J. London Math. Soc. (2) 64 (2001) 501–512
DOI: 10.1112/S0024610701002332

ON SIMPLEX SHAPE SPACES

HUILING LE AND DENNIS BARDEN

Abstract

The right-invariant Riemannian metric on simplex shape spaces in fact makes them particular Riemannian symmetric spaces of non-compact type. In the paper, the general properties of such symmetric spaces are made explicit for simplex shape spaces. In particular, a global matrix coordinate representation is suggested, with respect to which several geometric features, important for shape analysis, have simple and easily computable expressions. As a typical application, it is shown how to locate the Fréchet means of a class of probability measures on the simplex shape spaces, a result analogous to that for Kendall's shape spaces.

1. *Introduction*

In [**1**] and [**2**], Bookstein proposed a method for the representation of the shapes of labelled triangles as points in the Poincaré half plane – a space of constant negative curvature, with the collinear triangles situated on the infinite horizon formed by the x-axis. This representation has a geometrically appealing property: the geodesic distance between any two triangle shapes is equal to a multiple of the logarithm of the strain ratio of the affine transformation mapping one triangle to the other. In [**13**], Small extended the Bookstein model by representing the shapes of labelled simplexes in $\mathbb{R}^n$ on a Riemannian manifold in such a way that the Riemannian distance between two simplex shapes is invariant under simultaneous transformation of the simplexes by a common affine transformation and is a function of the principal strains of the affine transformation mapping one simplex onto the other. These manifolds are quite distinct from those proposed by D. G. Kendall (cf. [**5**]) based upon procrustean arguments, and are designed mainly for applications, such as the biological ones, where it is reasonable to assume that there is some geometrical interdependence among the vertices which prevents the simplex they span being degenerate. In such cases, it would be more appropriate to use Bookstein's model than Kendall's model to measure the difference between shapes, since, in the former, the shapes of degenerate simplexes are infinitely far from each non-degenerate one.

In [**11**], using the theory of Riemannian submersions, Le and Small obtained an explicit formula for the Riemannian metric on simplex shape space and, hence, also for the Riemannian distance between simplex shapes. In the case of Kendall shape spaces, much more geometric information is known (cf. [**6**]), such as a formula for the geodesic between two given shapes and a formula for the full curvature tensor at each shape giving, in particular, bounds on the sectional curvature. All of this information has practical applications in the statistical analysis of Shape. For example, the geodesics have been used to model change of shape over time (cf. [**9**]); the knowledge of the curvature enables one to make a quantitative assessment of the validity of the common approximation for concentrated data of working in a

Received 21 July 1999; revised 28 September 2000.

2000 *Mathematics Subject Classification* 60D05.

tangent space to the shape space and it also plays an important role in the study of existence and location of Fréchet mean shapes (cf. [**8**]). In this paper, we take a different approach to that of [**11**] in order to obtain further explicit geometric information for simplex shape spaces and we also illustrate its practical application.

In the next section, we describe our model of simplex shape space. We explain how the requirement that the distance between simplex shapes is invariant under simultaneous affine transformation implies that this metric be invariant under a natural action of the special linear group on the shape space. If this invariant metric is induced by a Riemannian metric, this further implies that our simplex shape space is in fact an irreducible Riemannian globally symmetric space of non-compact type (cf. [**3**, p. 518]). This means that it has a number of geometric properties which are of significance in the study of simplex shapes. For example, the right-invariant Riemannian metric that we require is unique up to a homothety and, hence, the choice that we shall make must agree up to a global scalar multiple with those described rather differently in [**11**] and [**13**]. In fact, it agrees with that in [**11**] and is $\sqrt{n}/2$ times that in [**13**]. We describe this and other geometric properties of symmetric spaces that we shall use and also, since they are not always easy to locate explicitly in the literature, give readers some idea of how they arise.

In Section 3 we use this basic information to obtain a choice of matrix coordinates for the simplex shape space in terms of which the various geometric features have particularly simple expressions. In particular, we obtain explicit expressions for the induced distance function, the geodesics and the global isometric involutions. All of these formulae are elementary and easily susceptible to computation.

This lays the foundation for practical results analogous to those mentioned above for Kendall shape spaces and, in the final section of the paper, we give an example of such an application: namely, the proof of the existence and uniqueness of Fréchet means, as well as their identification, for appropriate probability measures on simplex shape spaces.

2. *Simplex shape space as a symmetric space*

Our model of simplex shape space starts from the observation that, since the vertices of a simplex are ordered, or equivalently labelled, a non-degenerate simplex in $\mathbb{R}^n$ can be represented by an $n \times (n+1)$ matrix whose jth column is the location of the jth vertex of the simplex. By moving the first vertex to the origin and then omitting it, the quotient space of *labelled* non-degenerate simplexes in $\mathbb{R}^n$ modulo translations is then identified with

$$\mathbf{GL}(n) = \{Y \mid Y \text{ is a } (n \times n)\text{-matrix such that } \det(Y) \neq 0\}.$$

The absolute value of the determinant of a matrix in $\mathbf{GL}(n)$ is, up to a constant multiple, equal to the volume of the corresponding simplex. Thus, by making the natural choice of the volume as a measure of the 'size' of the simplex, the quotient space of simplexes in $\mathbb{R}^n$ modulo translations, reflections and scaling may be identified with

$$\mathbf{SL}(n) = \{Y \in \mathbf{GL}(n) \mid \det(Y) = 1\}.$$

Thus, since the shape of a simplex is also invariant under rotations, the space of simplex shapes in $\mathbb{R}^n$ is homeomorphic with the quotient $\mathbf{M}(n) = \mathbf{SL}(n)/\mathbf{SO}(n)$ of $\mathbf{SL}(n)$ by the *left* action of $\mathbf{SO}(n)$. In the following, we shall write π for the quotient map from $\mathbf{SL}(n)$ to $\mathbf{M}(n)$ so that $\pi(Y)$ is the left coset $\mathbf{SO}(n)Y = \{Y' \in \mathbf{SL}(n) \mid Y' = TY$ for some $T \in \mathbf{SO}(n)\}$. This is the element of $\mathbf{M}(n)$ representing the shape of all simplexes that, after translation, reflection and re-scaling, are identified with $Y \in \mathbf{SL}(n)$.

Recall that the basic requirement in Bookstein's and Small's models is that the distance between the shapes of two simplexes should depend only on the affine transformation that takes one onto the other. Since shape is unaffected by translation, reflection and scalar multiplication, this means that for each T in $\mathbf{SL}(n)$, regarded as the linear part of an affine transformation that preserves volume and orientation, we require the distance $d(\pi(Y), \pi(TY))$ between the shapes $\pi(Y)$ and $\pi(TY)$ to be independent of the choice of Y, and so equal to the distance $d(\pi(I), \pi(T))$ between the shapes of two simplexes that, after standardising with respect to translation, reflection and scaling, are identified with I and T respectively. Then, varying Y will show that we need this metric to be invariant under the *right* action of $\mathbf{SL}(n)$ on $\mathbf{M}(n)$.

In correlating the results we claim with those in the literature, two technical points need to be observed. Both are consequences of the facts that we represent the shape of a simplex in $\mathbb{R}^n$ by a *left* coset of $\mathbf{SO}(n)$ in $\mathbf{SL}(n)$ and that we are seeking a *right*-invariant metric on $\mathbf{M}(n)$. Firstly, in order to have a group action of $\mathbf{SL}(n)$ on itself, for $T \in \mathbf{SL}(n)$, we should define the corresponding right multiplication $\tilde{\rho}_T$ by $\tilde{\rho}_T : Y \mapsto YT^{-1}$. Similarly, for $\mathbf{M}(n)$, we define $\rho_T : \pi(Y) \mapsto \pi(YT^{-1})$. Throughout this paper, the term *right invariance* will refer to the appropriate one of these right actions. Secondly, it will be convenient to identify the tangent space to $\mathbf{SL}(n)$ at the identity I with the Lie algebra of *right-invariant* vector fields with the Lie product corresponding to the Poisson bracket of vector fields. These are the opposites of the conventions usually adopted, however all the proofs remain valid *mutatis mutandis*.

If we further require that the metric on $\mathbf{M}(n)$, invariant under the right action of $\mathbf{SL}(n)$, be induced by a Riemannian metric, as was indeed the case in [**11**] and [**13**], then $\mathbf{M}(n)$ becomes a Riemannian symmetric space of non-compact type. Such spaces are classified (cf. [**3**] and [**14**]) and their main geometric properties are well understood. In the remainder of this section, we explain how two basic properties of symmetric spaces, the essentially unique right-invariant metric and the resulting geodesics, arise in our specific context.

The group $\mathbf{GL}(n)$ has a natural differential structure as an open subset of Euclidean space and $\mathbf{SL}(n)$ is a smooth submanifold of $\mathbf{GL}(n)$ for which the right translations $\tilde{\rho}_T$ are diffeomorphisms. If we give $\mathbf{M}(n)$ the quotient topology, then the induced right translations ρ_T will be homeomorphisms. However, $\mathbf{M}(n)$ is an example of a coset manifold, and general theory (cf. [**12**, 307ff]) tells us that it has a unique differential structure that makes the natural projection, or quotient map, $\pi : \mathbf{SL}(n) \to \mathbf{M}(n)$ a submersion: that is, at each point of $\mathbf{SL}(n)$, the derivative of π has rank equal to the dimension of $\mathbf{M}(n)$. In terms of this structure, the homeomorphisms ρ_T form a group of diffeomorphisms which acts transitively on $\mathbf{M}(n)$. With respect to any right-invariant metric, these diffeomorphisms will necessarily be isometries which will be important for our calculations and applications.

Any right-invariant vector field, in the Lie algebra of $\mathbf{SL}(n)$, is of course determined by its value in the tangent space at the identity I, its value at other points being obtained from that by right translation. The tangent space to $\mathbf{CL}(n)$ at I is the space of all $n \times n$ real-valued matrices, and the subspace tangent to $\mathbf{SL}(n)$, the kernel of the derivative of the determinant map there, is the subspace of matrices of trace zero. Thus, we may further identify the Lie algebra $\mathfrak{sl}(n)$ of $\mathbf{SL}(n)$ with this subspace, which we shall still denote by $\mathfrak{sl}(n)$, and, similarly, the Lie algebra of $\mathbf{SO}(n)$ may be identified with the subalgebra

$$\mathfrak{so}(n) = \{X \in \mathfrak{sl}(n) \mid X^t = -X\}.$$

In terms of these matrices, the Lie product is given by $[X_1, X_2] = X_1X_2 - X_2X_1$.

For any element T of $\mathbf{SL}(n)$, the conjugation isomorphism $C_T : Y \mapsto TYT^{-1}$ determines a linear isomorphism $\mathrm{Ad}(T) = d(C_T)_I$ of the tangent space $\mathfrak{sl}(n)$ at the identity. Since, in terms of the standard coordinates on $\mathbf{GL}(n)$, both left and right multiplication are linear, it follows that $\mathrm{Ad}(T)$ is the linear map $X \mapsto TXT^{-1}$ of $\mathfrak{sl}(n)$ into itself. Using the trivial decomposition $X = \frac{1}{2}\{(X - X^t) + (X + X^t)\}$ for any $X \in \mathfrak{sl}(n)$, we have

$$\mathfrak{sl}(n) = \mathfrak{so}(n) \oplus \mathfrak{p}(n),$$

where $\mathfrak{p}(n)$ is the subspace $\{X \in \mathfrak{sl}(n) \mid X^t = X\}$, which is clearly $\mathrm{Ad}(\mathbf{SO}(n))$-invariant. This means that $\mathbf{M}(n)$ will be what is termed a *reductive* coset manifold. When equipped with an $\mathbf{SL}(n)$-invariant metric, it becomes a reductive homogeneous space and the possible $\mathbf{SL}(n)$-invariant metrics may be determined as in the following lemma, which is a special case of a result which holds for all reductive coset manifolds (cf. [**12**]).

LEMMA 1. *There is a bijection between* $\mathrm{Ad}(\mathbf{SO}(n))$*-invariant scalar products on* $\mathfrak{p}(n)$ *and* $\mathbf{SL}(n)$*-invariant Riemannian metrics on* $\mathbf{M}(n)$*, determined by the requirement that* $d\pi_I : \mathfrak{p}(n) \to \mathscr{T}_{\pi(I)}(\mathbf{M}(n))$ *be an isometry.*

One particular choice of $\mathrm{Ad}(\mathbf{SO}(n))$-invariant inner product on $\mathfrak{p}(n)$ is $\langle X_1, X_2\rangle = \mathrm{tr}(X_1X_2^t)$ for which $\mathrm{Ad}(\mathbf{SO}(n))$-invariance follows from the fact that, for R in $\mathbf{SO}(n)$, $\mathrm{Ad}(R)$ acts by conjugation. This inner product is in fact a constant multiple of the restriction to $\mathfrak{p}(n)$ of the Killing form on $\mathfrak{sl}(n)$ and is essentially unique. This uniqueness arises from the fact that the decomposition of $\mathfrak{sl}(n)$ as $\mathfrak{so}(n) \oplus \mathfrak{p}(n)$, together with an $\mathrm{Ad}(\mathbf{SO}(n))$-invariant inner product on $\mathfrak{p}(n)$ and the involution of $\mathfrak{sl}(n)$ that is $+\mathrm{id}$ on $\mathfrak{so}(n)$ and $-\mathrm{id}$ on $\mathfrak{p}(n)$, satisfies a couple of other technical conditions that make it an irreducible orthogonal involutive Lie algebra. The classification of such algebras implicitly involves the following lemma. See, for example, [**14**], where the result we require falls under class 3 of Theorem 8.2.9. However, given our explicit knowledge of $\mathrm{Ad}(\mathbf{SO}(n))$-action on $\mathfrak{p}(n)$, it is possible, by choosing an obvious basis of $\mathfrak{p}(n)$ and generators of $\mathbf{SO}(n)$, to give a direct elementary proof for our case.

LEMMA 2. *The only* $\mathrm{Ad}(\mathbf{SO}(n))$*-invariant inner products on* $\mathfrak{p}(n)$ *are positive scalar multiples of the restriction of the Killing form.*

From Lemma 2, together with Lemma 1, we see that the $\mathbf{SL}(n)$-invariant Riemannian metric on $\mathbf{M}(n)$ is unique up to homothety, that is, multiplication by a constant. The choice we make is that corresponding to the multiple $1/2n$ of the Killing form. We now use it to characterise the geodesics on simplex shape space and, in the next section, we shall use it to obtain a simple formula for the distance between two shapes as well as for the geodesics themselves.

In order to describe the geodesics, we consider the Lie exponential map from $\mathfrak{sl}(n)$ to $\mathbf{SL}(n)$ given, in terms of matrices, by $\exp : X \mapsto \sum_{n=0}^{\infty} X^n/n!$. This series converges and so $\exp(X)$ is defined for all X. Since $\exp(YXY^{-1}) = Y\exp(X)Y^{-1}$, conjugation of X to the Jordan normal form shows that $\det(\exp(X)) = \exp(\operatorname{tr}(X))$. Thus, $\exp(X)$ is indeed in $\mathbf{SL}(n)$ for all X in $\mathfrak{sl}(n)$. Moreover, when X_1 and X_2 commute, $\exp(X_1 + X_2) = \exp(X_1)\exp(X_2)$ so that, for real s and t, $\exp(sX)\exp(tX) = \exp((s+t)X)$. This means that $\alpha_X : t \mapsto \exp(tX)$ is a 1-parameter subgroup of $\mathbf{SL}(n)$. It is also the integral curve starting at I of the right-invariant vector field X. On the other hand, since $\mathbf{M}(n)$ is a Riemannian manifold, for each $v \in \mathscr{T}_{\pi(Y)}(\mathbf{M}(n))$, there is a geodesic γ_v, a curve in $\mathbf{M}(n)$ starting at $\pi(Y)$ determined by the requirement that $\gamma_v'(0) = v$ and $D_{v(t)}v(t) = 0$, where D is the Levi–Civitá connection determined by the Riemannian metric and $v(t) = \gamma_v'(t)$ is the tangent vector field along γ_v. We then have the following, which is, again, a special case of a more general result. See, for example, [**12**], Proposition 3.1 on p. 317.

LEMMA 3. *For $X \in \mathfrak{p}(n)$, the geodesic $\gamma_{d\pi_I(X)}$, with respect to the above $\mathbf{SL}(n)$-invariant metric on $\mathbf{M}(n)$, is the image $\pi \circ \alpha_X$ of the 1-parameter subgroup of $\mathbf{SL}(n)$ generated by X.*

Since $\exp(tX)$ is defined for all t, this lemma confirms the completeness of $\mathbf{M}(n)$, another property common to all Riemannian symmetric spaces.

3. *Some explicit geometry of simplex shape spaces*

The results of the previous section enable us to give an elementary account in our context of the features of symmetric spaces which interest us. Firstly, we describe a convenient set of matrix coordinates for simplex shape space, in terms of which we then give explicit formulae for the distances and geodesics between shapes and for the global isometric involutions which characterise it as a symmetric space. We close with some comments on the sectional curvature of our spaces.

Our coordinates will lie in the submanifold

$$\mathbf{P}(n) = \{P \in \mathbf{SL}(n) \mid P^t = P \text{ and } P \text{ is positive definite}\},$$

a relatively open subset of a linear subspace, of $\mathbf{SL}(n)$. From the unique polar factorisation of $Y \in \mathbf{SL}(n)$ as $Y = R_Y P_Y$, where $R_Y \in \mathbf{SO}(n)$ and $P_Y \in \mathbf{P}(n)$, it follows that P_Y is the unique element of $\mathbf{P}(n)$ in the $\mathbf{SO}(n)$-coset $\pi(Y)$. Thus, the restriction of the quotient map π to $\mathbf{P}(n)$, mapping P_Y to $\pi(P_Y) = \pi(Y)$, is bijective and is easily checked to be a diffeomorphism.

This diffeomorphism $\bar{\pi}$ of $\mathbf{P}(n)$ onto $\mathbf{M}(n)$ has particularly convenient properties, which we now elucidate. Note first that exp maps the subspace $\mathfrak{p}(n)$ into the submanifold $\mathbf{P}(n)$. In fact, if P is in $\mathbf{P}(n)$, then all the eigenvalues of P are positive, so that P may be written as $P = T\Lambda T^t$ with T orthogonal and Λ a diagonal matrix with positive entries and $\det(\Lambda) = 1$. Thus, $\Lambda = \exp(X)$ with $X \in \mathfrak{p}(n)$ and so $P = \exp(TXT^t)$ with TXT^t also in $\mathfrak{p}(n)$. This shows that the restricted map $\overline{\exp} : \mathfrak{p}(n) \longrightarrow \mathbf{P}(n)$ is surjective and it may similarly be seen to be injective and so a diffeomorphism between the two spaces. Thus, the composite $\bar{\pi} \circ \overline{\exp}$ is a diffeomorphism of $\mathfrak{p}(n)$ onto $\mathbf{M}(n)$. However, the Riemannian exponential Exp of $\mathbf{M}(n)$ at $\pi(I)$ is given by

$$\operatorname{Exp}(td\pi_I(X)) = \gamma_{td\pi_I(X)}(1) = \gamma_{d\pi_I(X)}(t),$$

so Lemma 3 says that $\bar{\pi} \circ \overline{\exp} = \operatorname{Exp} \circ d\bar{\pi}_I$. Since, by choice, in order to apply Lemma 1, $d\pi_I$ is an isometry of $\mathfrak{p}(n)$ with $\mathscr{T}_{\pi(I)}\mathbf{M}((n))$, it follows that Exp is also a diffeomorphism between $\mathscr{T}_{\pi(I)}(\mathbf{M}(n))$ and $\mathbf{M}(n)$ and, by homogeneity, a similar property holds at all points of $\mathbf{M}(n)$. This fact will be used in Section 4 when applying the result of Lemma 4 to simplex shape spaces.

In view of the above diffeomorphism $\bar{\pi}$, we shall, when it is convenient, identify $\mathbf{M}(n)$ with $\mathbf{P}(n)$ and so denote $\pi(Y)$ just by P_Y or P. This gives another system of local coordinates on the simplex shape space, which is different from those arising from the upper-triangular representation given in [**13**] or from the representation based on the singular-values decomposition used in [**11**]. Our present coordinates have the advantage that by Lemma 3, when $\mathbf{M}(n)$ has the $\mathbf{SL}(n)$-invariant metric induced by the inner product $\langle X_1, X_2\rangle = \mathrm{tr}(X_1 X_2^t)$ on $\mathfrak{p}(n)$, the Riemannian geodesics on $\mathbf{M}(n)$ starting at $\pi(I)$ correspond in $\mathbf{P}(n)$ to the 1-parameter groups

$$\gamma_X(t) = \exp(tX), \qquad t \in \mathbb{R}, \tag{1}$$

for $X \in \mathfrak{p}(n)$. Since $\overline{\exp} : \mathfrak{p}(n) \to \mathbf{P}(n)$ is bijective, this shows in particular that the Riemannian distance from $P = \overline{\exp}(X)$ to I, corresponding to that in $\mathbf{M}(n)$, is given by $d(I,P)^2 = \mathrm{tr}(XX^t)$. However, if $\lambda_1, \ldots, \lambda_n$ are the eigenvalues of P, then the eigenvalues of X are $\log\lambda_1, \ldots, \log\lambda_n$, and so we have

$$d(I,P)^2 = \mathrm{tr}(XX^t) = \sum_{i=l}^{n} (\log\lambda_i)^2.$$

Note that $P \in \mathbf{P}(n)$ implies that all the λ_i are positive and $\prod_{i=1}^n \lambda_i = 1$. The right invariance of the metric on $\mathbf{M}(n)$, together with the fact that the eigenvalues of the positive definite symmetric factor in the polar decomposition of $Y \in \mathbf{SL}(n)$ are the positive square roots of the eigenvalues of YY^t, gives us the following result.

PROPOSITION 1. *Let the general points P_1 and P_2 of $\mathbf{P}(n)$ represent the simplex shapes $\bar{\pi}(P_1)$ and $\bar{\pi}(P_2)$ in $\mathbf{M}(n)$. Then, the squared distance $d(\bar{\pi}(P_1), \bar{\pi}(P_2))^2$ between these two shapes, induced by the right-invariant Riemannian metric on $\mathbf{M}(n)$, is $\sum_{i=1}^n (\log\lambda_i)^2$, where the λ_i are the positive square roots of the eigenvalues of $P_1P_2^{-1}(P_1P_2^{-1})^t$.*

Now, to correlate this distance function with that given in [**11**], for any $\tilde{Y} \in \mathbf{GL}(n)$, we define

$$Y = \mathrm{diag}\{1, \ldots, 1, \mathrm{sign}(\det(\tilde{Y}))\}\frac{\tilde{Y}}{(|\det(\tilde{Y}|)^{1/n}}.$$

so that Y is in $\mathbf{SL}(n)$. If $\tilde{\gamma}_1 \geqslant \ldots \geqslant \tilde{\gamma}_n > 0$ and $\gamma_1 \geqslant \ldots \geqslant \gamma_n > 0$ are the singular-values of $\tilde{Y}$ and Y respectively, then we have

$$\gamma_I = \frac{\tilde{\gamma}_i}{\left(\prod_{j=1}^n \tilde{\gamma}_j\right)^{1/n}},$$

and the point P_Y in $\mathbf{P}(n)$ that corresponds to $\pi(Y) \in \mathbf{M}(n)$ has eigenvalues γ_i, $i = 1, \ldots, n$. Thus, the squared Riemannian distance from the simplex shape $\pi(Y)$ to the shape $\pi(I)$ is

$$d(\pi(I), \pi(Y))^2 = \sum_{i=1}^n (\log\gamma_i)^2 = \sum_{i=1}^n \left\{ \log\tilde{\gamma}_i - \frac{1}{n}\sum_{j=1}^n \log\tilde{\gamma}_j \right\}^2,$$

which is equivalent to the formula given in [**11**].

We have already given the simple formula (1) for geodesics on $\mathbf{M}(n)$ starting at $\pi(I)$ in terms of their coordinates in $\mathbf{P}(n)$. For geodesics starting from other points, a little care is needed. Firstly, we note that, if we define an inner product on $\mathfrak{sl}(n)$ by the same formula, $\langle X_1, X_2\rangle = \operatorname{tr}(X_1 X_2^t)$, that we are using on $\mathfrak{p}(n)$ and then, regarding $\mathfrak{sl}(n)$ as the tangent space to $\mathbf{SL}(n)$ at the identity, extend this to a right-invariant Riemannian metric on $\mathbf{SL}(n)$ using the derivatives of right translations, then the projection π of $\mathbf{SL}(n)$ onto $\mathbf{M}(n)$ will be a Riemannian submersion with respect to the $\mathbf{SL}(n)$-invariant metric on $\mathbf{M}(n)$. This follows from the fact that $\mathfrak{so}(n)$ and $\mathfrak{p}(n)$ are orthogonal subspaces of $\mathfrak{sl}(n)$ so that, as $\mathfrak{so}(n)$ is the tangent space to the fibre $\mathbf{SO}(n)$, $\mathfrak{p}(n)$ is the horizontal subspace at $I \in \mathbf{SL}(n)$. Then, since right translations preserve fibres and are isometric, $\mathscr{H}_Y = d\tilde{\rho}_{Y^{-1}}(\mathfrak{p}(n))$ is the horizontal subspace at $Y \in \mathbf{SL}(n)$. However, the projection $d\pi_Y : \mathscr{H}_Y \to \mathscr{T}_{\pi(Y)}(M(n))$ is the composite $d\rho_{Y^{-1}} \circ d\pi_1 \circ d\tilde{\rho}_Y$ of maps all of which are isometries by the definitions.

Given points P_1 and P_2 in $\mathbf{P}(n)$, representing the simplex shapes $\bar{\pi}(P_1)$ and $\bar{\pi}(P_2)$ in $\mathbf{M}(n)$, let $P_2P_1^{-1} = RP_{P_2P_1^{-1}}$, where $R \in \mathbf{SO}(n)$ and $P_{P_2P_1^{-1}} \in P(n)$. If $X \in \mathscr{T}_I(P(n)) = \mathfrak{p}(n)$ is such that $P_{P_2P_1^{-1}} = \exp(X)$, then $P(t) = \exp(tX)$ is a, necessary horizontal, geodesic in $\mathbf{P}(n)$ from I to $P_{P_2P_1^{-1}}$ and so

$$\tilde{\rho}_{P_1^{-1}}(P(t)) = P(t)P_1 = \exp(tX)P_1$$

is a horizontal geodesic in $\mathbf{SL}(n)$ from P_1 to R^tP_2. Thus, $\pi(\tilde{\rho}_{P_1^{-1}}(P(t)))$ is the geodesic in $\mathbf{M}(n)$ from $\bar{\pi}(P_1)$ to $\bar{\pi}(P_2)$. Since

$$\frac{d}{dt}(\tilde{\rho}_{P_1^{-1}}(P(t)))|_{t=0} = XP_1,$$

we have

$$\tilde{\rho}_{P_1^{-1}}(P(t)) = \exp(tX)P_1 = \widetilde{\operatorname{Exp}}_{P_1}(tXP_1),$$

where $\widetilde{\operatorname{Exp}}_{P_1}$ is the Riemannian exponential map on $\mathbf{SL}(n)$ at P_1. Hence, the required geodesic $\pi(\tilde{\rho}_{P_1^{-1}}(P(t))$ is $\operatorname{Exp}_{\bar{\pi}(P_1)}(td\pi(XP_1))$, where, as before, $\operatorname{Exp}_{\bar{\pi}(P_1)}$ is the Riemannian exponential map on $\mathbf{M}(n)$ at $\bar{\pi}(P_1)$. Thus, we have established the following result.

Proposition 2. *Writing the general tangent vector V at $\bar{\pi}(P_1) \in \mathbf{M}(n)$, where $P_1 \in \mathbf{P}(n)$, as the image $V = d\pi(XP_1)$ of the horizontal vector XP_1 at P_1 in $\mathbf{SL}(n)$, where $X \in \mathfrak{p}(n)$, the geodesic $\operatorname{Exp}_{\bar{\pi}(P_1)}(tV)$ in $\mathbf{M}(n)$ is given by $\bar{\pi}(P(t))$, where $\exp(tX)P_1 = R(t)P(t)$ with $R(t) \in \mathbf{SO}(n)$ and $P(t) \in \mathbf{P}(n)$. This geodesic joins $\bar{\pi}(P_1)$ and $\bar{\pi}(P_2)$ when X is chosen such that $P_2P_1^{-1} = R\exp(X)$ with R in $\mathbf{SO}(n)$.*

The defining property of a Riemannian symmetric space is the existence at each point $\pi(Y)$ of an isometric involution fixing $\pi(Y)$ whose derivation is $-\mathrm{id}$ on the tangent space at $\pi(Y)$ (cf. [**4**] and [**12**]). These involutions will play an important role in Theorem 1 below. They are easily identified in our context: if $\sigma = \overline{\exp} \circ \imath \circ \overline{\exp}^{-1}$ is the 'polar' diffeomorphism that the involution $\imath = -\mathrm{id}$ of $\mathfrak{p}(n)$ induces on $\mathbf{P}(n)$, then the involution $\zeta = \bar{\pi} \circ \sigma \circ \bar{\pi}^{-1}$ that σ induces on $\mathbf{M}(n)$ via $\bar{\pi}$ is the required isometric involution at $\pi(I)$. That $d\zeta_{\pi(I)} = -\mathrm{id}$ follows immediately from the fact that $d\exp(0) = \mathrm{id}$ and $d\bar{\pi}_I$ is an isomorphism. Writing $\mathbf{P} = \exp(X)$ for X in $\mathfrak{p}(n)$, we see that $\sigma(P) = P^{-1} = (P^t)^{-1}$, since $P \in \mathbf{P}(n)$, so that σ is the restriction of the involutive automorphism $\tilde{\sigma} : Y \mapsto (Y^{-1})^t$ of $\mathbf{SL}(n)$. From the polar decomposition

$Y = R_Y P_Y$ of Y, we see that

$$\pi \circ \tilde{\sigma}(Y) = \pi(R_Y (P_Y)^{-1}) = \bar{\pi} \circ \sigma(P_Y) = \zeta \circ \pi(Y).$$

We then find that, for all $P \in \mathbf{P}(n)$, $\rho_{\bar{\pi}(P^{-1})} \circ \zeta = \zeta \circ \rho_{\bar{\pi}(P)}$ on $\mathbf{M}(n)$. This result follows from the bijectivity of $\bar{\pi}$ and the fact that, for all P' in $\mathbf{P}(n)$,

$$\begin{aligned} \zeta \circ \rho_{\bar{\pi}(P)}(\bar{\pi}(P')) &= \zeta \circ \pi(P'P^{-1}) = \pi \circ \tilde{\sigma}(P'P^{-1}) \\ &= \pi((P')^{-1}P) = \rho_{\bar{\pi}(P^{-1})} \circ \bar{\pi}((P')^{-1}) = \rho_{\bar{\pi}(P^{-1})} \circ \zeta(\bar{\pi}(P')). \end{aligned}$$

Thus, for any $v \in \mathscr{T}_{\bar{\pi}(P)}(\mathbf{M}(n))$, writing $v_0 = d\rho_{\bar{\pi}(P)}(v) \in \mathscr{T}_{\pi(I)}(\mathbf{M}(n))$, we get

$$\begin{aligned} \langle d\zeta(v), d\zeta(v) \rangle &= \langle d\zeta \circ d\rho_{\bar{\pi}(P^{-1})}(v_0), d\zeta \circ d\rho_{\bar{\pi}(P^{-1})}(v_0) \rangle \\ &= \langle d\rho_{\bar{\pi}(P)} \circ d\zeta(v_0), d\rho_{\bar{\pi}(P)} \circ d\zeta(v_0) \rangle \\ &= \langle d\zeta(v_0), d\zeta(v_0) \rangle = \langle -v_0, -v_0 \rangle = \langle v, v \rangle. \end{aligned}$$

Thus, ζ is the induced global isometric involution of $\mathbf{M}(n)$ fixing $\pi(I)$.

In terms of our chosen coordinates, ζ is just represented by the polar diffeomorphism σ of $\mathbf{P}(n)$. As right translations are isometries and the composite $\rho_{\bar{\pi}(P_0^{-1})} \circ \zeta \circ \rho_{\bar{\pi}(P_0)}$ has derivative at $\bar{\pi}(P_0)$ equal to $-\mathrm{id}$, it must be the global isometric involution at $\bar{\pi}(P_0)$. Since the coordinate representation of the right translation $\rho_{\bar{\pi}(P_0)}$ is $P_1 \mapsto P_{P_1 P_0^{-1}}$, we have the following result.

PROPOSITION 3. *The global isometric involution* $\mathbf{M}(n)$ *at the shape* $\bar{\pi}(P_0)$ *has coordinate representation*

$$P_1 \mapsto P_{\left(P_{P_1 P_0^{-1}}\right)^{-1} P_0}, \tag{2}$$

where P_Y *is the positive definite symmetric factor in the polar decomposition* $R_Y P_Y$ *of* Y.

Finally in this section we briefly mention the sectional curvature of simplex shape spaces, referring to [**3**], for example, for details. Since it is a symmetric space of non-compact type of rank $n-1$, we know that $\mathbf{M}(n)$ has non-positive sectional curvature with a totally geodesic flat submanifold of dimension $n-1$ through each point. There is also, through each point, a totally geodesic submanifold of maximally negative curvature. When calculated with respect to the metric induced by the Killing form, this minimal value is $-(2-3/n)$. Since the metric we are using is induced by $1/2n$ times the Killing form, it follows that, for our choice, the minimal curvature is $6-4n$. The curvature will influence the range over which it is reasonable to make linear approximations, projecting for example onto the tangent space at a convenient point. However, for any desired level of accuracy, the different metrics will give the same result since any two given shape points will be deemed to be $\sqrt{2n}$ times further apart with respect to the 'Killing metric' than with respect to ours; the domains on which the approximation is valid will have a different measure, but will contain the same shape points.

4. *Fréchet means of certain probability measures on simplex shape spaces*

We now use the geometric properties of the simplex shape space consequent on its structure as a symmetric space, when it is equipped with its unique $\mathbf{SL}(n)$-invariant

Riemannian metric, to study a probabilistic issue that arises in the statistical analysis of shape.

Recall that a Fréchet mean, a generalisation of the mean or the expectation of a probability distribution on a Euclidean space to that of a probability measure μ on a general metric space ($\mathbf{M}$, dist), is defined to be *any point* that achieves the *global* minimum of the function

$$F(p) = \tfrac{1}{2}\int_{\mathbf{M}} \operatorname{dist}(p,q)^2 d\mu(q).$$

Fréchet means of probability measures on shape spaces have been used in studying the shape of the means, although such Fréchet means are in general not unique (cf. [**6**] and [**7**]). Nevertheless, when it is equipped with its $\mathbf{SL}(n)$-invariant Riemannian structure, $\mathbf{M}(n)$ has properties from which it will follow that any probability measure on it has *unique* Fréchet mean with respect to this Riemannian metric and that there is a simple criterion to identify this mean. The properties of $\mathbf{M}(n)$ that we require are its non-positive curvature and the fact that the Riemannian exponential map at each point is a diffeomorphism of the tangent space there onto $\mathbf{M}(n)$. A point with the latter property is called a *pole*. As for $\mathbf{M}(n)$, though not necessarily by homogeneity, it suffices that one point of a Riemannian manifold be a pole for all its points to be poles. The following, then, are the implications for Fréchet means of these two properties of a manifold.

LEMMA 4. *Let $\mathbf{M}$ be a Riemannian manifold of non-positive sectional curvature for which every point is a pole. Then, $p \in \mathbf{M}$ is a Fréchet mean, with respect to the Riemannian distance d on $\mathbf{M}$, of a given probability measure μ on $\mathbf{M}$ if and only if $0 \in \mathscr{T}_p(\mathbf{M})$ is a Fréchet mean of the induced probability measure $\mu \circ \operatorname{Exp}_p$ on $\mathscr{T}_p(\mathbf{M})$ with respect to the distance $\tilde{d}$ on $\mathscr{T}_p(\mathbf{M})$ determined by the Riemannian metric $\mathfrak{g}_p$ at p. Furthermore, such Fréchet means are unique.*

Proof. A Fréchet mean of the induced probability measure on $\mathscr{T}_p(\mathbf{M})$ with respect to $\tilde{d}$ is a global minimum of the function

$$\tilde{F}_p(X) = \tfrac{1}{2}\int_{\mathscr{T}_p(\mathbf{M})} \|X - X'\|_p^2 d\mu(\operatorname{Exp}_p(X')), \quad X \in \mathscr{T}_p(\mathbf{M}),$$

where $\|\cdot\|_p$ is the norm determined by the inner product $\mathfrak{g}_p$.

Firstly, since Exp_p is a diffeomorphism between $\mathscr{T}_p(\mathbf{M})$ and $\mathbf{M}$ and, since $\tilde{d}(0, X) = d(p, \operatorname{Exp}_p(X))$ for all X in $\mathscr{T}_p(\mathbf{M})$, we have

$$\tilde{F}_p(0) = F(p) \tag{3}$$

and, in particular, one is defined if and only if the other is. Secondly, it follows from Gauss' lemma together with the Rauch comparison theorem (cf. [**4**]) that, since $\mathbf{M}$ has non-positive sectional curvature, Exp_p is distance-increasing: $d(\operatorname{Exp}_p(X_1), \operatorname{Exp}_p(X_2)) \geqslant \tilde{d}(X_1, X_2)$ for all $X_1, X_2 \in \mathscr{T}_p(\mathbf{M})$. This implies that

$$\tilde{F}_p(X) \leqslant F(\operatorname{Exp}_p(X)), \quad \forall\, X \in \mathscr{T}_p(\mathbf{M}). \tag{4}$$

Thus, it follows from (3) and (4) that, if 0 is a Fréchet mean of $\mu \circ \operatorname{Exp}_p$ on $\mathscr{T}_p(\mathbf{M})$ with respect to the distance $\tilde{d}$, then p is a Fréchet mean of μ on $\mathbf{M}$ with respect to the Riemannian distance d.

On the other hand, the global minimum of $\tilde{F}_p$ can only occur at a critical point and, since $\int_{\mathscr{T}_p(\mathbf{M})} d\mu(\mathrm{Exp}_p(X)) = 1$, we see that $\tilde{F}'_p(\hat{X}) = 0$ if and only if

$$\hat{X} = \int_{\mathscr{T}_p(\mathbf{M})} X d\mu(\mathrm{Exp}_p(X)). \tag{5}$$

We also find that $\tilde{F}''_p(X) = \mathfrak{g}_p$ so that this unique critical point is a local, and hence global, minimum. However, since $\mathrm{grad}_p(d(p,q)^2) = -2\mathrm{Exp}_p^{-1}(q)$, we have

$$\mathrm{grad}F(p) = -\int_{\mathbf{M}} \mathrm{Exp}_p^{-1}(q) d\mu(q) = -\int_{\mathscr{T}_p(\mathbf{M})} X d\mu(\mathrm{Exp}_p(X)) = -\hat{X}.$$

Thus, when p is a critical point, in particular a global minimum, of F, then the unique global minimum of $\tilde{F}_p$ occurs at the origin:

$$\tilde{F}_p(0) < \tilde{F}_p(X), \qquad \forall\ X \in \mathscr{T}_p(\mathbf{M})\backslash\{0\}. \tag{6}$$

Then, (3), (4) and (6) imply that p is the unique global minimum of F. $\square$

It is clear from (5) that $\hat{X}$ is the Fréchet mean of $\mu \circ \mathrm{Exp}_p$ on $\mathscr{T}_p(\mathbf{M})$ with respect to the distance $\tilde{d}$ if and only if $\hat{X}$ is the Euclidean mean of $\mu \circ \mathrm{Exp}_p$ on $\mathscr{T}_p(\mathbf{M})$. Note also that this lemma gives us a context in which Fréchet means are unique and, together with our proof of it, two further sufficient conditions for such a Fréchet mean to exist: it will exist if, for any point p, the corresponding $\tilde{F}_p$ has its unique global minimum at the origin or if, on any domain on which the integral defining it converges, F has any critical point p and, in each case, that point p will be the unique Fréchet mean. However, since it is usually difficult to express the Riemannian exponential map explicitly, their identification still remains a hard problem in general. On the other hand, in the case of simplex shape space, our detailed knowledge of the geometry enables us to make further progress in that context. Identifying $\mathbf{M}(n)$ with $\mathbf{P}(n)$ and $\mathscr{T}_P(\mathbf{P}(n))$ with the horizontal subspace $\mathscr{H}_P(\mathbf{SL}(n))$, the function $\tilde{F}_P$ defined in the proof of Lemma 4 can be expressed as

$$\begin{aligned}\tilde{F}_p(XP) &= \tfrac{1}{2}\int_{\mathscr{H}_P} \|XP - X'P\|_P^2 d\mu \circ \widetilde{\mathrm{Exp}}_P(X'P) \\ &= \tfrac{1}{2}\int_{\mathscr{H}_P} \|XP - X'P\|_P^2 d\mu(\exp(X')P) \\ &= \tfrac{1}{2}\int_{\mathfrak{p}(n)} \|X - X'\|_I^2 d\mu(\exp(X')P).\end{aligned}$$

Let $\hat{X}_p$ be the Euclidean mean of $\tilde{\mu}_p$ on $\mathfrak{p}(n)$ defined by

$$d\tilde{\mu}_P(X') = d\mu(\exp(X')P),$$

which is clearly computable once μ is given. Then, $\hat{X}_P P$ is the Euclidean mean of $\mu \circ \widetilde{\mathrm{Exp}}_P$ on $\mathscr{H}_P$. Thus, the criterion of Lemma 4 for $\bar{\pi}(P)$ to be a Fréchet mean of μ is effectively calculatable, at least for simple μ.

We can go further than this to obtain an implementable algorithm for locating the Fréchet mean of μ, given that it exists. Suppose that $\bar{\pi}(P)$ is not the Fréchet mean of μ, so that $\hat{X}_p \neq 0$. Then,

$$G_{\bar{\pi}(P)}(t) = F(\mathrm{Exp}_{\bar{\pi}(P)}(td\pi(\hat{X}_P P)))$$

is a convex function, which is again directly computable using Proposition 2. From the proof of Lemma 4, we see that $G_{\bar{\pi}(p)}$ is decreasing at zero. Thus, there is a unique t_0 in (0, 1] such that $G_{\bar{\pi}(P)}(t_0) \leqslant G_{\bar{\pi}(P)}(t)$ for all t in [0, 1]. In particular,

$$F(\bar{\pi}(P)) = G_{\bar{\pi}(P)}(0) > G_{\bar{\pi}(P)}(t_0) = F(\mathrm{Exp}_{\bar{\pi}(P)}(t_0 d\pi(\hat{X}_P P))).$$

Repeating the calculation at $\bar{\pi}(P') = \mathrm{Exp}_{\bar{\pi}(P)}(t_0 d\pi(\hat{X}_p P))$, we obtain a sequence of points at which F takes a strictly decreasing sequence of values. Since these values are bounded below by the assumed global minimum value of F, they must converge and the above argument implies that they can only do so at a point $\bar{\pi}(\hat{P})$ such that $\hat{X}_{\hat{P}} = 0$, that is, by Lemma 4, they can only do so at the Fréchet mean.

In certain cases, making use of the global isometric involutions of $\mathbf{M}(n)$, we can also locate the Fréchet mean immediately without any need for such an approximating algorithm, as in the following result which is analogous to one obtained for Kendall shape spaces (cf. [**7**]) and is a generalisation of a result in [**10**].

THEOREM 1. *Suppose that* $\mathbf{M}(n)$ *is equipped with the unique* $\mathbf{SL}(n)$*-invariant Riemannian metric and that the Radon–Nikodým derivative* f*, with respect to the volume element* v*, of a given probability measure* μ *on the simplex shape space is a function of distance to a fixed point* $\pi(Y_0)$*. If any Fréchet mean, with respect to the Riemannian distance, of* μ *exists, then it is* $\pi(Y_0)$ *and is unique.*

Proof. By Lemma 4, we know the Fréchet mean of μ is unique. Suppose that $\pi(Y_1) \neq \pi(Y_0)$ is the Fréchet mean of μ. Then, if ζ is the global isometric involution of $\mathbf{M}(n)$ centred on $\pi(Y_0)$ and if $\zeta(\pi(Y_1)) = \pi(Y_2)$, we have $\pi(Y_1) \neq \pi(Y_2)$, again since $\mathbf{M}(n)$ is complete and has non-negative curvature. Then, since $\zeta(\pi(Y_0)) = \pi(Y_0)$,

$$\begin{aligned} F(\pi(Y_2)) &= \int_{\mathbf{M}(n)} d(\pi(Y_2), \pi(Y))^2 f(d(\pi(Y_0), \pi(Y))) dv(\pi(Y)) \\ &= \int_{\mathbf{M}(n)} d(\pi(Y_1), \zeta^{-1}(\pi(Y)))^2 f(d(\pi(Y_0), \zeta^{-1}(\pi(Y)))) dv(\zeta^{-1}(\pi(Y))) \\ &= \int_{\mathbf{M}(n)} d(\pi(Y_1), \pi(Y))^2 f(d(\pi(Y_0), \pi(Y))) dv(\pi(Y)) = F(\pi(Y_1)). \end{aligned}$$

This contradicts the fact that $F(\pi(Y_2)) > F(\pi(Y_1))$. □

Note that this proof only involves the single involution ζ and so we actually only need to require that μ be invariant with respect to ζ. Such probability measures are easily identifiable, since the explicit coordinate representation (2) that we have given for this involution only requires the computation of the polar decomposition of a matrix which is readily available in mathematical computer packages. Note also that, if the Radon–Nikodým derivative f in the statement of Theorem 1 has compact support, then the Fréchet mean of the corresponding probability measure does exist.

References

1. F. L. BOOKSTEIN, 'Size and shape spaces for landmark data in two dimensions' *Statist. Sci.* 1 (1986) 181–242 (with discussion).
2. F. L. BOOKSTEIN, *Morphometric tools for landmark data: geometry and biology* (Cambridge University Press, Cambridge, 1991).
3. S. HELGASON, *Differential geometry, Lie groups and symmetric spaces* (Academic Press, New York, 1978).

4. J. Jost, *Riemannian geometry and geometric analysis* (Springer, Berlin, 1998).
5. D. G. Kendall, 'Shape manifolds, procrustean metrics, and complex projective spaces,' *Bull. London Math. Soc.* 16 (1984) 81–121.
6. D. G. Kendall, D. Barden, T. K. Carne, and H. Le, *Shape and shape theory* (John Wiley, Chichester, 1999).
7. H. Le, 'On consistency of procrustean mean shapes', *Adv. Appl. Prob.* 30 (1998) 53–63.
8. H. Le, 'Locating Fréchet means with application to shape spaces,' *Adv. Appl. Prob.* (2001).
9. H. Le and A. Kume, 'Detection of shape changes in biological features,' *J. Microscopy* 200 (2000) 140–147.
10. H. Le and A. Kume, 'The Fréchet mean shape and the shape of the means,' *Adv. Appl. Prob.* 32 (2000) 101–113.
11. H. Le and C. G. Small, 'Multidimensional scaling of simplex shapes,' *Pattern Recognition* 32 (1999) 1601–1613.
12. B. O'Neill, *Semi-Riemannian geometry* (Academic Press, Orlando, 1983).
13. C. G. Small, *The statistical theory of shape* (Springer, New York, 1996).
14. J. A. Wolf, *Spaces of constant curvature* (McGraw–Hill, New York, 1967).

School of Mathematical Sciences
University of Nottingham
University Park
Nottingham NG7 2RD

Girton College
Cambridge CB3 0JG