\alpha)}$-affine coordinate system for $\hat{\mathcal{M}}$, since (9) gives

$$\left(\hat{\nabla}^{(\alpha)}_{\partial_i} \frac{\partial}{\partial \xi^j}\right)^{(\alpha)} = \frac{\partial^2 \ell_\alpha(\rho)}{\partial \xi^i \partial \xi^j} = \frac{\partial X_j}{\partial \xi^i} = 0.$$

Therefore, $\hat{\mathcal{M}}$ is $\hat{\nabla}^{(\alpha)}$-flat, even though its submanifold $\mathcal{M}$ is not $\nabla^{(\alpha)}$-flat.

3. Duality and the WYD Metric

- Dual connections: two connections ∇ and ∇^* on a Riemannian manifold $(\mathcal{M}, g)$ are dual with respect to g if and only if

$$Xg(Y,Z) = g\left(\nabla_X Y, Z\right) + g\left(Y, \nabla^*_X Z\right), \tag{10}$$

for any vector fields X, Y, Z on $\mathcal{M}$. Equivalently, if $\tau_{\gamma(t)}$ and $\tau^*_{\gamma(t)}$ are the respective parallel transports along a curve $\{\gamma(t)\}_{0 \leq t \leq 1}$ on $\mathcal{M}$, with $\gamma(0) = \rho$, then ∇ and ∇^* are dual with respect to g if and only if for all $t \in [0,1]$,

$$g_\rho(Y,Z) = g_{\gamma(t)}\left(\tau_{\gamma(t)} Y, \tau^*_{\gamma(t)} Z\right). \tag{11}$$

- Dual coordinate systems: two coordinate systems $\theta = (\theta^i)$ and $\eta = (\eta_i)$ on a Riemannian manifold $(\mathcal{M}, g)$ are dual with respect to g if and only if their natural bases for $T_\rho\mathcal{M}$ are *biorthogonal* at every point $\rho \in \mathcal{M}$, that is,

$$g\left(\frac{\partial}{\partial\theta^i}, \frac{\partial}{\partial\eta_j}\right) = \delta^i_j.$$

Equivalently, $\theta = (\theta^i)$ and $\eta = (\eta_i)$ are dual with respect to g if and only if

$$g_{ij} = \frac{\partial\eta_i}{\partial\theta^j} \quad \text{and} \quad g^{ij} = \frac{\partial\theta_i}{\partial\eta^j}$$

at every point $\rho \in \mathcal{M}$, where, as usual, $g^{ij} = (g_{ij})^{-1}$.

Theorem 12 (Amari, 1985) *When a Riemannian manifold* $(\mathcal{M}, g)$ *has a pair of dual coordinate systems* (θ, η)*, there exist potential functions* $\Psi(\theta)$ *and* $\Phi(\eta)$ *such that*

$$g_{ij}(\theta) = \frac{\partial^2 \Psi(\theta)}{\partial \theta^i \partial \theta^j} \quad \text{and} \quad g^{ij} = \frac{\partial^2 \Phi(\eta)}{\partial \eta_i \partial \eta_j}.$$

Conversely, when either potential function Ψ *or* Φ *exists from which the metric is derived by differentiating it twice, there exist a pair of dual coordinate systems. The dual coordinate systems and the potential functions are related by the folloing Legendre transforms*

$$\theta^i = \frac{\partial \Phi(\eta)}{\partial \eta_i}, \quad \eta_i = \frac{\partial \Psi(\theta)}{\partial \theta^i}$$

and

$$\Psi(\theta) + \Phi(\eta) - \theta^i \eta_i = 0$$

Theorem 13 (Amari, 1985) *Suppose that* ∇ *and* ∇^* *are two flat connections on a manifold* $\mathcal{M}$*. If they are dual with respect to a Riemannian metric* g *on* $\mathcal{M}$*, then there exists a pair* (θ, η) *of dual coordinate systems such that* θ *is* ∇*-affine and* η *is a* ∇^**-affine.*

Let us now consider the definition of a Riemannian metric for our manifold $\mathcal{M}$ of density matrices. Using the α-representation to obtain a concrete realization of tangent vectors on $\mathcal{M}$ in terms of operators in $\mathcal{A}$, a Riemannian metric on $\mathcal{M}$ is deemed to be provided by the smooth assignment of an inner product $\langle \cdot, \cdot \rangle_\rho$ in $\mathcal{A} \subset B(\mathcal{H}^N)$ for each point $\rho \in \mathcal{M}$.

For a fixed $\alpha \in (-1, 1)$, the *WYD* (Wigner-Yanase-Dyson) metric on $\mathcal{M}$ is given by

$$g_\rho^{(\alpha)}(A, B) := \mathrm{Tr}\left(A^{(\alpha)} B^{(-\alpha)}\right), \qquad A, B \in T_\rho \mathcal{M}. \tag{14}$$

In a coordinate system $(\theta^1, \ldots, \theta^n)$ for $\mathcal{M}$, we have that

$$\begin{aligned} g_{ij}^{(\alpha)}(\theta) &:= g_\rho^{(\alpha)}\left(\frac{\partial}{\partial\theta^i}, \frac{\partial}{\partial\theta^j}\right) = \mathsf{Tr}\left(\frac{\partial \ell_\alpha(\rho)}{\partial\theta^i}\frac{\partial \ell_{-\alpha}(\rho)}{\partial\theta^j}\right) \qquad (15) \\ &= \mathsf{Tr}\left(\rho\frac{\partial^c \log\rho}{\partial\theta^i}\frac{\partial^c \log\rho}{\partial\theta^j}\right) + \frac{4}{1-\alpha^2}\mathsf{Tr}\left[\rho^{\frac{1-\alpha}{2}}, \Delta_i\right]\left[\rho^{\frac{1+\alpha}{2}}, \Delta_j\right]. \end{aligned}$$

It is clear that $g_{ij}^{(\alpha)} = g_{ji}^{(\alpha)} = g_{ij}^{(-\alpha)}$.

Observe also that for the extreme cases $\alpha \to \pm 1$, formula (14) leads to the familiar *BKM* (Bogoliubov-Kubo-Mori) metric

$$g_\rho^{(\pm 1)}(A, B) = g_\rho^B(A, B) = \text{Tr}\left(A^{(-1)} B^{(1)}\right) \qquad (16)$$

where $A^{(\pm 1)}, B^{(\pm 1)}$ are the ± 1-representations of the tangent vectors $A, B \in T_\rho \mathcal{M}$. In coordinates, the *BKM* metric assumes the form

$$\begin{aligned} g_{ij}^B(\theta) &:= g_\rho^B\left(\frac{\partial}{\partial\theta^i}, \frac{\partial}{\partial\theta^j}\right) = \text{Tr}\left(\frac{\partial \log\rho}{\partial\theta^i}\frac{\partial\rho}{\partial\theta^j}\right) \\ &= \text{Tr}\left(\rho\frac{\partial^c \log\rho}{\partial\theta^i}\frac{\partial^c \log\rho}{\partial\theta^j}\right) + \text{Tr}[\log\rho, \Delta_i][\rho, \Delta_j]. \qquad (17) \end{aligned}$$

It follows directly from the definition (14), that the $\pm\alpha$-connections are dual with respect to the metric $g^{(\alpha)}$ for each fixed value of $\alpha \in (-1, 1)$ (just as the ± 1-connections are dual with respect to the *BKM* metric). Our purpose is to discover what other metrics have the same property.

From now on confine our attention to those metrics on $\mathcal{M}$ which are obtained as restrictions of metrics on the extended manifold $\hat{\mathcal{M}}$, which is $\hat{\nabla}^{(\pm\alpha)}$-flat, and treat the latter as our primary objects.

Observe first that the *WYD* metric extends quite naturally to $\hat{\mathcal{M}}$, simply using the $\pm\alpha$-representations of tangent vectors $\widehat{A}, \widehat{B}$ (that is, the representation induced by the $\pm\alpha$-embedding of $\hat{\mathcal{M}}$ into $\mathcal{A}$):

$$\widehat{g}_\sigma^{(\alpha)}\left(\widehat{A}, \widehat{B}\right) := \mathrm{Tr}\left(\widehat{A}^{(\alpha)}\widehat{B}^{(-\alpha)}\right), \qquad \widehat{A}, \widehat{B} \in T_\sigma\hat{\mathcal{M}}. \tag{18}$$

Lemma 19 *If* $(\theta^1, \ldots, \theta^{n+1})$ *is a* $\widehat{\nabla}^{(\alpha)}$*-affine coordinate system for the extended manifold* $\hat{\mathcal{M}}$*, then the function*

$$\tilde{\Psi}_\alpha(\theta) = \frac{2}{1+\alpha} Tr\sigma(\theta), \qquad \sigma(\theta) \in \hat{\mathcal{M}} \tag{20}$$

satisfies

$$\hat{g}_{ij}^{(\alpha)}(\theta) = \frac{\partial^2 \tilde{\Psi}_\alpha(\theta)}{\partial\theta^i \partial\theta^j}. \tag{21}$$

Moreover,

$$\tilde{\eta}_i = \frac{\partial \tilde{\Psi}_\alpha(\theta)}{\partial\theta^i} \tag{22}$$

is a $\widehat{\nabla}^{(-\alpha)}$*-affine coordinate system for* $\hat{\mathcal{M}}$*.*

Theorem 23 *For a fixed value of* $\alpha \in (-1, 1)$*, suppose that the connections* $\nabla^{(\alpha)}$ *and* $\nabla^{(-\alpha)}$ *are dual with respect to a Riemannian metric* $\widehat{g}$ *on* $\hat{\mathcal{M}}$*. Then there exist a constant (that is, independent of* σ*)* $(n+1) \times (n+1)$ *matrix* M*, such that* $(\widehat{g}_\sigma)_{ij} = \sum_{k=1}^{n+1} M_i^k (\widehat{g}_\sigma^{(\alpha)})_{kj}$*, in some* α*-affine coordinate system.*

Proof: (*Ato 1, primo movimento*) Since the two connections are flat on the extend manifold $\hat{\mathcal{M}}$, theorem 13 tell us that there exist dual coordinate systems (θ, η) such that θ is $\nabla^{(\alpha)}$-affine and η is $\nabla^{(-\alpha)}$-affine.

(*Act 1, secondo movimento*) Using lemma 19, we know that the function $\tilde{\Psi}_\alpha(\theta) = \frac{2}{1+2}\text{Tr}\sigma(\theta)$ satisfies

$$\hat{g}_{ij}^{(\alpha)}(\theta) = \frac{\partial^2 \tilde{\Psi}_\alpha(\theta)}{\partial\theta^i \partial\theta^j} \tag{24}$$

and also that

$$\tilde{\eta}_i = \frac{\partial \tilde{\Psi}_\alpha(\theta)}{\partial\theta^i} \tag{25}$$

is a another $\hat{\nabla}^{(-\alpha)}$-affine coordinate system for $\hat{\mathcal{M}}$.

(*Intermezzo*) Therefore, the coordinate systems (η) and $(\tilde{\eta})$ are related by an affine transformation, so there must exist a matrix M and numbers $(a_1, \ldots, a_{n+1})$ such that

$$\eta_i = \sum_{k=1}^{n} M_i^k \tilde{\eta}_k + a_i. \tag{26}$$

(*Ato 2, movimento unico*) But from theorem 12, there exists a potential function $\Psi(\theta)$ such that

$$\widehat{g}_{ij}(\theta) = \frac{\partial^2 \Psi(\theta)}{\partial \theta^i \partial \theta^j}$$

and

$$\eta_i = \frac{\partial \Psi(\theta)}{\partial \theta^i}.$$

Equation (26) then gives

$$\frac{\partial \Psi(\theta)}{\partial \theta^i} = \sum_{k=1}^{n+1} M_i^k \frac{\partial \tilde{\Psi}_\alpha(\theta)}{\partial \theta^k} + a_i.$$

(*Finale*) Differentiating this equation with respect to θ^j leads to

$$\widehat{g}_{ij}(\theta) = \frac{\partial^2 \Psi(\theta)}{\partial \theta^i \partial \theta^j} = \sum_{k=1}^{n} M_i^k \frac{\partial^2 \tilde{\Psi}_\alpha(\theta)}{\partial \theta^j \partial \theta^k} = \sum_{k=1}^{n} M_i^k \widehat{g}_{kj}^{(\alpha)}(\theta). \qquad (27)$$