View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by Elsevier - Publisher Connector

Available at WWW.MATHEMATICSWEB.ORG
POWERED BY SCIENCE @ DIRECT®

ACADEMIC PRESS

Journal of Multivariate Analysis 92 (2005) 97–115

Journal of Multivariate Analysis

http://www.elsevier.com/locate/jmva

Rao distances

Charles A. Micchelli[a,1] and Lyle Noakes[b,*,2]

[a] Department of Mathematics and Statistics, State University of New York at Albany, Albany, New York 12222, USA

[b] School of Mathematics and Statistics, The University of Western Australia, Nedlands, WA 6907, Australia

Received 14 March 2003

This paper is dedicated to Boris Korenblum with friendship and esteem on the occasion of his 80th birthday.

Abstract

We determine Riemannian distances between a large class of multivariate probability densities with the same mean, where the Riemannian metric is induced by a weighted Fisher information matrix. We reduce the evaluation of distances to quadrature and in some cases give closed form expressions.
© 2003 Elsevier Inc. All rights reserved.

AMS 2000 subject classifications: 62H99; 53C99

Keywords: Fisher information; Riemannian distances

1. Introduction

The problem of measuring the distance between probability densities is pervasive in applied sciences. Among other applications it comes up in applied statistics, speech recognition and image analysis. There are a number of approaches to this problem, but the one which is our focus of study here is the method introduced by Rao [12]. Generally speaking, the Rao method puts a Riemannian structure on the parameter space which determines the family of probability densities under

*Corresponding author.

E-mail addresses: cam@math.albany.edu (C. A. Micchelli), lyle@maths.uwa.edu.au (L. Noakes).

[1] Partially supported by NSF Grant DMS-9973427, and the University of Western Australia Distinguished Visitors Fund.

[2] Partially supported by an Australian Research Council Small Grant.

0047-259X/$ - see front matter © 2003 Elsevier Inc. All rights reserved.
doi:10.1016/S0047-259X(03)00132-5

consideration and then the distance between two probability densities is measured by the distance between their corresponding parametric realizations. Needless to say these distances are difficult to identify. There are a few cases for which they have been obtained. For example, the Riemannian distance induced by the Fisher information matrix between two multivariate normal densities with the *same* mean was identified in [2] as well as the distance between *any* two univariate normals. Other cases for which distances have been computed are multivariate normals with a non-Rao Riemannian metric [8], and elliptical densities with the same mean and the Rao Riemannian metric [3]. Other sources of interest in this connection are [1,6,11,13].

Recently one of us extended some results in [2] to a certain family of weighted Fisher information matrices as proposed in [4]. Our purpose here is to give improvements of this result in two directions. First, we consider a wide class of *elliptical densities* which include as a special case normal densities and secondly measure their Riemannian distances by a weighted Fisher information matrix which includes all the cases considered in [9]. In this generality, we shall demonstrate here that the computation of the Riemannian distance reduces to quadrature, that is, the computation of *univariate* integrals, and in many cases of interest can be obtained *explicitly*. In this regard, we exploit the invariance of the elliptical densities as reflected in the Riemannian structure on the parameter space.

Let us begin by establishing necessary terminology and notation. For any integer $m \geqslant 1$, let M be a C^∞ manifold of dimension m. We choose a $\theta \in M$, a real number $a \in R$ and a C^∞ function $\omega : R \to M$ satisfying the equation $\omega(a) = \theta$. The velocity $v = \dot{\omega}(a)$ of the curve at time a is said to be *tangent to* M at θ and the set of all such velocities is a real vector space of dimension m called the *tangent space* TM_θ *of* M at θ. Evidently, TM_θ is independent of the choice of $a \in R$. A *Riemannian metric* on M is a C^∞ assignment of an inner product on TM_θ for each $\theta \in M$,

$$\theta \in M \mapsto < \cdot, \cdot | \theta >$$

and with this metric is associated a *Riemannian norm* $|| \cdot |\theta||$ defined for $v \in TM_\theta$ by the equation $||v \ |\theta||^2 := \langle v, v|\theta \rangle$. For any nonempty finite interval $I := (a, b)$ of the real numbers R, the *Riemannian length* of ω restricted to I is defined as the integral

$$l(\omega; I) := \int_I ||\dot{\omega}(t)|\omega(t)||dt$$

and the *Riemannian distance* between $\theta_0, \theta_1 \in M$ is then defined to be

$$d(\theta_0, \theta_1) = \inf\{l(\omega; I) \colon \omega \in C^\infty(\theta_0, \theta_1, I, M)\}, \tag{1}$$

where $C^\infty(\theta_0, \theta_1, I, M)$ is the space of all C^∞ curves $\omega : I \to M$ satisfying the equations $\omega(a) = \theta_0$ and $\omega(b) = \theta_1$. Notice that the distance $d(\theta_0, \theta_1)$ is independent of the choice of I which we often take to be $[0, 1]$ and the distance d is said to be *induced* by the Riemannian metric on M. In this case, we simply write $C^\infty(\theta_0, \theta_1, M)$ for $C^\infty(\theta_0, \theta_1, [0, 1], M)$ and $l(\omega)$ for $l(\omega; [0, 1])$. The manifold M is said to be *complete* with respect to the Riemannian metric whenever, for every $\theta_0, \theta_1 \in M$ and every finite interval I there exists a curve $\gamma \in C^{\infty}(\theta_0, \theta_1, I, M)$ such that $d(\theta_0, \theta_1) = l(\gamma; I)$. In such a case there is no loss of generality in supposing γ *parameterised proportionally to arc-length*, namely $||\dot{\gamma}||$ constant, since lengths are unaffected by reparameterisations. Then γ minimises the *energy*

$$E(\gamma) := \int_I ||\dot{\omega}(t)|\omega(t)||^2 dt$$

among all curves in $C^{\infty}(\theta_0, \theta_1, I, M)$, as well as minimising length.

Definition 1. A curve $\gamma \in C^{\infty}(\theta_0, \theta_1, M)$ parameterised proportionally to arc-length is called a *geodesic* whenever for any $c \in I$ there exists a subinterval $J = (c_-, c_+)$ of I containing c such that

$$d(\gamma(c_-), \gamma(c_+)) = l(\gamma; J).$$

So a geodesic minimises the distance between sufficiently nearby points. In particular, given $\theta_0, \theta_1 \in M$, any $\gamma \in C^{\infty}(\theta_0, \theta_1, M)$ parameterised proportionally to arc-length and satisfying $d(\theta_0, \theta_1) = l(\gamma)$ is a geodesic. On the other hand, not all geodesics minimise length, and unless M is complete there might be no length-minimising curve joining given points in M.

The example described in the next section is central to our investigation. To prepare for it, we let $\{e_1, e_2, \ldots, e_n\}$ be the standard basis of R^n, $GL(n)$ be the group of invertible $n \times n$ real matrices, $SL(n)$ the subgroup of matrices of determinant 1, $O(n)$ denote the subgroup of orthogonal matrices and I the identity $n \times n$ matrix. We use R^n_+ for the *positive* orthant (all vectors with positive coordinates) in R^n and $C^{\infty}(M, N)$ for all C^{∞} functions from the manifold M to a manifold N and when $M = N$ we simply write $C^{\infty}(M)$ for $C^{\infty}(M, M)$.

2. Riemannian metrics

Let P_n be the space of $n \times n$ real symmetric positive definite matrices. The tangent space for any point (μ, Λ) on the manifold $M := R^n \times P_n$ is $R_n \oplus S_n$ where S_n is the vector space of *all* $n \times n$ real symmetric matrices. We consider a family of Riemannian norms induced by three functions a, b, c in $C^{\infty}(R_+, R)$. Specifically, at any point (μ, Λ) in M and any point (γ, Γ) in the tangent space we define $||(\gamma, \Gamma)|(\mu, \Lambda)||^2$ to be

$$a(\det \Lambda)\gamma^T \Lambda \gamma + b(\det \Lambda)(Tr(\Lambda^{-1}\Gamma))^2 + c(\det \Lambda)Tr((\Lambda^{-1}\Gamma)^2). \tag{2}$$

The choice of this norm comes from Statistics and will be explained in detail in the next section. For later use, we shall first discuss here some properties of this quadratic form.

A necessary and sufficient condition to ensure that this is indeed a norm on the tangent space for all points (μ, Λ) of M is that for all $t \in R_+$, there holds the inequalities $a(t) > 0$, $c(t) > 0$ and $nb(t) + c(t) > 0$. The necessity of this condition follows from simple choices of (μ, Λ) and (γ, Γ). The sufficiency of this assertion requires

$$Tr\, C^2 - n^{-1}(Tr\, C)^2 \geqslant 0, \tag{3}$$

valid for any square matrix C, with strict equality if and only if C is a multiple of I. This follows from Cauchy–Schwarz.

We have two comments to make about the Riemannian norm (2). First, the distance of any two points in M with the same *second* coordinate, that is, $d((\mu_0, \Lambda_0), (\mu_1, \Lambda_0))$ is the Euclidean distance between the vectors $a((\det \Lambda_0)^{\frac{1}{2}})\Lambda_0^{\frac{1}{2}}\mu_0$ and $a((\det \Lambda_0)^{\frac{1}{2}})\Lambda_0^{\frac{1}{2}}\mu_1$. The second comment concerns the computation of the distance between two points in M with the same *first* coordinate, that is, $d((\mu_0, \Lambda_0), (\mu_0, \Lambda_1))$. Since the first components are the same, we see that this distance is the same as the Riemannian distance between Λ_0, Λ_1 relative to the norm

$$||\Gamma|\Omega||^2 = b(\det \Lambda)(Tr(\Lambda^{-1}\Gamma))^2 + c(\det \Lambda)Tr((\Lambda^{-1}\Gamma)^2) \tag{4}$$

on the tangent space S_n of Λ as an element of the manifold P_n. By our previous comment for the metric (2), we see that (4) is a norm for all $\Lambda \in P_n$ and $\Gamma \in S_n$ if and only if for all $t \in R_+$ there holds the inequalities $c(t) > 0$ and $b(t) > -n^{-1}c(t)$. When the pair of functions (b, c) satisfy these conditions we say they are *acceptable*. Unless otherwise stated (b, c) will always be assumed to be acceptable.

There are important observations to be made about the computation of the Riemannian distances $d(\Lambda_0, \Lambda_1)$ induced by norm (4). These observations take the form of *reductions* which terminate at a calculation of Riemannian length in the plane R^2. Let us explain in detail what we have in mind. Every $\Omega \in SL(n)$ determines an automorphism on P_n given by $\Lambda \to \tilde{\Lambda} := \Omega^T \Lambda \Omega$ for $\Lambda \in P_n$ which takes the metric (4) into itself.

Now, let us explain how to choose Ω. We consider the matrix $\Gamma := \Lambda_1^{-\frac{1}{2}}\Lambda_0\Lambda_1^{-\frac{1}{2}} \in P_n$ and choose $U \in O(n)$ such that $\Delta_1 := U^T \Gamma U$ is a *diagonal* matrix. In this case, the diagonal elements of Δ_1 are the eigenvalues $\lambda_1, \ldots, \lambda_n$ of the matrix $\Lambda_0^{-1}\Lambda_1$. With this choice of U we set $\Omega := (\det \Lambda_0)^{1/2}\Lambda_0^{-\frac{1}{2}}U$ so that $\tilde{\Lambda}_0 = (\det \Lambda_0)I$ and $\tilde{\Lambda}_1 = (\det \Lambda_0)\Delta_1$ and we conclude that $d(\Lambda_0, \Lambda_1) = \tilde{d}(\tilde{\Lambda}_0, \tilde{\Lambda}_1)$, that is, it suffices to compute the distance between *a scalar multiple of the identity* matrix and a *diagonal* matrix.

Next, we restrict the Riemannian metric (4) to the manifold D_n of all *diagonal* matrices in P_n and let $\bar{d}$ be the corresponding distance in the *submanifold* D_n. Since the distance $\bar{d}$ is calculated as the infimum of lengths of curves in D_n we have for all $\Delta_0, \Delta_1 \in D_n$ that $d(\Delta_0, \Delta_1) \leqslant \bar{d}(\Delta_0, \Delta_1)$. We shall show that indeed

$$d(\Delta_0, \Delta_1) = \bar{d}(\Delta_0, \Delta_1). \tag{5}$$

Since P_n is not necessarily complete we do not prove this result by using geodesics. Instead, for $\varepsilon > 0$ and $\omega \in C^\infty(\Delta_0, \Delta_1, P_n)$ we choose
$\hat{\omega} \in C^\infty(\Delta_0, \Delta_1, P_n)$, so that

$$l(\hat{\omega}) < l(\omega) + \varepsilon$$

and the eigenvalues of $\hat{\omega}(t) \in P_n$ are *distinct* for all $t \in \bar{T}$ where $\bar{T} := [0,1] \backslash T$ for some finite subset T of $(0,1)$. From the definition of distance it is enough to find $\bar{\omega} \in C^\infty(\Delta_0, \Delta_1, D_n, \bar{T})$ with $l(\bar{\omega}) \leqslant l(\hat{\omega})$. To this end, using continuity of the spectrum of $\omega(t)$ as a set-valued function of t we write $\hat{\omega}(t)$ in the form

$$\hat{\omega}(t) = U(t)^T \bar{\omega}(t) U(t),$$

where $U \in C^\infty(\bar{T}, O(n))$ and $\bar{\omega} \in C^\infty(\Delta_0, \Delta_1, D_n, \bar{T})$ is C^∞, see [7]. For $t \notin int(T)$, $U^T(t)\dot{U}(t)$ is skew-symmetric and it follows that

$$Tr(\hat{\omega}^{-1}(t)\dot{\hat{\omega}}(t)) = Tr(\bar{\omega}^{-1}(t)\dot{\bar{\omega}}(t)).$$

Therefore, after some calculation we conclude that

$$Tr((\hat{\omega}^{-1}(t)\dot{\hat{\omega}}(t))^2) = Tr((\bar{\omega}^{-1}(t)\dot{\bar{\omega}}(t))^2) + 4Tr((AB)^2),$$

where $A := \hat{\omega}^{-1}(t)$ and B is the symmetric part of $\dot{U}^T \bar{\omega}(t) U$. We recall Eq. (3) to obtain $Tr((AB)^2) \geqslant 0$ and conclude that $l(\bar{\omega}) \leqslant l(\hat{\omega})$. This completes the proof of Eq. (5). The argument demonstrates that geodesics in D_n are also geodesics in P_n.

Identify R^n with D_n by $x = (x_1, x_2, \dots, x_n)^T \mapsto \text{diag}(e^{x_1}, e^{x_2}, \dots, e^{x_n})$, where diag means diagonal matrix. Let e be the vector $(1,1,\dots,1) \in R^n$. Then $(\text{logdet}\,\Delta_0)e$ corresponds to Δ_0, and we suppose $x^1 \in R^n$ corresponds to Δ_1. Define functions $\bar{b}, \bar{c}$ by $b(e^t) = \bar{b}(t), c(e^t) = \bar{c}(t)$ for $t \in R$. For a path $x(t)$ in R^n the Riemannian norm of $\dot{x}(t)$, induced by the norm on D_n, is

$$\begin{aligned}&\bar{b}(x_1 + x_2 + \cdots + x_n)(\dot{x}_1 + \dot{x}_2 + \cdots + \dot{x}_n)^2 + \bar{c}(x_1 + x_2 + \cdots + x_n) \\ &\quad \times (\dot{x}_1^2 + \dot{x}_2^2 + \cdots + \dot{x}_n^2).\end{aligned}$$

Let $H \in O(n)$ rotate e to $\sqrt{n}e_1$, and suppose also that H rotates x^1 into the plane spanned by e_1, e_2. Let $y = \sqrt{n}Hx$ and similarly $y^1 = \sqrt{n}Hx^1$. Then the Riemannian norm is

$$(\bar{b}(y_1) + n^{-1}\bar{c}(y_1))\dot{y}_1^2 + n^{-1}\bar{c}(y_1)(\dot{y}_2^2 + \cdots + \dot{y}_n^2). \tag{6}$$

This proves most of a theorem, whose statement requires the following definition.

Definition 2. Let $\chi \in C^\infty(R, R_+)$ be continuous. The associated χ-*Riemannian metric* $<,>_\chi$ on R^2 is defined by $||\dot{z}|z||^2 = \chi(z_1)(\dot{z}_1^2 + \dot{z}_2^2)$. □

Theorem 1. *Given functions* $b, c \in C^\infty(R_+, R)$ *satisfying* $nb(t) + c(t) > 0$, *for all* $t \in R_+$, *there is* $\chi \in C^\infty(R, R_+)$ *such that, for every* $\Lambda_0, \Lambda_1 \in P_n$ *there are* $z^0, z^1 \in R^2$ *with*

$$d(\Lambda_0, \Lambda_1) = d_\chi(z^0, z^1).$$

Moreover, $z^0 = (\log\det \Lambda_0)(1,1)$ *and* $z^1 = n(m, \sigma)$ *where*

$$m = \frac{1}{n}\sum_{i=1}^n \rho_i, \quad \sigma^2 = \frac{1}{n}\sum_{i=1}^n (\rho_i - m)^2,$$

$\rho_i = \log\lambda_i$ *for* $i = 1, 2, \dots, n$ *and the* λ_i *are the eigenvalues of* $\Lambda_0^{-1}\Lambda_1$.

Proof. Define $q : R^n \to R^n$ by $q(v) = (q_1(v), q_2(v), \dots, q_n(v))$, where $q_i(v) = v_i$ for $2 \leqslant i \leqslant n$, and

$$q_1(v) = \int_0^{v_1} \sqrt{1 + \frac{n\bar{b}(u)}{\bar{c}(u)}}\, du,$$

where the integral exists by hypothesis, and $v = (v_1, v_2, \dots, v_n) \in R^n$. Let $z^1 = q(y^1)$, where y^1 is defined in the discussion preceding Definition 2. For a curve $z(t)$ corresponding to $x(t)$, expression (6) reduces to

$$\chi(z_1)(\dot{z}_1^2 + \dot{z}_2^2 + \cdots + \dot{z}_n^2),$$

where $\chi(q_1(v)) = n^{-1}\bar{c}(v_1)$. Because $y^1 \in R^2 \times \{0\}$, it follows that $z^1 \in R^2 \times \{0\}$. So the Riemannian distance $d(0, y^1)$ in R^n, corresponding to the norm in (6), is achieved as an infimum of paths entirely contained in $R^2 \times \{0\}$. $\square$

The computation of d_χ is complex and the subject of Section 5. We raise, in passing, the following question. Given $F \in C(R^3, R_+)$, when is $F(\det \Lambda_0, m, \sigma)$ a distance on P_n? Next we turn to the statistical motivation for the Riemannian metric studied in Section 2.

3. Rao metrics

For us here a *probability density* on R^n is a measurable function $p : R^n \to R_+$ such that $\int_{R^n} p(x)\, dx = 1$. Let M be a C^∞ manifold of dimension m, and $P(\theta, \cdot), \theta \in M$ be a family of probability densities parameterised by M, that is, $P : M \times R^n \to R$ with $P(\theta, \cdot) \in C^\infty(R^n), \theta \in M$. The associated *log-likelihood* $L : M \times R^n \to R$ is given by $L := \log P$. For functions $\psi'' \in C^\infty(R_+, R)$ and $\omega \in C^\infty(R, M)$, we write $\omega(0) = \theta_0, \dot{\omega}(0) = v \in TM_{\theta_0}$ and define the (weighted) *Fisher information* of v as

$$\int_{R^n} \left(\frac{dL(\omega(t), x)}{dt}\Big|_{t=0} \right)^2 \psi''(-L(\theta_0, x))\, dx, \tag{7}$$

whenever the integral exists. For some families P and function ψ the Fisher information defines a *norm* on the vector space TM_{θ_0} for every choice of $\theta_0 \in M$. To elaborate on this point, we let U be an open neighbourhood of θ_0 in M, and $\varphi : U \to R^m$ a chart diffeomorphism satisfying $\varphi(\theta_0) = 0$. In chart coordinates $\theta \in M, v$ and L are represented respectively by

$$\bar{\theta} := \varphi(\theta), \quad \bar{v} := \frac{d(\varphi(\omega(t)))}{dt}\Big|_{t=0},$$

and $\bar{L}(\bar{\theta}, x) := L(\theta, x)$ for all $\theta \in M$ and $x \in R^n$. We conclude that the Fisher information of the v is the quadratic form $\bar{v}^T g(\theta_0)\bar{v}$ where $g(\theta_0)$ is the matrix whose entries are given by

$$g_{ij}(\theta_0) := \int_{R^n} \frac{\partial \bar{L}(\bar{\theta}_0, x)}{\partial \bar{\theta}_i} \frac{\partial \bar{L}(\bar{\theta}_0, x)}{\partial \bar{\theta}_j} \psi''(-\bar{L}(\bar{\theta}_0, x))\, dx, \quad i,j = 1, \ldots, m.$$

When $g(\theta_0) \in P_n$ for all $\theta_0 \in P_n$ and ψ is *strictly convex* the Fisher information is a norm. When this is the case we set

$$||v|\theta||^2 := \bar{v}^T g(\theta) \bar{v}, \quad v \in R^m, \quad \theta \in M. \tag{8}$$

Note that when the functions $\frac{\partial \bar{L}(\bar{\theta}_0, \cdot)}{\partial \bar{\theta}_i}$, $i = 1, \ldots, m$ are linearly independent on R^n for all $\theta \in M$ then $g(\theta_0) \in P_n$. Alternatively, integrating by parts we may express the elements of the matrix $g(\theta_0)$ in the form

$$g_{ij}(\theta_0) = \frac{\partial^2}{\partial \bar{\theta}_i \partial \bar{\theta}_j} \int_{R^n} \psi(-\bar{L}(\bar{\theta}, x))\, dx + \int_{R^n} \psi'(-\bar{L}(\bar{\theta}, x)) \frac{\partial^2 \bar{L}(\bar{\theta}, x)}{\partial \bar{\theta}_i \partial \bar{\theta}_j} dx \tag{9}$$

whenever the boundary terms are zero, and the right-hand side is evaluated at $\bar{\theta} = 0$. When $\psi(t) = e^{-t}, t \in R$, the weighted Fisher information is the usual Fisher information given by

$$g_{ij}(\theta_0) = -\int_{R^n} P(\bar{\theta}, x) \frac{\partial^2 \bar{L}(\bar{\theta}, x)}{\partial \bar{\theta}_i \partial \bar{\theta}_j} dx. \tag{10}$$

Let $SA(n)$ be the subgroup of invertible affine transformations of R^n whose linear parts have determinant ± 1, acting on the left of R^n in the standard way, and acting on the right of M.

Definition 3. We say L is *ample* when for all $h \in SA(n)$, $\theta \in M$ and $x \in R^n$ we have that

$$L(\theta h, x) = L(\theta, hx).$$

If L is ample then every $h \in SA(n)$ defines a diffeomorphism $R(h) : M \to M$ by the equation $R(h)(\theta) := \theta h, \theta \in M$. We let

$$dR(h)_{\theta_0} : TM_\theta \to TM_{R(h)\theta}$$

be its derivative at θ. For any $h \in SA(n), \theta \in M$ and $v \in TM_{\theta_0}$ we conclude from (7) and the change of variables formula for integration that

$$||v|\theta||^2 = ||dR(h)_\theta(v)|\theta||^2. \tag{11}$$

Let us now give a concrete example of an ample family of probability densities.

Definition 4. Let $M = R^n \times P_n$. For all $(\mu, \Lambda) \in M$ and $x \in R^n$ the family P of probability densities parameterised by M is *elliptical* when

$$\log P((\mu, \Lambda), x) = f(\det \Lambda, \frac{1}{2}(x - \mu)^T \Lambda (x - \mu)) \tag{12}$$

and $f : R^2 \to R_+$.

Since the function $P((\mu,\Lambda),\cdot)$ has integral 1 for all $(\mu,\Lambda)\in M$ Eq. (12) constrains f. In particular, we have for all $t\in R_+$ that

$$\int_{R_+} e^{f(t,\frac{r^2}{2})} r^{n-1} dr = ts_n, \tag{13}$$

where s_n is the $(n-1)$-dimensional area of the unit sphere S^{n-1} in R^n, namely

$$s_n = \frac{2\pi^{n/2}}{\Gamma(n/2)}.$$

Here we use the formula

$$\int_{R^n} g\left(\frac{1}{2}(x-\mu)^T\Lambda(x-\mu)\right) dx = s_n(\det\Lambda)^{-1/2}\int_{R_+} g\left(\frac{r^2}{2}\right) r^{n-1}\, dr, \tag{14}$$

where $g: R_+ \to R$ has the property that the integral on the right of (14) is absolutely convergent.

When $P((\mu,\lambda),\cdot)$ is elliptical the mean is μ. However, Λ^{-1} is not always the covariance (although it is for the normal density). Every elliptical family P is ample with right action of $SA(n)$ on $M = R^n \times P_n$ given by $(\mu,\Lambda)h = (h^{-1}\mu, \tilde{h}^T\Lambda\tilde{h})$. We also have that

$$dR(h)_{(\mu,\Lambda)}(\gamma,\Gamma) = (\tilde{h}^{-1}\gamma, \tilde{h}^T\dot{\Gamma}\tilde{h}),$$

where $\tilde{h}\in SL(n)$ is the linear part of $h\in SA(n)$. Consequently, Eq. (11) says that

$$||(\gamma,\Gamma)|(\mu,\Lambda)|| = ||(\tilde{h}^{-1}\gamma, \tilde{h}^T\Gamma\tilde{h})|(h^{-1}\mu, \tilde{h}^T\Lambda\tilde{h})||. \tag{15}$$

Theorem 2. *The Rao Riemannian metric of an ample family has form* (2).

The proof of this theorem turns on (15) and is given following the next two lemmas.

Lemma 1. *There exists a function $a\in C^\infty(R_+,R_+)$ such that for all $\lambda\in R_+, \Lambda\in P_n$ and $\gamma,\mu\in R^n$, we have that $||(\gamma,0)|(\mu,\Lambda)||^2 = a(\det\Lambda)\gamma^T\Lambda\gamma$.*

Proof. Given any $(\mu,\Lambda)\in R^n\times P_n$ we choose $U\in O(n)$ such that $\Delta := U^T\Lambda U$ is in D_n and set $\Omega = \det\Lambda^{1/2}U\Delta^{-1/2}$. Therefore, we conclude that

$$(\Omega^{-1})^T\Omega^{-1} = (\det\Lambda)^{-1}\Lambda \tag{16}$$

and

$$\Omega^T\Lambda\Omega = \det\Lambda I. \tag{17}$$

We now apply (15) with $h=\Omega$ and use (17) to obtain that

$$||(\gamma,0)|(\mu,\Lambda)||^2 = ||(\Omega^{-1}\gamma,0)|(\Omega^{-1}\mu,\det\Lambda I)||^2.$$

For any $(\mu, r) \in R^n \times R_+$ let $Q(\mu, r) \in P_n$ be such that for all $\gamma, \mu \in R^n$ we have that $||(\gamma, 0)|(\mu, rI)||^2 = \gamma^T Q(\mu, r)\gamma$. Consequently, we obtain that

$$||(\gamma, 0)|(\mu, \Lambda)||^2 = \gamma^T (\Omega^{-1})^T Q(\Omega^{-1}\mu, \det \Lambda)\Omega^{-1}\gamma. \tag{18}$$

Taking $h \in O(n)$ in (15), $Q(\mu, r) = hQ(\mu, r)h^T$. So $Q(\mu, r) = a_0(\mu, r)I$ where $a_0 > 0$. Taking h to be translation by μ in (15), $a_0(\mu, r) = a_0(0, r)$. Then from (18), (16)

$$||(\gamma, 0)|(\mu, \Lambda)||^2 = \gamma^T (\Omega^{-1})^T Q(\Omega^{-1}\mu, \det \Lambda)\Omega^{-1}\gamma = a(\det \Lambda)\gamma^T \Lambda\gamma, \tag{19}$$

where $a(\lambda) = \lambda^{-1} a_0(0, \lambda)$. $\square$

Lemma 2. *There exist functions $b, c \in C^\infty(R_+, R)$ such that for all $v \in R^n, \Lambda \in P_n$ and $\Gamma \in S_n$ it follows that*

$$||(0, \Gamma)|(\mu, \Lambda)||^2 = b(\det \Lambda)(Tr(\Lambda^{-1}\Gamma))^2 + c(\det \Lambda)Tr((\Lambda^{-1}\Gamma)^2).$$

Proof. As in the proof of Lemma 1, we observe for any $V \in O(n)$ that

$$\begin{aligned} ||(0, \Gamma)|(\mu, \Lambda)||^2 &= ||(0, \Omega^T \Gamma \Omega)|(\Omega^{-1}\mu, \det \Lambda I)||^2 \\ &= ||(0, \Omega^T \Gamma \Omega)|(0, \det \lambda I)||^2 = ||(0, V^T \Omega^T \Gamma \Omega V)|(0, \det \Lambda I)||^2. \end{aligned}$$

We choose $U \in O(n)$ so that $U^T \Omega^T \Gamma \Omega U$ is a diagonal matrix which we denote by $\bar{\Delta}$ and observe that

$$||(0, \Gamma)|(\mu, \Lambda)||^2 = ||(0, \bar{\Delta})|(0, \det \Lambda I)||^2. \tag{20}$$

Moreover, the diagonal entries of $\bar{\Delta}$ are the eigenvalues of the matrix $\Omega^T \Gamma \Omega$. Under conjugating by $U\Delta^{-1/2}$, this matrix is transformed to $\Lambda^{-1}\Gamma$. Therefore, we see that its eigenvalues $\bar{\lambda}_1, \bar{\lambda}_2, \ldots, \bar{\lambda}_n$ are the eigenvalues of $\Lambda^{-1}\Gamma$. The $\bar{\lambda}_i$ could occur in any order along the diagonal, depending on the choice of U. So the right-hand side of (20) is independent of the order, and quadratic in the $\bar{\lambda}_i$, namely

$$||(0, \Gamma)|(\mu, \Lambda)||^2 = b(\det \Lambda)(Tr(\bar{\Delta}))^2 + c(\det \Lambda)Tr(\bar{\Delta}^2),$$

where $b, c \in C^\infty(R_+, R)$. Now $Tr(\bar{\Delta}) = Tr(\Omega\Omega^T \Gamma) = \det \Lambda Tr(\Lambda^{-1}\Gamma)$, by (16). Similarly $Tr(\bar{\Delta}^2) = \det \Lambda^2 Tr((\Lambda^{-1}\Gamma)^2)$. This proves the lemma. $\square$

Proof of Theorem 2. Because of Lemmas 1 and 2 we need only show that

$$||(\gamma, \Gamma)|(\mu, \Lambda)||^2 = ||(-\gamma, \Gamma)|(\mu, \Lambda)||^2.$$

As in the proofs of Lemmas 1 and 2, it suffices to take $\mu = 0$. Then apply (15) with $h = -I$. $\square$

Invariance (15) can also be used to help determine the functions a, b, c for the elliptical family P. We consider this next.

4. Riemannian norm for elliptical families

In this section we demonstrate how to compute the functions a, b, c appearing in the Riemannian metric (2) generated by a Rao metric corresponding to an elliptical family of probability densities. Our main result in this connection is the following theorem.

Theorem 3. *For n an integer greater than 1, the Rao Riemannian metric of an elliptical family has form (2) with*

(i) $$a(t) = s_n t^{-1/2} \int_{R_+} r^{n-1}\left(f_v + \frac{r^2}{n} f_{vv}\right)\psi' \, dr,$$

(ii) $$b(t) = s_n t^{-1/2} \int_{R_+} r^{n-1}\left(\frac{\psi}{4} + t f_u \psi' + t^2 f_u^2 \psi'' + t\frac{r^2}{n} f_{uv}\psi' + \frac{r^4}{4n(n+2)} f_{vv}\psi'\right) dr,$$

(iii) $$c(t) = s_n t^{-1/2} \int_{R_+} r^{n-1}\left(\frac{\psi}{2} + \frac{r^4}{2n(n+2)} f_{vv}\psi'\right) dr,$$

where ψ and its derivatives are evaluated at $-f$. Also f and its partials are evaluated at $(t, \frac{r^2}{2})$, and $t \in R_+$.

The proof of this result is a difficult computation which requires some preparation. We need two distinct types of formulas. The first type concerns derivatives of $\det \Lambda$ for $\Lambda \in P_n$ as a function of the elements $\Lambda_{ij} = \Lambda_{ji}$ parameterising positive definite symmetric matrices Λ. of the matrix Λ.

Lemma 3. *For every $\Lambda = (\Lambda_{ij})_{i,j=1,\dots,n} \in P_n$, we have that*

$$\frac{\partial \det \Lambda}{\partial \Lambda_{ij}} = (2 - \delta_{ij})(\Lambda^{-1})_{ij} \det \Lambda \tag{21}$$

and

$$\frac{\partial^2 \det \Lambda}{\partial \Lambda_{kl}\partial \Lambda_{ij}} = \frac{(2-\delta_{ij})(2-\delta_{kl})}{2} \Psi \det \Lambda,$$

where

$$\Psi = (2(\Lambda^{-1})_{ij}(\Lambda^{-1})_{kl} - (\Lambda^{-1})_{ik}(\Lambda^{-1})_{lj} - (\Lambda^{-1})_{il}(\Lambda^{-1})_{kj}). \tag{22}$$

Proof. The proof of (21) is by Cramer's rule for the inverse of Λ and Laplace's expansion by minors for $\det \Lambda$. One first differentiates in nontangent directions, permitting Λ to move freely in the space of all $n \times n$ matrices, and giving (21) without the factor $(2-\delta_{ij})$. When Λ is constrained to P_n a second term appears (equal to the first since matrices in P_n are symmetric) except when $i=j$. This explains the factor $(2-\delta_{ij})$. Differentiating both sides of (21) and simplification gives (22). The second type of computation reduces integrals of certain spherically symmetric functions on R^n to integrals over R_+. $\square$

Lemma 4.

(i) $$\int_{R^n} x_1^2 g\left(\frac{1}{2}x^T x\right) dx = \frac{s_n}{n}\int_{R_+} g\left(\frac{r^2}{2}\right) r^{n+1}\, dr,$$

(ii) $$\int_{R^n} x_1^4 g\left(\frac{1}{2}x^T x\right) dx = \frac{3s_n}{n(n+2)}\int_{R_+} g\left(\frac{r^2}{2}\right) r^{n+3}\, dr,$$

(iii) $$\int_{R^n} x_1^2 x_2^2 g\left(\frac{1}{2}x^T x\right) dx = \frac{s_n}{n(n+2)}\int_{R_+} g\left(\frac{r^2}{2}\right) r^{n+3}\, dr,$$

where $g : R_+ \to R$ is of rapid decrease.

Proof. For the proof of (i) we use

$$\tau \int_{R^n} g\left(\frac{1}{2}(\tau^2 x_1^2 + x_2^2 + x_3^2 + \cdots + x_n^2)\right) dx = s_n \int_{R_+} g\left(\frac{r^2}{2}\right) r^{n-1}\, dr,$$

where $\tau \in R_+$, obtained from (14) with $\mu = 0$ and $\Lambda = \mathrm{diag}(\tau, 1, 1, \ldots, 1)$. Differentiate both sides with respect to τ, set $\tau = 1$, and simplify to obtain

$$\int_{R^n} x_1^2 g\left(\frac{1}{2}x^T x\right) dx = -s_n \int_{R_+} h\left(\frac{r^2}{2}\right) r^{n-1}\, dr,$$

where $h(t) = -\int_t^\infty g(s)\, ds$, $t \in R_+$. Integration by parts on the right completes the proof of (i). To prove (iii) reinsert the scale τ into (i), namely

$$\tau^3 \int_{R^n} x_1^2 g\left(\frac{1}{2}(\tau^2 x_1^2 + x_2^2 + x_3^2 + \cdots + x_n^2)\right) dx = \frac{s_n}{n}\int_{R_+} g\left(\frac{r^2}{2}\right) r^{n+1}\, dr,$$

and, as before, we obtain (iii). Formula (ii) is obtained in a similar fashion, but scaling x_2 instead of x_1. $\square$

More general formulae might of course be derived, and by other means, but these are all we need. This completes the preliminaries needed for the proof of Theorem 3.

Proof of Theorem 3. To compute a use Lemma 1 with $\gamma = (1, 0, 0, \ldots, 0)^T$, $\mu = (0, 0, \ldots, 0)^T$ and $\Lambda = \sigma I$, where $\sigma \in R_+$. Consequently, we have that

$\sigma a(t) = ||(\gamma, 0)|(\mu, \Lambda)||^2$ where $\sigma^n = t$. By (9), it follows that

$$\sigma a(t) = \frac{\partial^2 H}{\partial \mu_1^2} + \sigma \int_{R^n} f_v \psi' \, dx + \sigma^2 \int_{R^n} x_1^2 f_{vv} \psi' \, dx, \tag{23}$$

where

$$H = \int_{R^n} \psi(-f(\log \det \Lambda, \frac{1}{2} x^T \Lambda x)) \, dx. \tag{24}$$

The integrals on the right of (23) are evaluated by Lemma 4, proving (i). We now discuss the computation of b and c. For $1 \leqslant i \leqslant j \leqslant n$, let $\Xi_{ij} \in S_n$ have entries 1 in row i and column j, 1 in row j and column i, and zeroes elsewhere. Then $\{\Xi_{ij}\colon 1 \leqslant i \leqslant j \leqslant n\}$ is a basis for S_n. From (9), we have that

$$\begin{aligned} ||(0, \Xi_{ij})|(0, \sigma I)||^2 &= \frac{\partial^2 H}{\partial \Lambda_{ij}^2} \\ &+ \int_{R^n} (\delta_{ij} \sigma^{2n-2} f_{uu} - 2(1 - \delta_{ij}) \sigma^{n-2} f_u + \delta_{ij} \sigma^{n-1} f_{uv} x_i x_j \\ &+ \frac{1}{4} (4 - 3\delta_{ij}) f_{vv} x_i^2 x_j^2) \psi' \, dx \end{aligned} \tag{25}$$

where ψ, ψ' are evaluated at $-f$. Also f and its partials are evaluated at $(t, \frac{\sigma}{2} x^T x)$. By Lemma 2

$$\sigma^2 ||(0, \Xi_{ij})|(0, \sigma I)||^2 = b(t)\delta_{ij} + c(t)(2 - \delta_{ij}), \tag{26}$$

where $t = \sigma^n$. We specialize this to $i = j = 1$ and to $i = 1, j = 2$, obtaining two equations for b, c yielding

$$\begin{aligned} b(t) = \sigma^2 \left(\frac{\partial^2 H}{\partial \Lambda_{11}^2} - \frac{1}{2} \frac{\partial^2 H}{\partial \Lambda_{12}^2} \right) + \int_{R^n} \bigg(& t f_u + t^2 f_{uu} + t \sigma f_{uv} x_1^2 + \frac{\sigma^2}{4} f_{vv} x_1^4 \\ & - \frac{\sigma^2}{2} f_{vv} x_1^2 x_2^2 \bigg) \psi' \, dx, \end{aligned}$$

$$c(t) = \frac{\sigma^2}{2} \frac{\partial^2 H}{\partial \Lambda_{12}^2} + \int_{R^n} \left(-t f_u + \frac{\sigma^2}{2} f_{vv} x_1^2 x_2^2 \right) \psi' \, dx.$$

Now we reduce these expressions to integrals over R_+. For $\Lambda \in P_n$ set $\lambda = \det \Lambda$. Then, from (14) we have

$$H = \lambda^{-1/2} s_n \int_{R_+} \psi \left(-f \left(\lambda, \frac{r^2}{2} \right) \right) r^{n-1} \, dr.$$

Therefore, we have that

$$\frac{\partial H}{\partial \Lambda_{ij}} = -s_n \frac{\partial \lambda}{\partial \Lambda_{ij}} \int_{R_+} r^{n-1} \left(\frac{1}{2} \lambda^{-3/2} \psi + \lambda^{-1/2} f_u \psi' \right) dr$$

and

$$\frac{\partial^2 H}{\partial \Lambda_{ij}^2} = -s_n \frac{\partial^2 \lambda}{\partial \Lambda_{ij}^2} \int_{R_+} r^{n-1} \left(\frac{1}{2} \lambda^{-3/2} \psi + \lambda^{-1/2} f_u \psi' \right) dr$$
$$+ s_n \left(\frac{\partial \lambda}{\partial \Lambda_{ij}} \right)^2 \int_{R_+} r^{n-1} \left(\frac{3}{4} \lambda^{-5/2} \psi + \lambda^{-3/2} f_u \psi' + \lambda^{-1/2} f_u^2 \psi'' - \lambda^{-1/2} f_{uu} \psi' \right) dr.$$

On the right ψ and its derivatives are evaluated at $-f$, and f and its partials are evaluated at $(\lambda, \frac{r^2}{2})$.

We use these formulas for the derivatives of H to evaluate the right-hand side of the expressions for $b(t)$ and $c(t)$ at $\Lambda = \sigma I$, proving the theorem. $\square$

Corollary 1. *Let* $f(u, v) = \frac{1}{2}\log u - \alpha - \beta v^{\gamma}$ *where* $\beta > 0, \gamma > 0$, *and*

$$\alpha = \frac{n}{2}\log(2\pi) - \frac{n}{2\gamma}\log\beta - \log\gamma - \log\Gamma\left(\frac{n}{2}\right) + \log\Gamma\left(\frac{n}{2\gamma}\right),$$

(*so that* (13) *is satisfied*), *and for any* $\delta, t \in R_+$ *define* $\psi(t) = e^{-\delta t}$. *Then, for all* $t \in R_+$,

$$(a(t), b(t), c(t)) = t^{(\delta-1)/2} w,$$

where $w \in R^3$ *depends on* n, β, γ, δ. *Moreover, there exist* $p, q \in R$ *such that* $\chi(t) = pe^{qt}$ *where* $t \in R$. *For normal distributions and* $\delta = 1$, *we have that* $(a, b, c) = (1, 0, \frac{1}{2})$. $\square$

5. Computing distances

Given a C^{∞} function $\chi : R \to (0, \infty)$, we refer to geodesics of $<,>_{\chi}$ as χ-*geodesics* (note that $<,>_{\chi}$ is conformal to the Euclidean inner product). The associated distance function d_{χ} is called the χ-*distance*. Corollary 1 reduces calculation of Rao distances to χ-distances. The next result reduces computation of d_{χ} to finding χ-geodesics, and calculating their lengths with respect to $<,>_{\chi}$.

Theorem 4. *Let* $x, y \in R^2$. *Then* $d_{\chi}(x, y)$ *is either the length of a shortest* χ-*geodesic from* x *to* y, *or*

$$\left(\int_{x_1}^{\infty} + \int_{y_1}^{\infty} \right) \sqrt{\chi(t)}\, dt, \quad \text{or} \quad \left(\int_{-\infty}^{x_1} + \int_{-\infty}^{y_1} \right) \sqrt{\chi(t)}\, dt,$$

whichever is smallest.

Proof. Define $\xi : R \to R_+$ by $\xi(t) = \sqrt{\chi(t)}$. Consider first the case where ξ is not bounded uniformly away from 0 on the whole of R. Then either $\liminf_{t\to\infty} \xi(t) = 0$ or $\liminf_{t\to-\infty} \xi(t) = 0$. If $\liminf_{t\to\infty} \xi(t) = 0$, choose an increasing sequence $\{t_i : i = 1, 2, \ldots\}$, where $\lim_{i\to\infty} t_i = \infty$, $\lim_{i\to\infty} \xi(t_i) = 0$ and $t_1 > 1/3$. Define a sequence $\Omega_+ = \{\omega^{(i)} : i \geq 1\}$ of piecewise-C^1 curves $\omega_+^{(i)} : [0, 1] \to R^2$

from x to y by

$$\omega_+^{(i)}(t) = \begin{cases} (x_1 + 3t(t_i - x), x_2) & \left(0 \leqslant t \leqslant \frac{1}{3}\right), \\ (t_i, x_2 + (3t-1)(y_2 - x_2)) & \left(\frac{1}{3} \leqslant t \leqslant \frac{2}{3}\right), \\ ((3t-2)y_1 + 3(1-t)t_i, y_2) & \left(\frac{2}{3} \leqslant t \leqslant 1\right). \end{cases}$$

The length $\lambda_\chi(\omega_+^{(i)})$ of $\omega^{(i)}$ with respect to the Riemannian metric $<,>_\chi$ is

$$\int_{x_1}^{t_i} \xi(t)\,dt + \xi(t_i)|y_2 - x_2| + \int_{y_1}^{t_i} \xi(t)\,dt.$$

Therefore, and because $\lim_{i\to\infty} \xi(t_i) = 0$,

$$d_\chi(x,y) \leqslant \lim_{i\to\infty} \lambda_\chi(\omega_+^{(i)}) = \left(\int_{x_1}^{\infty} + \int_{y_1}^{\infty}\right)\xi(t)\,dt.$$

Similarly $d_\chi(x,y) \leqslant (\int_{-\infty}^{x_1} + \int_{-\infty}^{y_1})\xi(t)dt$ when $\liminf_{t\to-\infty} \xi(t) = 0$. It follows, whether ξ is bounded away from 0 or not, that

$$d_\chi(x,y) \leqslant \delta \equiv \min\left\{\left(\int_{x_1}^{\infty} + \int_{y_1}^{\infty}\right)\xi(t)\,dt, \left(\int_{-\infty}^{x_1} + \int_{-\infty}^{y_1}\right)\xi(t)\,dt\right\}.$$

This proves the theorem, except when $d_\chi(x,y) < \delta$.

In such a case set $\varepsilon = \frac{1}{2}(\delta - d_\chi(x,y))$. Then $\varepsilon > 0$. Let $\boldsymbol{\omega}$: $[0,1] \to R^2$ be a piecewise-C^1 curve from x to y, and let $t_0 \in [0,1]$. If $\boldsymbol{\omega}_1(t_0) \geqslant \max\{x_1, y_1\}$ then

$$\begin{aligned} \lambda_\chi(\omega) &\geqslant \int_0^1 \xi(\omega_1(t))|\dot{\omega}_1(t)|\,dt = \left(\int_0^{t_0} + \int_{t_0}^1\right)\xi(\omega_1(t))|\dot{\omega}_1(t)|\,dt \\ &\geqslant \left(\int_{x_1}^{\omega_1(t_0)} + \int_{y_1}^{\omega_1(t_0)}\right)\xi(y)\,dy. \end{aligned}$$

Similarly, if $\omega_1(t_0) \leqslant \min\{x_1, y_1\}$ then $\lambda_\chi(\omega) \geqslant (\int_{\boldsymbol{\omega}_1(t_0)}^{x_1} + \int_{\boldsymbol{\omega}_1(t_0)}^{y_1})\xi(y)\,dy$. Using these facts, choose M so large that, for any piecewise-C^1 curve $\omega : [0,1] \to R^2$ from x to y, and any Riemannian metric $<,>_{\hat{\chi}}$ which agrees with $<,>_\chi$ over $[-M, M] \times R$, either

- $\omega[0,1] \subset [-M, M] \times R$, or
- $\lambda_{\hat{\chi}}(\omega) \geqslant d_\chi(x,y) + \varepsilon$.

Let Ω be a sequence of piecewise-C^1 curves $\omega^{(j)} : [0,1] \to [-M, M] \times R$ from x to y, such that

$$\lim_{j\to\infty} \lambda_\chi(\omega^{(j)}) = d_\chi(x,y).$$

Let $\hat{\chi} : R \to R$ be a C^2 extension of $\chi \mid [-M, M]$ with the property that, for some $0 < \iota \leqslant \sigma$, we have $\iota \leqslant \hat{\chi}(t) \leqslant \sigma$ for all $t \in R$. For vectors v whose norms are measured at

$z \in R^2$, $\imath ||v|| \leqslant ||v||_{\hat{\chi}(z_1)} \leqslant \sigma ||v||$. Riemannian and Euclidean distances are related in the same way, namely

$$\imath ||x-y|| \leqslant d_{\hat{\chi}}(x,y) \leqslant \sigma ||x-y||. \tag{27}$$

By (27), and because R^2 is complete with respect to the Euclidean metric, R^2 is complete with respect to $d_{\hat{\chi}}$. By the Hopf–Rinow Theorem [10] Theorem 10.9, $d_{\hat{\chi}}(x,y) = \lambda_{\hat{\chi}}(\hat{\omega})$, where $\hat{\boldsymbol{\omega}} : [0,1] \to R^2$ is a $\hat{\chi}$-geodesic from x to y. Because the $\omega^{(i)}$ are curves in $[-M, M] \times R$, and $d_{\hat{\chi}}$ agrees with d_χ over $[-M, M] \times R$, $d_\chi(x,y) = \lambda_{\hat{\chi}}(\hat{\boldsymbol{\omega}})$.

Because $\lambda_{\hat{\chi}}(\hat{\omega}) < d_\chi(x,y) + \varepsilon$, $\hat{\omega}$ is a curve in $[-M, M] \times R$, and therefore a χ-geodesic with

$$\lambda_\chi(\hat{\omega}) = \lambda_{\hat{\chi}}(\hat{\omega}) = d_\chi(x,y). \qquad \square$$

Notice, in the statement of Theorem 4, that there might be no χ-geodesic from x to y. Also, either or both of the improper integrals might be infinite.

Corollary 2. *If χ is bounded away from* 0 *on the whole of R, then $d_\chi(x,y)$ is the length of a shortest χ-geodesic joining x, y.* $\square$

This begs the question of how to find χ-geodesics from x to y. Sometimes closed form expressions can be found, but in general the problem reduces to calculations of univariate integrals, as follows.

The form of the Riemannian metric $\langle \ , \ \rangle_\chi$ can be used to simplify the Euler–Lagrange equations for geodesics, for instance using Clairaut patches [5, 26.2]. Alternatively, we can proceed directly, as follows. The Lagrangian for χ-geodesics $\gamma : [0,1] \to R^2$ is $L = \chi(x_1)(\dot{x}_1^2 + \dot{x}_2^2)$, and the Euler–Lagrange equations are

$$2\chi\ddot{\gamma}_1 = -\chi'(\dot{\gamma}_1^2 - \dot{\gamma}_2^2) \quad \text{and} \quad \frac{d}{dt}(2\chi\dot{\gamma}_2) = 0,$$

where χ' is the derivative of $\chi, \dot{\gamma}$ is the derivative of $\gamma(t)$ with respect to $t \in [0,1]$, and χ, χ' are evaluated at $\gamma_1(t)$. These equations integrate to give

$$\dot{\gamma}_1^2 + \dot{\gamma}_2^2 = \frac{a^2}{\chi} \quad \text{and} \quad \dot{\gamma}_2 = \frac{b}{\chi},$$

where $a, b \in R$ are constants of integration, and $a \geqslant 0$ is the length $l_\chi(\gamma)$ of γ with respect to the Riemannian metric $\langle \ , \ \rangle_\chi$. Set $x = \gamma(0)$, $y = \gamma(1)$ and suppose $x \neq y$. Then $a > 0$. When $b = 0$, $\dot{\gamma}_2$ is constant and the χ-geodesic γ is said to be *horizontal*. Define $\xi = \sqrt{\chi} : R \to R_+$. When γ is horizontal, $x_2 = y_2$, $l_\chi(\gamma) = |\int_{x_1}^{y_1} \xi(u)du|$, and $\gamma_1(t)$ is given implicitly, by the equation

$$\int_{x_1}^{\gamma_1(t)} \xi(u)du = \pm at.$$

So it remains only to calculate lengths of non-horizontal χ-geodesics. We describe how to do this in the simplest non-trivial case, where χ' is everywhere positive. Then the range $R(\chi)$ of χ is an open interval of positive numbers. Define $F, G : R^2 \times R \to R$ by

$$F(c,d,v) = \int_d^v \frac{\chi(u)}{\sqrt{\chi(u)-c}}\,du \quad \text{and} \quad G(c,d,v) = \int_d^v \frac{1}{\sqrt{\chi(u)-c}}\,du,$$

where the improper integrals converge because every $c \in R(\chi)$ is a regular value of χ.

Theorem 5. *Suppose $\chi'(v) > 0$ for all $v \in R$. Let $\gamma : [0,1] \to R^2$ be a non-horizontal χ-geodesic from x to y, where $y_2 > x_2$. Then, for some $c \in (0, \min\{\chi(x_1), \chi(y_1)\})$, either*

(i) *$c \in R$, $G(c, \chi^{-1}(c), x_1) + G(c, \chi^{-1}(c), y_1) = \frac{y_2 - x_2}{\sqrt{c}}$, and γ has length*

$$|F(c, \chi^{-1}(c), x_1) + F(c, \chi^{-1}(c), y_1)|,$$

or

(ii) *$|G(c, x_1, y_1)| = \frac{y_2 - x_2}{\sqrt{c}}$, and γ has length $|F(c, x_1, y_1)|$.*

Proof. Call $t \in (0,1)$ a *fold* of γ when $\dot{\gamma}_1(t) = 0$. Then $\chi(\gamma_1(t)) = \frac{b^2}{a^2}$. From the Euler–Lagrange equations,

$$2\chi(\gamma(t))\ddot{\gamma}_1(t) = \chi'\dot{\gamma}_2(t)^2 = b^2\frac{\chi'}{\chi^2}$$

whose sign is that of χ'. So $\ddot{\gamma}_1(t) \neq 0$, with the same sign as χ'. So the folds of γ comprise a discrete subset D of $(0,1)$. We claim that D has cardinality at most 1. Suppose, to the contrary, that D has more than one element. If $t_0, t_1 \in D$ with $t_0 < t_1$ and $(t_0, t_1) \cap D = \emptyset$, then

$$\chi(\gamma_1(t_0)) = \frac{b^2}{a^2} = \chi(\gamma_1(t_1))$$

and $\chi'(\gamma_1(s))\dot{\gamma}_1(s) = 0$ for some $s \in (t_0, t_1)$. Because χ' is nowhere-zero, s is a fold. The contradiction proves our claim, and γ has at most one fold.

Because the χ-geodesic γ is not horizontal, $\dot{\gamma}_2$ has constant sign, and the coordinate γ_2 may be used to parameterise γ. Call this the *vertical parameterisation* of γ. With respect to the vertical parameterisation

$$\frac{d\gamma_1}{d\gamma_2} = \pm\sqrt{\frac{a^2\chi(\gamma_1) - b^2}{b^2}}. \tag{28}$$

For a non-empty open interval $(s_0, s_1) \subset [0,1]$ that does not contain folds, we have

$$\int_{\gamma_1(s_0)}^{\gamma_1(s_1)} \frac{\chi(u)du}{\sqrt{\chi(u) - \frac{b^2}{a^2}}} = \sigma a(s_1 - s_0) \tag{29}$$

and

$$\int_{\gamma_1(s_0)}^{\gamma_1(s_1)} \frac{du}{\sqrt{\chi(u) - \frac{b^2}{a^2}}} = \sigma\frac{a(\gamma_2(s_1) - \gamma_2(s_0))}{b}, \tag{30}$$

where $\sigma = \pm 1$ is the sign of $\gamma_1(s_1) - \gamma_1(s_0)$. Set $c = \frac{b^2}{a^2}$, and consider first the case where γ has a fold $s \in (0, 1)$. Then s is a point of local minimum of γ_1, and $\chi(\gamma_1(s)) = c$. Applying (29), (30) along the interval $(0, s)$,

$$F(c, \chi^{-1}(c), x_1) = as \quad \text{and} \quad G(c, \chi^{-1}(c), x_1) = \frac{\gamma_2(s) - x_2}{\sqrt{c}}.$$

Applying (29), (30) along the interval $(s, 1)$ we similarly obtain

$$F(c, \chi^{-1}(c), y_1) = a(1 - s) \quad \text{and} \quad G(c, \chi^{-1}(c), y_1) = \frac{y_2 - \gamma_2(s)}{\sqrt{c}}.$$

Addition of the pairs of equations gives (i). Suppose next that γ has no fold. Applying (29), (30) along the interval $(0, 1)$ gives (ii). The theorem is proved. □

To apply Theorem 5, in either of the cases (i), (ii), the first equation is solved for c, and then c is substituted in the second equation to give the length of γ. In connection with Corollary 1, we have the following result, proved also in [9].

Corollary 3. *Let $\chi(t) = e^{2t}$, for $t \in R$. Then for $x, y \in R^2$, the Riemannian distance $d_\chi(x, y)$ is*

$$\sqrt{e^{2x_1} + e^{2y_1} - 2e^{x_1+y_1}\cos(y_2 - x_2)} \quad \text{or} \quad e^{x_1} + e^{y_1},$$

according as $|y_2 - x_2| < \pi$ or not.

Proof. $F(c, d, v) = \sqrt{e^{2v} - c} - \sqrt{e^{2d} - c}$, and

$$G(c, d, v) = \frac{\arccos(\frac{\sqrt{c}}{e^v}) - \arccos(\frac{\sqrt{c}}{e^d})}{\sqrt{c}}.$$

The conditions in Theorem 5(i), (ii) are then

$$\left| \arccos\left(\frac{\sqrt{c}}{e^{x_1}}\right) \pm \arccos\left(\frac{\sqrt{c}}{e^{y_1}}\right) \right| = y_2 - x_2, \tag{31}$$

where the + sign is taken in case (i), and − for (ii). When $0 < y_2 - x_2 < \pi$, (31) has solution

$$c = \frac{e^{2x_1+2y_1}\sin^2(y_2 - x_2)}{e^{2x_1} + e^{2y_1} - 2e^{x_1+y_1}\cos(y_2 - x_2)}.$$

To tell whether to apply case (i) or (ii) of Theorem 5, check the sign for which (31) holds. Then, by Theorem 5, the length of the χ-geodesic from x to y is

$$\sqrt{e^{2x_1} + e^{2y_1} - 2e^{x_1+y_1}\cos(y_2 - x_2)}$$

in either case. Applying Theorem 4 we obtain the corollary. □

Corollary 4. *Let* $f(u,v)=\frac{1}{2}\log u-\alpha-\beta v^{\gamma}$ *where* $\beta>0,\gamma>0$, *and*

$$\alpha=\frac{n}{2}\log(2\pi)-\frac{n}{2\gamma}\log\beta-\log\gamma-\log\Gamma\left(\frac{n}{2}\right)+\log\Gamma\left(\frac{n}{2\gamma}\right),$$

so that (13) *is satisfied. For any* $\delta,t\in R_+$ *let* $\psi(t)=e^{-\delta t}$, *where* $t\in R_+$. *Define* $\tilde{\chi}(t)=e^{2t}$. *Then, for some factor* $\phi\in R_+$ *depending on* n,β,γ,δ, *and vectors* $\tilde{z}^0,\tilde{z}^1\in R^2$ *computable from* $\Lambda_0,\Lambda_1,n,\beta,\gamma,\delta$, *we have*

$$d(\Lambda_0,\Lambda_1)=d_{\tilde{\chi}}(\tilde{z}^0,\tilde{z}^1).$$

Proof. A calculation shows that, for some $w\in R^3$, depending on n,β,γ,δ,

$$(a(t),b(t),c(t))=t^{(\delta-1)/2}w \quad \text{where} \quad t\in R_+.$$

From Corollary 1,

$$\bar{b}(t)=b(e^t)=w_2e^{t(\delta-1)/2} \quad \text{and} \quad \bar{c}(t)=w_3e^{t(\delta-1)/2}.$$

So, in the proof of Theorem 1,

$$\chi(t)=\phi_2e^{\phi_3t},$$

where $\phi_1=\sqrt{1+\frac{nw_2}{w_3}}$, $\phi_2=n^{-1}w_3$ and $\phi_3=(\delta-1)/(2\phi_1)$. Consider the affine transformation given by $\tilde{z}=(\phi_3/2)z+(\log(4\phi_2/\phi_3^2),0)/2$, where $z=(z_1,z_2)\in R^2$. Then $d_\chi(z^0,z^1)=d_{\tilde{\chi}}(\tilde{z}^0,\tilde{z}^1)$, and $\tilde{\chi}(\tilde{z}_1)(\dot{\tilde{z}}_1{}^2+\dot{\tilde{z}}_2{}^2)=\chi(z_1)(\dot{z}_1^2+\dot{z}_2^2)$. $\square$

6. Addendum

In Corollary 3 there is a simple formula for $d_\chi(x,y)$, but this is an exceptional case and closed form expressions are usually not available. The case treated in Corollary 3 is also exceptional in another sense, namely the (sectional) curvature $\kappa:R^2\to R$ is 0, as can be calculated directly. Alternatively, define a local isometry from $(R^2,<,>_\chi)$ onto a punctured cone $C'=\{(u,v,w):u^2+v^2=w^2,w>0\}$ in Euclidean 3-space R^3, by

$$(z_1,z_2)\mapsto\frac{e^{z_1}}{\sqrt{2}}(\cos\sqrt{2}z_2,\cos\sqrt{2}z_2,1).$$

Of course C' is isometric to an open subset of R^2 but incomplete. An alternative proof of Corollary 3 can be constructed based on these remarks.

Modulo affine transformations, the other exceptional cases that we know about, where closed form expressions are available for d_χ, are

- $\chi(t)=\frac{1}{\cosh^2 t}$, for which κ is identically 1, and lengths of χ-geodesics can be calculated by comparison with the geometry of the unit sphere S^2 embedded in R^3,
- $\chi(t)=\dfrac{1}{t^2}$, with κ identically -1, and $R_+\times R$ with the Riemannian metric $<,>_\chi$ is isometric to the Poincaré upper half-plane [5,14]. Of course, in this case, the

Riemannian metric $<,>_\chi$ is not defined over the whole of R^2. For $x, y \in R_+ \times R$ with $x_2 \neq y_2$, set

$$c = \frac{1}{2}\left(x_2 + y_2 + \frac{x_1^2 - y_1^2}{x_2 - y_2}\right).$$

Then $d_\chi(x, y)$ is $|\log(\frac{y_1}{x_1})|$, or

$$\left|\log\left(\frac{y_1(\sqrt{(x_2 - c)^2 + x_1^2} - |x_2 - c|)}{x_1(\sqrt{(y_2 - c)^2 + y_1^2} - |y_2 - c|)}\right)\right|,$$

according as $x_2 = y_2$ or not.

In these cases there are also comparisons to be made, using isometric immersions in R^3, between $<,>_\chi$ and the first fundamental forms of well-studied surfaces in R^3. This gives an alternative method of computing $d_\chi(x, y)$.

References

[1] S.I. Amari, O.E. Barndorff-Nielson, R.E. Kass, S.L. Lauritzen, C.R. Rao, Differential geometry in statistical inference Vol. 10, Institute of Statistics, Hayward, CL, 1987.
[2] C. Atkinson, A.F.S. Mitchell, Rao's distance measure, Sankhyā 4 (1981) 345–365.
[3] M. Berkane, K. Oden, P.M. Bentler, Geodesic estimation in elliptical distributions, J. Multivariate Anal. 63 (1997) 35–46.
[4] J. Burbera, C.R. Rao, Entropy, differential metric, distance and divergence measures in probability spaces: a unified approach, J. Multivariate Statist. 12 (1982) 575–596.
[5] A. Gray, Modern Differential Geometry of Curves and Surfaces with Mathematica, CRC Press LLC, Boca Raton, 1998.
[6] A.T. James, The variance information manifold and the functions on it, in: Multivariate Analysis, Vol. III, Academic Press, New York, 1973, pp. 157–169.
[7] T. Kato, Perturbation Theory for Linear Operators, Springer, New York, 1966.
[8] M. Lovric, M. Min-Oo, E.A. Ruh, Multivariate normal distributions parametrized as a Riemannian symmetric space, J. Multivariate Anal. 74 (2000) 36–48.
[9] C.A. Micchelli, On a measure of dissimilarity for normal probability densities, Ann. Numer. Math. 4 (1997) 461–478.
[10] J. Milnor, Morse Theory, Princeton University Press, Princeton, 1963.
[11] M.K. Murray, J.W. Rice, Differential Geometry and Statistics, Chapman & Hall, London, 1993.
[12] C.R. Rao, On the distance between two populations, Sankhyā 9 (1949) 246–248.
[13] L.T. Skovgard, A Riemannian geometry of the multivariate normal model, Research Report 81/3, Statistical Research Unit, Danish Medical Research Council, Danish Social Science Research Council, 1981.
[14] W.P. Thurston, S. Levy (Ed.), Three-Dimensional Geometry and Topology, Vol. 1, Princeton University Press, Princeton, NJ, 1997.