Metadata of the chapter that will be visualized in SpringerLink

Book Title	Geometric Science of Information	
Series Title		
Chapter Title	Information Geometry Under Monotone Embedding. Part II: Geometry	
Copyright Year	2017	
Copyright HolderName	Springer International Publishing AG	

Corresponding Author	Family Name	**Naudts**
	Particle	
	Given Name	**Jan**
	Prefix	
	Suffix	
	Division	
	Organization	Universiteit Antwerpen
	Address	Antwerpen, Belgium
	Email	jan.naudts@uantwerpen.be
Author	Family Name	**Zhang**
	Particle	
	Given Name	**Jun**
	Prefix	
	Suffix	
	Division	
	Organization	University of Michigan
	Address	Ann Arbor, MI, USA
	Email	junz@umich.edu

Abstract	The rho-tau embedding of a parametric statistical model defines both a Riemannian metric, called "rho-tau metric", and an alpha family of rho-tau connections. We give a set of equivalent conditions for such a metric to become Hessian and for the ± 1-connections to be dually flat. Next we argue that for any choice of strictly increasing functions $\rho(u)$ and $\tau(u)$ one can construct a statistical model which is Hessian and phi-exponential. The metric derived from the escort expectations is conformally equivalent with the rho-tau metric.

Keywords (separated by '-')	Hessian geometry - Dually-flat - rho-tau embedding - phi-exponential family - Escort probability

Information Geometry Under Monotone Embedding. Part II: Geometry

Jan Naudts[1(✉)] and Jun Zhang[2]

[1] Universiteit Antwerpen, Antwerpen, Belgium
jan.naudts@uantwerpen.be
[2] University of Michigan, Ann Arbor, MI, USA
junz@umich.edu

Abstract. The rho-tau embedding of a parametric statistical model defines both a Riemannian metric, called "rho-tau metric", and an alpha family of rho-tau connections. We give a set of equivalent conditions for such a metric to become Hessian and for the ± 1-connections to be dually flat. Next we argue that for any choice of strictly increasing functions $\rho(u)$ and $\tau(u)$ one can construct a statistical model which is Hessian and phi-exponential. The metric derived from the escort expectations is conformally equivalent with the rho-tau metric. AQ1

Keywords: Hessian geometry · Dually-flat · rho-tau embedding · phi-exponential family · Escort probability

1 Introduction

Amari [1,2] introduced the alpha family of connections $\Gamma^{(\alpha)}$ for a statistical model belonging to the exponential family. He showed that $\Gamma^{(\alpha)}$ and $\Gamma^{(-\alpha)}$ are each others dual and that for $\alpha = \pm 1$ the corresponding geometries are flat. Both the notions of an alpha family of connections and that of an exponential family of statistical models have been generalized. The present paper combines two general settings, that of the alpha family of connections determined by rho-tau embeddings [3] and that of phi-deformed exponential families [4].

Let $\mathcal{M}$ denote the space of probability density functions over the measure space $(\mathcal{X}, \mathrm{d}x)$. A parametric model p^θ is a map from some open domain in $\mathbb{R}^n$ into $\mathcal{M}$. It becomes a parametric statistical model if $\theta \to p^\theta$ is a Riemannian manifold with metric tensor $g(\theta)$.

Throughout the paper it is assumed that two strictly increasing functions ρ and τ are given. The rho-tau divergence (see Part I) induces a metric tensor g on finite-dimensional manifolds of probability distributions and makes them into Riemannian manifolds.

J. Naudts and J. Zhang contributed equally to this paper.

F. Nielsen and F. Barbaresco (Eds.): GSI 2017, LNCS 10589, pp. 1–8, 2017.
https://doi.org/10.1007/978-3-319-68445-1_25

2 The Metric Tensor

The rho-tau divergence $D_{\rho,\tau}(p,q)$ can be used [3,5,6] to define a metric tensor $g(\theta)$ by

$$g_{i,j}(\theta) = \partial_j\partial_i D_{\rho,\tau}(p,p^\theta)\Big|_{p=p^\theta}, \tag{1}$$

with $\partial_i = \partial/\partial\theta^i$. A short calculation gives

$$g_{ij}(\theta) = \int_{\mathcal{X}} \mathrm{d}x\, \left[\partial_i\tau(p^\theta(x))\right]\left[\partial_j\rho(p^\theta(x))\right]. \tag{2}$$

Because $\tau = f'\circ\rho$, the rho-tau metric $g(\theta)$ also takes the form:

$$\begin{aligned} g_{ij}(\theta) &= \int_{\mathcal{X}} \mathrm{d}x\, \left[\partial_i f'(\rho(p^\theta(x)))\right]\left[\partial_j\rho(p^\theta(x))\right] \\ &= \int_{\mathcal{X}} \mathrm{d}x\, f''(\rho(p^\theta(x)))\left[\partial_i\rho(p^\theta(x))\right]\left[\partial_j\rho(p^\theta(x))\right]. \end{aligned}$$

This shows that the matrix $g(\theta)$ is symmetric. Moreover, it is positive-definite, because the derivatives ρ' and f'' are strictly positive and the matrix with components $\left(\partial_j p^\theta(x)\right)\left(\partial_i p^\theta(x)\right)$ has eigenvalues 0 and 1 (assuming $\theta\to p^\theta$ has no stationary points). Finally, $g(\theta)$ is covariant, so g is indeed a metric tensor on the Riemannian manifold p^θ. From (2) follows that it is invariant under the exchange of ρ and τ.

The rho-tau entropy $S_{\rho,\tau}$ of the parametric family p^θ can be written as

$$S_{\rho,\tau}(p^\theta) = -\int_{\mathcal{X}} \mathrm{d}x\, f(\rho(p^\theta(x))). \tag{3}$$

So its second derivative

$$h_{ij}(\theta) = -\partial_i\partial_j S_{\rho,\tau}(p^\theta)$$

is symmetric in i,j. When positive-definite, $h(\theta)$ can also serve as a metric tensor as is found sometimes in the Physics literature.

Note that $h(\theta)$ differs from $g(\theta)$ in general: the former is induced by the entropy function $S_{\rho,\tau}(p)$, whose definition depends on the single function $f\circ\rho$, the latter is derived from the function $D_{\rho,\tau}(p,q)$.

3 Gauge Freedom

Write the rho-tau metric g_{ij} as

$$g_{ij}(\theta) = \int_{\mathcal{X}} \mathrm{d}x\, \frac{1}{\phi(p^\theta)}\left[\partial_i p^\theta(x)\right]\left[\partial_j p^\theta(x)\right], \tag{4}$$

where $\phi(u) = 1/(\rho'(u)\tau'(u))$. So despite of the two independent choices of embedding functions ρ and τ, the metric tensor g_{ij} is determined by one function ϕ only. More remarkably,

$$\begin{aligned} g_{ij}(\theta) &= \int_{\mathcal{X}} \mathrm{d}x\, f''(\rho(p^\theta(x))) \left[\partial_i(\rho(p^\theta(x))\right] \left[\partial_j \rho(p^\theta(x))\right] \\ &= \int_{\mathcal{X}} \mathrm{d}x\, (f^*)''(\tau(p^\theta(x))) \left[\partial_i \tau(p^\theta(x))\right] \left[\partial_j \tau(p^\theta(x))\right], \end{aligned}$$

so the gauge freedom in g_{ij} exists independent of the embedding – there is freedom in choosing an arbitrary function f in the case of the ρ-embedding and an arbitrary function f^* in the case of the τ-embedding of p^θ.

Without loss of generality, we choose τ-embedding and denote $X^\theta(x) = \tau(p^\theta(x))$. From the form of the rho-tau metric

$$g_{ij}(\theta) = \int_{\mathcal{X}} \mathrm{d}x\, \frac{\rho'(p^\theta(x))}{\tau'(p^\theta(x))} \left[\partial_i \tau(p^\theta(x))\right] \left[\partial_j \tau(p^\theta(x))\right],$$

we introduce a bilinear form $\langle \cdot, \cdot \rangle$ defined on pairs of random variables $u(x), v(x)$

$$\langle u, v \rangle_\theta = \int_{\mathcal{X}} \mathrm{d}x\, \frac{\rho'(p^\theta(x))}{\tau'(p^\theta(x))}\, u(x)\, v(x).$$

For any random variable u it holds that

$$\partial_j \int_{\mathcal{X}} \mathrm{d}x\, \rho(p^\theta(x)) u(x) = \int_{\mathcal{X}} \mathrm{d}x\, \frac{\rho'(p^\theta(x))}{\tau'(p^\theta(x))} \partial_j \tau(p^\theta(x)) u(x) = \langle \partial_j X^\theta, u \rangle_\theta$$

Following [2], $\partial_j X^\theta$ is then, by definition, tangent to the rho-representation $\rho(p^\theta)$ of the model p^θ. We also have

$$-\partial_j S_{\rho,\tau}(p^\theta) = \langle \partial_j X^\theta, X^\theta \rangle_\theta. \tag{5}$$

The difference of the metrics $g(\theta)$ and $h(\theta)$ can be readily appreciated:

$$g_{ij}(\theta) = \langle \partial_j X^\theta, \partial_i X^\theta \rangle_\theta$$

whereas

$$\begin{aligned} h_{ij}(\theta) = -\partial_i \partial_j S_{\rho,\tau}(p^\theta) &= \partial_i \langle \partial_j X^\theta, X^\theta \rangle_\theta \\ &= g_{ij}(\theta) + \int_{\mathcal{X}} \mathrm{d}x\, \tau(p^\theta(x)) \partial_i \partial_j \rho(p^\theta(x)). \end{aligned} \tag{6}$$

4 The Hessian Case

We now consider the condition under which the rho-tau metric g becomes Hessian.

Theorem 1. *Let be given a C^∞-manifold of probability distributions p^θ. For fixed strictly increasing functions ρ and τ, let the metric tensor $g(\theta)$ be given by (2). Then the following statements are equivalent:*

(i) g is Hessian, i.e., there exists $\Phi(\theta)$ such that

$$g_{ij}(\theta) = \partial_i\partial_j\Phi(\theta).$$

(ii) There exists a function $V(\theta)$ such that

$$\frac{\partial^2 V}{\partial\theta^i\partial\theta^j} = -\int_{\mathcal{X}} \mathrm{d}x\, \tau(p^\theta(x))\partial_i\partial_j\rho(p^\theta(x)). \tag{7}$$

(iii) There exists a function $W(\theta)$ such that

$$\frac{\partial^2 W}{\partial\theta^i\partial\theta^j} = -\int_{\mathcal{X}} \mathrm{d}x\, \rho(p^\theta(x))\partial_i\partial_j\tau(p^\theta(x)). \tag{8}$$

(iv) There exist coordinates $\eta_i(\theta)$ for which

$$g_{ij}(\theta) = \partial_j\eta_i.$$

(v) There exist coordinates ξ_i such that

$$\partial_j\xi_i(\theta) = -\int_{\mathcal{X}} \mathrm{d}x\, \tau(p^\theta(x))\partial_i\partial_j\rho(p^\theta(x)). \tag{9}$$

(vi) There exist coordinates ζ_i such that

$$\partial_j\zeta_i(\theta) = -\int_{\mathcal{X}} \mathrm{d}x\, \rho(p^\theta(x))\partial_i\partial_j\tau(p^\theta(x)). \tag{10}$$

Proof.
(i) $\longleftrightarrow$ (iv) This is well-known: the existence of a strictly convex function Φ is equivalent to the existence of dual coordinates η_i.
(ii) $\longleftrightarrow$ (v) From (ii) to (v): Given the existence of $V(\theta)$ satisfying (7), choose $\xi_i = \partial_i V$, and (9) is satisfied. From (v) to (ii): Since the right-hand side of (9) is symmetric with respect to i, j, we have $\partial_j\xi_i = \partial_i\xi_j$. Hence there exists a function $V(\theta)$ such that $\xi_i = \partial_i V$; this is the V function satisfying (7).
(iii) $\longleftrightarrow$ (vi) The proof is similar to the previous paragraph, by simply changing V to W and ξ to ζ.
(i) $\longleftrightarrow$ (ii) From the identity (6), the existence of $\Phi(\theta)$ to represent g_{ij} as its second derivatives allows us to choose the function V as $V = \Phi + S$. So from (i) we obtain (ii). Conversely when the integral term can be represented by the second derivative of $V(\theta)$, we can choose $\Phi = V - S$ that would satisfy (6). This yields (i) from (ii).
(i) $\longleftrightarrow$ (iii) The proof is similar to that of the previous paragraph, except that we will invoke the following identity instead of (6):

$$-\partial_i\partial_j S^*_{\rho,\tau}(p^\theta) = g_{ij}(\theta) + \int_{\mathcal{X}} \mathrm{d}x\, \rho(p^\theta(x))\partial_i\partial_j\tau(p^\theta(x)).$$

$\square$

The case when g is Hessian is very special, because of the existence of various bi-orthogonal coordinates.

The η_i are the dual coordinates of the θ^i. The ζ_i are called *escort coordinates.* They are linked to η_i by

$$\zeta_i = -\int_{\mathcal{X}} \mathrm{d}x\, \rho(p^\theta(x))\partial_i\tau(p^\theta(x)) + \eta_i = \partial_i S^*_{\rho,\tau}(p^\theta) + \eta_i. \tag{11}$$

They satisfy

$$\partial_j\partial_k\zeta_i = -\langle \partial_k X^\theta, \partial_i\partial_j X^\theta\rangle.$$

The *dual escort coordinates* ξ_i are given by

$$\xi_j(\theta) = \partial_j S_{\rho,\tau}(p^\theta) + \eta_j. \tag{12}$$

The Hessian of the function $V(\theta)$, when it does not vanish, causes a discrepancy between a metric tensor h defined as minus the Hessian of the entropy and the metric tensor g as defined by (2).

5 Zhang's rho-tau Connections

Given a pair of strictly increasing functions ρ and τ and a model p^θ, Zhang introduced the following connections [3]

$$\begin{aligned}\Gamma^{(\alpha)}_{ij,k} &= \frac{1+\alpha}{2}\int_{\mathcal{X}} \mathrm{d}x\, \left[\partial_i\partial_j\rho(p^\theta(x))\right]\left[\partial_k\tau(p^\theta(x))\right] \\ &\quad + \frac{1-\alpha}{2}\int_{\mathcal{X}} \mathrm{d}x\, \left[\partial_i\partial_j\tau(p^\theta(x))\right]\left[\partial_k\rho(p^\theta(x))\right],\end{aligned} \tag{13}$$

where $\Gamma^{(\alpha)}_{ij,k} \equiv (\Gamma^{(\alpha)})^l_{ij} g_{lk}$. One readily verifies

$$\Gamma^{(\alpha)}_{ij,k} + \Gamma^{(-\alpha)}_{jk,i} = \partial_i g_{jk}(\theta). \tag{14}$$

This shows that, by definition, $\Gamma^{(-\alpha)}$ is the dual connection of $\Gamma^{(\alpha)}$.

The coefficients of the connection $\Gamma^{(-1)}$ vanish identically if

$$\int_{\mathcal{X}} \mathrm{d}x\, \left[\partial_i\partial_j\tau(p^\theta(x))\right]\left[\partial_k\rho(p^\theta(x))\right] = 0. \tag{15}$$

This condition can be written as

$$\partial_j\partial_k\zeta_i = -\langle \partial_i\partial_j X^\theta, \partial_k X^\theta\rangle_\theta = 0. \tag{16}$$

It states that the escort coordinates are affine functions of θ and expresses that the second derivatives $\partial_i\partial_j X^\theta$ are orthogonal to the tangent plane of the statistical manifold. If satisfied then the dual of $\Gamma^{(-1)}$ satisfies

$$\Gamma^{(1)}_{ij,k} = \partial_i g_{jk}(\theta). \tag{17}$$

Likewise, the coefficients of the connection $\Gamma^{(1)}$ vanish identically if

$$\int_{\mathcal{X}} \mathrm{d}x \, \left[\partial_i\partial_j\rho(p^\theta(x))\right] \left[\partial_k\tau(p^\theta(x))\right] = 0. \tag{18}$$

Proposition 1. *With respect to conditions (15) and (18),*

1. *When (15) holds, the coordinates θ^i are affine coordinates for $\Gamma^{(-1)}$; the dual coordinates η_i are affine coordinates for $\Gamma^{(1)}$;*
2. *When (18) holds, the coordinates θ^i are affine coordinates for $\Gamma^{(1)}$; the dual coordinates η_i are affine coordinates for $\Gamma^{(-1)}$;*
3. *In either case above, $g(\theta)$ is Hessian.*

Proof.
One recalls that when $\Gamma = 0$ under a coordinate system θ, then θ^i's are affine coordinates – the geodesics are straight lines:

$$\theta(t) = (1-t)\theta_{(t=1)} + t\theta_{(t=0)}.$$

The geodesics of the dual connection Γ^* satisfies the Euler-Lagrange equations

$$\frac{\mathrm{d}^2}{\mathrm{d}t^2}\theta^i + \Gamma^i_{km}\left(\frac{\mathrm{d}}{\mathrm{d}t}\theta^k\right)\left(\frac{\mathrm{d}}{\mathrm{d}t}\theta^m\right) = 0. \tag{19}$$

Its solution is such that the dual coordinates η are affine coordinates:

$$\eta(t) = (1-t)\eta_{(t=1)} + t\eta_{(t=0)}.$$

For Statement 1, we apply the above knowledge, taking $\Gamma = \Gamma^{(-1)}$ and $\Gamma^* = \Gamma^{(1)}$; for Statement 2, taking $\Gamma = \Gamma^{(1)}$ and $\Gamma^* = \Gamma^{(-1)}$.
To prove Statement 3 observe that

$$\partial_k g_{ij}(\theta) = \int_{\mathcal{X}} \mathrm{d}x \, \left[\partial_i\tau(p^\theta(x))\right] \partial_j\partial_k\rho(p^\theta(x)) + \int_{\mathcal{X}} \mathrm{d}x \, \left[\partial_j\rho(p^\theta(x))\right] \partial_i\partial_k\tau(p^\theta(x)).$$

So the vanishing of either term, i.e., either (15) or (18) holding, will lead $\partial_k g_{ij}(\theta)$ to be symmetric in j,k or in i,k, respectively. This, in conjunction with the fact that g_{ij} is symmetric in i,j, leads to the conclusion that $\partial_k g_{ij}(\theta)$ is totally symmetric in an exchange of any two of the three indices i,j,k. This implies that η_i exist for which $g_{ij}(\theta) = \partial_j\eta_i$. That g is Hessian follows now from Theorem 1.

□

6 Rho-tau Embedding of phi-exponential Models

Let $\phi(u) = 1/(\rho'(u)\tau'(u))$ as before and fix real random variables $F_1, F_2, \cdots, F_n$. These functions determine a phi-exponential family $\theta \to p^\theta$ by the relation (see [4,7,8])

$$p^\theta(x) = \exp_\phi\left[\theta^k F_k(x) - \alpha(\theta)\right]. \tag{20}$$

The function $\alpha(\theta)$ is determined by the requirement that p^θ is a probability distribution and must be normalized to 1.

Assume that the integral

$$z(\theta) = \int_{\mathcal{X}} \mathrm{d}x\, \phi(p^\theta(x))$$

converges. Then the *escort family* of probability distributions $\tilde{p}^\theta$ is defined by

$$\tilde{p}^\theta(x) = \frac{1}{z(\theta)} \phi(p^\theta(x)).$$

The corresponding escort expectation is denoted $\tilde{\mathbb{E}}_\theta$. From the normalization of the p^θ follows that $\partial_i \alpha(\theta) = \tilde{\mathbb{E}}_\theta F_i$. Now calculate, starting from (4),

$$\begin{aligned} g_{ij}(\theta) &= \int_{\mathcal{X}} \mathrm{d}x\, \frac{1}{\phi(p^\theta(x))} \left[\partial_i p^\theta(x)\right] \left[\partial_j p^\theta(x)\right] \\ &= \int_{\mathcal{X}} \mathrm{d}x\, \phi(p^\theta(x)) \left[F_i - \partial_i \alpha(\theta)\right] \left[F_j - \partial_j \alpha(\theta)\right] \\ &= z(\theta) \left[\tilde{\mathbb{E}}_\theta F_i F_j - \tilde{\mathbb{E}}_\theta F_i \tilde{\mathbb{E}}_\theta F_j\right]. \end{aligned} \tag{21}$$

The latter expression is the metric tensor of the phi-exponential model as introduced in [4]. It implies that the rho-tau metric tensor is conformally equivalent with the metric tensor as derived from the escort expectation of the random variables F_i.

Finally, let $\eta_i = \mathbb{E}_\theta F_i$. A short calculation shows that

$$\begin{aligned} \partial_j \eta_i &= \int_{\mathcal{X}} \mathrm{d}x\, \phi(p^\theta(x)) \left[F_j - \partial_j \alpha(\theta)\right] F_i \\ &= z(\theta) \left[\tilde{\mathbb{E}}_\theta F_i F_j - \tilde{\mathbb{E}}_\theta F_i \tilde{\mathbb{E}}_\theta F_j\right] \\ &= g_{ij}(\theta). \end{aligned} \tag{22}$$

By (iv) of Theorem 1 this implies that the metric tensor g_{ij} is Hessian. Note that the η_i are dual coordinates. As defined here, they only depend on ϕ and not on the particular choice of embeddings ρ and τ. In particular, also the dually flat geometry does not depend on it.

One concludes that for any choice of strictly increasing functions $\rho(u)$ and $\tau(u)$ one can always construct statistical models for which the rho-tau metric is Hessian. These are phi-exponential models, with ϕ given by $\phi(u) = 1/\rho'(u)\tau'(u)$.

Conversely, given a phi-exponential model, its metric tensor is always a rho-tau metric tensor, with ρ, τ subject to the condition that $\rho'(u)\tau'(u) = 1/\phi(u)$. Two special cases are that either ρ or τ is the identity map, with the other being identified as the $\log_\phi$ function.

In the terminology of Zhang [3] the models of the phi-exponential family are called ρ-affine models where the normalization condition is, however, not imposed.

7 Discussion

This paper studies parametrized statistical models p^θ and the geometry induced on them by the choice of a pair of strictly increasing functions ρ and τ.

Theorem 1 gives equivalent conditions for the metric to be Hessian. It is shown that for the existence of a dually flat geometry the metric has to be Hessian.

The rho-tau metric tensor depends on a single function ϕ which is defined by $\phi(u) = 1/(\rho'(u)\tau'(u))$. If the model is phi-exponential for the same function ϕ then the rho-tau metric coincides with the metric used in the context of phi-exponential families and in particular the metric is Hessian. This shows that it is always possible to construct models which are Hessian for the given rho-tau metric.

Acknowledgement. The second author is supported by DARPA/ARO Grant W911NF-16-1-0383.

References

1. Amari, S.: Differential-Geometric Methods in Statistics. LNS, vol. 28. Springer, Heidelberg (1985). doi:10.1007/978-1-4612-5056-2
2. Amari, S., Nagaoka, H.: Methods of Information Geometry. Translations of Mathematical Monographs, vol. 191. Am. Math. Soc., Oxford University Press (2000). Originally in Japanese (Iwanami Shoten, Tokyo, 1993)
3. Zhang, J.: Divergence function, duality, and convex analysis. Neural Comput. **16**, 159–195 (2004)
4. Naudts, J.: Estimators, escort probabilities, and phi-exponential families in statistical physics. J. Ineq. Pure Appl. Math. **5**, 102 (2004)
5. Zhang, J.: Nonparametric information geometry: from divergence function to referential-representational biduality on statistical manifolds. Entropy **15**, 5384–5418 (2013)
6. Zhang, J.: On monotone embedding in information geometry. Entropy **17**, 4485–4499 (2015)
7. Naudts, J.: Generalised exponential families and associated entropy functions. Entropy **10**, 131–149 (2008)
8. Naudts, J.: Generalised Thermostatistics. Springer, London (2011). doi:10.1007/978-0-85729-355-8

Author Queries

Chapter 25

Query Refs.	Details Required	Author's response
AQ1	Please confirm if the corresponding author is correctly identified. Amend if necessary.	