

## PRINCIPAL SYMMETRIC SPACE ANALYSIS

CHARLES CURRY

Department of Mathematical Sciences, NTNU, Gjøvik, Norway

STEPHEN MARSLAND

School of Mathematics and Statistics, Victoria University of Wellington  
Wellington, New Zealand

ROBERT I MCLACHLAN

School of Fundamental Sciences, Massey University  
Palmerston North, New Zealand

**ABSTRACT.** Principal Geodesic Analysis is a statistical technique that constructs low-dimensional approximations to data on Riemannian manifolds. It provides a generalization of principal components analysis to non-Euclidean spaces. The approximating submanifolds are geodesic at a reference point such as the intrinsic mean of the data. However, they are local methods as the approximation depends on the reference point and does not take into account the curvature of the manifold. Therefore, in this paper we develop a specialization of principal geodesic analysis, *Principal Symmetric Space Analysis*, based on nested sequences of totally geodesic submanifolds of symmetric spaces. The examples of spheres, Grassmannians, tori, and products of two-dimensional spheres are worked out in detail. The approximating submanifolds are geometrically the simplest possible, with zero exterior curvature at all points. They can deal with significant curvature and diverse topology. We show that in many cases the distance between a point and the submanifold can be computed analytically and there is a related metric that reduces the computation of principal symmetric space approximations to linear algebra.

**1. Introduction.** Principal Components Analysis (PCA [12]), which is traditionally applied to data in a Euclidean space  $E^n$ , has many notable features that have made it one of the most widely-used of all statistical techniques. We single out the following properties:

1. The approximating subspaces (affine subspaces of  $E^n$ ) have zero extrinsic curvature;
2. any two affine subspaces of the same dimension are related by a Euclidean transformation;
3. the best approximations of each dimension are nested (that is, the best approximation by a  $k$ -dimensional subspace lies in the best approximation by a  $k + 1$ -dimensional subspace); and
4. the best approximations of each dimension from 0 to  $n - 1$  can be computed easily using linear algebra.

---

2010 *Mathematics Subject Classification.* Primary: 62H25, 53C35, 53C42.

*Key words and phrases.* PCA, PGA, totally geodesic submanifolds, subtori, symmetric spaces.

The underlying idea of PCA has been extended to deal with data on non-Euclidean manifolds. One such method is that of Principal Geodesic Analysis (PGA [6, 7, 9]). For data on a Riemannian manifold  $M$ , PGA relies on a chosen reference point  $x \in M$ , often taken to be the intrinsic mean of the data, i.e., a point that (if it exists) minimizes the sum of the squares of the distances from  $x$  to each data point. The approximating submanifolds in PGA are those submanifolds  $N$  that are “geodesic at  $x$ ”, namely those with the property that geodesics of  $N$  passing through  $x$  are also geodesics of  $M$ . Such submanifolds are the Riemannian exponentials of some subspace of  $T_x M$  and have the property that the extrinsic curvature of  $N$  is zero at  $x$ .<sup>1</sup> Fletcher et al. [6] introduced both this idea and (in view of the complexity of computing these submanifolds and their relationship to the data) a simplification now known as *linearized* PGA (or *tangent PCA*, tPCA), illustrated in Figure 1. Here the data is pulled back to the tangent space  $T_x M$  by the logarithm of the Riemannian exponential map at  $x$ . PCA can now be applied to the data on this Euclidean vector space.

Linearized PGA is now widely used for data on manifolds; see [22] for further background. However, it can be unsuitable if the manifold is highly curved, when two geodesics with common base points and distant tangent vectors may pass close to each other or intersect (see Figure 1). In this situation, nearby data points would become far apart in their linear approximation. For this reason, Sommer [22] implemented *exact* PGA as originally envisaged by Fletcher et al. [6]. Different variants either maximize the variance explained or minimize the residuals, and also differ in the selection of the reference point  $x$  [11, 19, 22]. Sequences of nested approximations can be constructed in either increasing or decreasing dimension [4]. Other approximating submanifolds have also been considered, such as nested spheres [13] and exponential barycentric subspaces [19].

In seeking to avoid the dependence on the reference point of PGA, and the curvature limitations of linearized PGA, in this work we focus on property (1) identified previously. A submanifold  $N$  of a Riemannian manifold  $M$  has zero extrinsic curvature if and only if it is totally geodesic (i.e., any geodesic of  $N$  is also a geodesic of  $M$ ). Such submanifolds provide excellent approximating spaces, being in a sense the flattest or simplest possible lower-dimensional representations of the data. One-dimensional totally geodesic submanifolds are geodesics, which are widely used for 1-dimensional interpolation and data fitting on manifolds [8, 14, 20]. Totally geodesic submanifolds are geodesic at all points; thus our approximating submanifolds are special cases of those used in (exact) PGA, that avoid the use of a reference point.

Generic manifolds have no totally geodesic submanifolds of dimension higher than 1, but Riemannian symmetric spaces have many. Examples of Riemannian symmetric spaces are Euclidean spaces, spheres, projective spaces, Grassmannians, compact Lie groups with bi-invariant Riemannian metrics, and products of these; they are all important examples of nonlinear domains for data. We will see that the structure of totally geodesic submanifolds offers rich possibilities for data reduction and for the discovery of hidden structure in data sets. Totally geodesic submanifolds of Riemannian symmetric spaces are themselves Riemannian symmetric, which offers the possibility of a nested structure as in property (3) previously.

---

<sup>1</sup>In the PGA literature, the term “geodesic subspace” is sometimes used as a synonym for “submanifold geodesic at a point.”

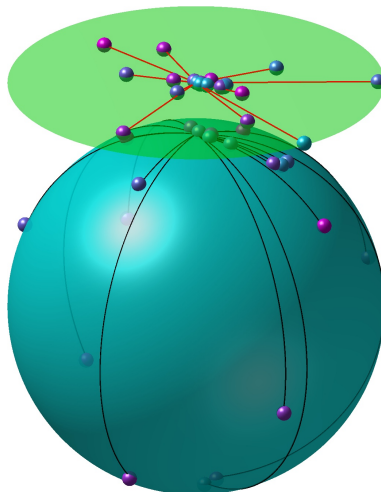


FIGURE 1. In linearized Principal Geodesic Analysis, the data (here 20 points on a sphere) is pulled back to the tangent space (shown here as a disk) of the intrinsic mean using geodesics. Data points near the mean are well represented, but data points far from the mean (here, near the south pole) become far apart in the linear approximation.

Although some form of nesting is desirable, we will see (e.g., in §6) that, given two best approximating totally geodesic submanifolds, one is not necessarily contained in the other. To overcome this, in this paper we define in Definition 1.2 the *symmetric space approximations* of a dataset in a Riemannian symmetric space. This is a set whose elements are best approximating totally geodesic submanifolds. Applying this construction recursively gives the *principal symmetric space approximation* (Definition 1.3) which is structured as a rooted tree. In this sense the nesting structure is retained, although it may be complicated in specific instances such as polyspheres (§6).

In Section 2 we review the relevant elements of symmetric spaces. In particular, the determination of totally geodesic submanifolds can be reduced (Theorem 2.4) to a purely algebraic equation (Equation 2) in the Lie algebra of the symmetry group of the symmetric space. Solving this equation may be difficult, however; it has been solved completely only in a few cases [2, 15, 24].

In the remainder of the paper, therefore, we proceed by example. Section 3 considers data on the  $n$ -dimensional sphere  $S^n$ . Section 4 considers data on the Grassmannian  $G(k, n)$  of  $k$ -planes in  $\mathbb{R}^n$ ; even here we need to restrict to the simple submanifolds  $G(k, m)$  of  $k$ -planes in  $\mathbb{R}^m$ . In both of these cases, we will show that the approximation problem can be linearised so that PCA-like nested sequences of approximating submanifolds can be determined using linear algebra.

This linearization is expressed in terms of a *compatible metric* (Definition 2.6), a non-Riemannian metric that approximates the Riemannian metric for nearby points. An example is the chordal (straight line) distance between points on a Euclidean sphere. The linearization does not depend on a reference point.

More complicated cases are handled in Section 5 on tori and Section 6 on polyspheres  $(S^2)^n$ . Each of these has an infinite number of distinct types of totally geodesic submanifolds and each reveals new features of the general situation.

**1.1. Principal symmetric space approximation.** We now introduce the central ideas of principal symmetric space approximation.

**Definition 1.1.** The set of connected totally geodesic submanifolds of the symmetric space  $M$  is denoted  $\mathcal{TG}(M)$ .

A group  $G$  acts on  $\mathcal{TG}(M)$  and partitions it into group orbits. The orbit through  $N \in \mathcal{TG}(M)$  is  $\{gx : g \in G, x \in N\}$ ; each element in this orbit is of the form  $gN := \{gx : x \in N\}$  for some  $g \in G$ . We regard the submanifolds in each orbit as being of equivalent structure and complexity, so that if there is a unique best approximation within an orbit, we choose it. However, the submanifolds from different orbits are different geometrically and are best regarded as representing different models. (For example, in the Euclidean case the affine subspaces of a given dimension form a single orbit. On the torus  $\mathbb{T}^2$ , geodesics of different winding numbers belong to different orbits.)

Let  $x \in M$  and let  $N \in \mathcal{TG}(M)$  be a totally geodesic submanifold of  $M$ . Let  $d(x, y)$  be the (geodesic) distance between  $x$  and  $y$  in  $M$ . We use the standard distance

$$d(x, N) = \min_{y \in N} d(x, y) \quad (1)$$

from a point to a submanifold as used in exact PGA [22]. As there, the distance  $d$  in Eq. (1) might not be well-defined, as the existence of the minimum depends on global properties of the manifolds. The residual (or residual error, or reconstruction error) of the data set  $X := (x_1, \dots, x_n)$ ,  $x_i \in M$  with respect to  $N$  is defined by

$$d(X, N) = \left( \sum_{i=1}^n d(x_i, N)^2 \right)^{1/2}.$$

**Definition 1.2.** The *symmetric space approximations* of  $X \in M^n$  with respect to  $M$  are the elements of

$$\mathcal{SSA}(X, M) := \left\{ gN : g = \operatorname{argmin}_{g \in G} d(X, gN), \quad N \in \mathcal{TG}(M) \right\}$$

where local minima are taken in the metric on  $G$ , that is, there is a neighbourhood  $U$  of  $g$  such that  $d(X, gN) \leq d(X, hN)$  for all  $h \in U$ .

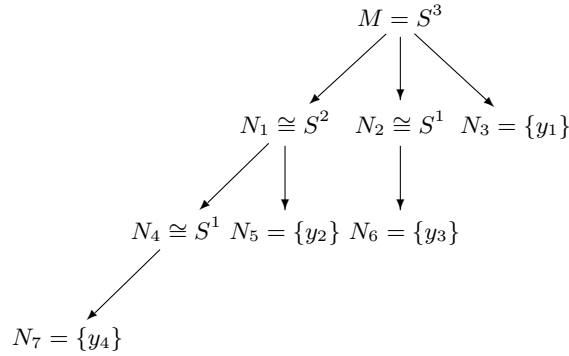
Thus, each element of  $\mathcal{SSA}(X, M)$  is a totally geodesic submanifold  $N$  of  $M$ , which best approximates the data in the sense that the approximation cannot be improved by passing to  $gN$  where  $g$  is close to the identity. Existence of symmetric space approximations depends on global properties of the manifolds.

As each  $N \in \mathcal{SSA}(X, M)$  is a Riemannian symmetric space, it typically has many totally geodesic submanifolds itself. These are already contained in  $\mathcal{TG}(M)$ . We can now calculate the symmetric space approximations of  $X$  with respect to each such  $N$ . Repeating this construction gives a tree of submanifolds. Each branch contains a nested sequence of approximations of decreasing dimensions, with each branch terminating in a submanifold of dimension 0, that is, a point.

**Definition 1.3.** A *principal symmetric space approximation*  $\mathcal{PSSA}(X, M)$  of  $X$  with respect to  $M$  is a rooted tree in which

1. each node is a totally geodesic submanifold of  $M$ ;
2. the root node is  $M$ ; and
3. each child of a node  $N$  is a proper submanifold of  $N$  and is a symmetric space approximation of  $X$  with respect to  $N$ .

Examples are the unbranched tree  $E^n \supset E^{n-1} \supset \dots \supset E^0$  found in Euclidean PCA, and the 2-node tree  $M \supset \{x\}$  for any Riemannian manifold  $M$ , where  $x$  is the intrinsic mean of  $X$ . A  $\mathcal{PSSA}(X, S^3)$  may look like:



where all of the  $N_i$  may be distinct. Here  $y_1$  is the intrinsic mean of the data. Depending on the application, it may not be relevant to compute the entire tree. For example, only the unbranched PSSA  $N_7 \subset N_4 \subset N_1 \subset M$  may be wanted, which can be computed in a backwards approximation (i.e., compute first  $N_1$ , then  $N_4$ , then  $N_7$ .)

**2. Symmetric spaces.** We give a brief account of symmetric spaces relevant to the sequel. The material presented here is standard, see for instance [16, Chapter XI].

**Definition 2.1.** A symmetric space is a triple  $(G, H, \sigma)$  where  $G$  is a connected Lie group,  $\sigma$  is an involutive automorphism of  $G$ , and  $H$  is a closed subgroup of  $G$  such that  $H$  lies between the isotropy subgroup  $G_\sigma$  and its identity component  $G_\sigma^o$ .

In particular, the manifold  $M = G/H$  is a canonically reductive homogeneous space (the canonical decomposition  $\mathfrak{g} = \mathfrak{h} + \mathfrak{m}$  arises by noting that  $\sigma$  induces an involution of  $\mathfrak{g}$ , for which we let  $\mathfrak{h}$  and  $\mathfrak{m}$  be the  $+1$  and  $-1$  eigenspaces respectively [16]) and hence comes equipped with a canonical linear connection. Let  $s_o$  be the automorphism of  $G/H$  induced by  $\sigma$ . For any point  $x = g.o$  where  $o$  is the origin, the mapping  $s_x = g.s_o.g^{-1}$  is independent of the choice of  $g$ . Moreover,  $s_x$  is a symmetry of the canonical connection for all  $x$ , i.e. a diffeomorphism of a neighbourhood of  $x$  onto itself sending  $\exp(X) \rightarrow \exp(-X)$  for any tangent vector  $X$ . We now present the infinitesimal picture.

**Definition 2.2.** A symmetric Lie algebra is a triple  $(\mathfrak{g}, \mathfrak{h}, \sigma)$  where  $\mathfrak{g}$  is a Lie algebra,  $\sigma$  is an involutive automorphism of  $\mathfrak{g}$ , and  $\mathfrak{h} \subset \mathfrak{g}$  is the Lie subalgebra of elements fixed by  $\sigma$ .

There is a one-to-one correspondence between effective symmetric Lie algebras and almost effective (i.e., the only normal subgroups of  $G$  are discrete) symmetric spaces with  $G$  simply connected and  $H$  connected.

Letting  $\mathfrak{g} = \mathfrak{h} + \mathfrak{m}$  be the canonical decomposition, we find the following relations, which suffice to characterize symmetric Lie algebras:

$$[\mathfrak{h}, \mathfrak{h}] \subset \mathfrak{h}, \quad [\mathfrak{h}, \mathfrak{m}] \subset \mathfrak{m}, \quad [\mathfrak{m}, \mathfrak{m}] \subset \mathfrak{h}.$$

**Example 1.** The oriented Grassmannian  $G_+(k, n)$  of oriented  $k$ -planes in  $\mathbb{R}^n$  is a symmetric space. The symmetric space structure is described by  $G_+(k, n) \simeq SO(n)/(SO(k) \times SO(n-k))$ , with automorphism  $\sigma(A) = SAS^{-1}$ ,

$$S = \begin{pmatrix} -I_k & 0 \\ 0 & I_{n-k} \end{pmatrix},$$

where  $I_k$  is the  $k \times k$  identity matrix. The case  $k = 1$  gives the symmetric space structure of the sphere  $S^n$ . The unoriented case  $G(k, n) \simeq O(n)/(O(k) \times O(n-k))$  is similar, and specialization to  $k = 1$  then gives projective spaces. We also note that there is a natural direct product of symmetric spaces:  $(G, H, \sigma) \times (G', H', \sigma') = (G \times G', H \times H', \sigma \times \sigma')$ .

A submanifold  $N \subset M$  is said to be totally geodesic if for all points  $x \in N$  and tangent vectors  $X \in T_x(N)$ , the geodesic  $\exp(tX)$  is contained in  $N$  for sufficiently small  $t$ . Where  $M$  is a Riemannian manifold, this is equivalent to requiring that the induced metric on  $N$  coincides with the metric on  $M$ .

**Definition 2.3.** A *Lie triple system*  $\mathfrak{n}$  is a subspace of a Lie algebra for which

$$[[\mathfrak{n}, \mathfrak{n}], \mathfrak{n}] \subset \mathfrak{n}. \quad (2)$$

The following result underlies our interest in totally geodesic submanifolds of symmetric spaces:

**Theorem 2.4.** [16, 24] *Let  $(G, H, \sigma)$  be a symmetric space with symmetric Lie algebra  $\mathfrak{g} = \mathfrak{h} + \mathfrak{m}$ . There is a one-to-one correspondence between complete totally geodesic submanifolds  $M'$  containing the origin and Lie triple systems  $\mathfrak{m}' \subset \mathfrak{m}$ . Moreover  $(G', H', \sigma')$  is a symmetric subspace, where  $G'$  is the largest connected Lie subgroup of  $G$  leaving  $M'$  invariant,  $H' = G' \cap H$ , and  $\sigma' = \sigma|_{G'}$ .*

Note that the proof constructs the symmetric subalgebra  $(\mathfrak{g}', \mathfrak{h}', \sigma')$ . Indeed, given such an  $\mathfrak{m}'$ , we take  $\mathfrak{h}' = [\mathfrak{m}', \mathfrak{m}']$ , then set  $\mathfrak{g}' = \mathfrak{h}' + \mathfrak{m}'$ .

The problem of classifying totally geodesic submanifolds of symmetric spaces is thus reduced to an algebraic one, although it remains a difficult task [2, 15, 24]. Moreover, there may exist complicated totally geodesic submanifolds that are of little physical relevance, so in some cases we restrict our attention to subfamilies of symmetric subspaces.<sup>2</sup>

The notion of a symmetric space approximation requires a distance function on the manifold. It is natural to specify this through a Riemannian metric. This makes most sense where our notion of totally geodesic submanifold coincides with the Riemannian geodesics, as summarized by the following definition.

**Definition 2.5.** A Riemannian symmetric space is a symmetric space for which the canonical connection coincides with the Riemannian (Levi-Civita) connection.

<sup>2</sup>For example, the Grassmannian of (oriented) two-dimensional subspaces of a vector space has six different families of totally geodesic submanifolds [15]. We will only consider one of these families in §4, namely the subspaces orthogonal to a given subspace.

This implies that the symmetries  $s_x$  are isometries. A symmetric space equipped with a metric is Riemannian symmetric if the metric is  $G$ -invariant. Given a symmetric space  $(G, H, \sigma)$  for which  $\text{ad}_{\mathfrak{g}}(H)$  is compact, a  $G$ -invariant Riemannian metric may be constructed in a canonical manner.

All of the symmetric spaces we consider are canonically Riemannian symmetric spaces. Nonetheless, for practical purposes we will minimize distances that differ from the Riemannian distance, typically to obtain a linearization of the minimization problem.

**Definition 2.6.** Two metrics  $d_1, d_2$  are *compatible* if they agree up to first order for nearby points, that is, if  $d_2(x, y) = d_1(x, y) + \mathcal{O}(d_1(x, y)^2)$ .

In a non-Riemannian metric space, the length of a curve is defined by a Riemann sum, and thus one still has the concept of geodesic and of totally geodesic submanifolds in this case. Moreover, the geodesics and totally geodesic submanifolds of a given smooth manifold equipped with two compatible metrics coincide. Thus, although perturbing the Riemannian metric of a Riemannian symmetric space changes the *specific* principal symmetric space approximation corresponding to a given set of data, it does not change the system of totally geodesic submanifolds itself.

**3. Spheres.** Datasets on high-dimensional spheres arise naturally whenever we have a set of measurements in a Euclidean space for which the magnitude is irrelevant. One important instance concerns directional data, see [13] and references therein for more examples.

The connected totally geodesic submanifolds of  $S^n$  are the spheres  $S^k$ , realized as the image of a standard sphere  $x_1^2 + \cdots + x_{k+1}^2 = 1$  in  $\mathbb{R}^{n+1}$  under an element of  $SO(n+1)$  [24, Thm 1].

We consider first the case of  $S^2$ , represented as the set of unit vectors in  $\mathbb{R}^3$ . Geodesics on  $S^2$  are precisely the great circles, which may be described as the set of points in  $S^2$  orthogonal to a given unit vector  $v$ . We call this great circle  $S_v := \{w \in S^2 : w \cdot v = 0\}$ . In this case the Riemannian distance from a point  $x$  to  $S_v$  is the angle between  $x$  and  $v$ , that is,

$$d(x, S_v) = \sin^{-1}(|x \cdot v|).$$

Note that the great circle with axis  $v$  consists of the intersection of  $S^2$  and the plane with normal vector  $v$ . More generally, the totally geodesic submanifolds of  $S^n$ , viewed as submanifolds of  $\mathbb{R}^{n+1}$ , are precisely the intersections  $S_N$  of  $S^n$  with linear subspaces  $N$  of  $\mathbb{R}^{n+1}$ .

**Lemma 3.1.** *Let  $N$  be a codimension- $m$  subspace of  $\mathbb{R}^{n+1}$  and let  $\{v_1, \dots, v_m\}$  be an orthogonal basis for  $N^\perp$ . Then  $S_N := S^n \cap N$  is a Euclidean sphere of dimension  $k = n - m$  and the distance between  $x \in S^n$  and  $S_N$  is*

$$d(x, S_N) = \sin^{-1} \left( \sqrt{\sum_i (x \cdot v_i)^2} \right).$$

*Proof.* The angle  $\theta (= d(x, S_N))$  between  $x$  and  $N$ , and the angle  $\hat{\theta}$  between  $x$  and  $N^\perp$ , are complementary angles. Likewise, the angle  $\theta$  between  $x$  and  $S_N$  and the angle  $\hat{\theta}$  between  $x$  and  $S^n \cap N^\perp$  are complementary angles. Let  $\hat{x}$  be the orthogonal

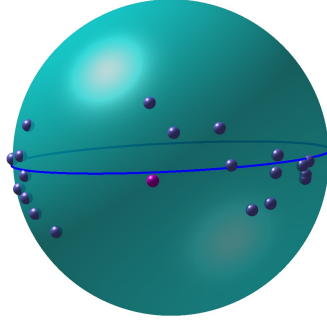


FIGURE 2. Twenty data points on  $S^2$  (dark blue) together with the point (pink) and the great circle (blue) that best approximate the data in the Riemannian metric, computed using nonlinear optimization. The best point (the intrinsic mean of the data) does not lie on the best great circle.

projection of  $x$  to  $N^\perp$ , that is,  $\hat{x} = \sum_{i=1}^m (x \cdot v_i) v_i$ . Then

$$\sin \theta = \cos \hat{\theta} = \|\hat{x}\| = \sqrt{\sum_{i=1}^m (x \cdot v_i)^2}.$$

□

Finding best approximations in the Riemannian metric requires nonlinear optimization and leads to a branched nested sequence of approximations (see Figure 2). We now linearize the metric so that best approximations can be determined using linear algebra. We call the *projection distance*  $d_p(x, v)$  between  $x$  and  $v \in S^n$  the shortest Euclidean distance from  $x$  to  $\text{span}(v)$  in  $\mathbb{R}^{n+1}$ . (Equivalently, from  $v$  to  $\text{span}(x)$ .)

**Lemma 3.2.** *Let  $N$  be a subspace of  $\mathbb{R}^{n+1}$  and let  $\{v_1, \dots, v_m\}$  be an orthogonal basis for  $N^\perp$ . Then the projection distance between  $x \in S^n$  and  $S_N$  is*

$$d_p(x, S_N) = \sqrt{\sum_{i=1}^m (x \cdot v_i)^2}.$$

*Proof.* Let  $\theta$  be the angle between  $x$  and  $v \in S^n$ , that is,  $\cos \theta = x \cdot v$ . Then  $d_p(x, v) = \sin \theta$ . The same construction as in Lemma 3.1, except measuring distances as  $\sin \theta$  instead of  $\theta$ , gives the result. □

Note that the projection distance between two points is nonlinear; its use is favoured here because it becomes linear when calculating distances to subspheres  $S_N$ . The projection distance is compatible with the Riemannian distance.

**Proposition 1.** *Let  $X$  be the matrix whose columns consists of the data points  $x_i \in S^n$ , where  $S^n$  is identified with the unit sphere in  $\mathbb{R}^{n+1}$ . Then for any  $m$  with  $0 < m < n$ , the best approximating  $(n-m)$ -sphere in the projection distance is given by  $S_N$ , where  $N$  is the span of the singular vectors corresponding to the smallest  $m$  singular values of  $X^\top$ .*

*Proof.* Let  $V = [v_1, \dots, v_m] \in \mathbb{R}^{n \times m}$ . We have  $d_p(x_j, S_N)^2 = \sum_{i=1}^m (x_j \cdot v_i)^2$ , and thus

$$d_p(X, S_N)^2 = \sum_{i=1}^m \sum_{j=1}^d (x_j \cdot v_i)^2 = \|X^T V\|_F^2.$$

We seek to minimize  $d_p(X, S_N)$  subject to the constraint that  $V$  is orthogonal. Introducing a Lagrange multiplier  $\Lambda \in \mathbb{R}^{m \times m}$  for the constraint, where  $\Lambda^T = \Lambda$ , we need to make

$$\|X^T V\|_F^2 + \text{tr} \Lambda (V^T V - I) = \text{tr}(V^T X X^T V) + \text{tr} \Lambda (V^T V - I)$$

stationary in  $W$ . The variational equations are

$$X X^T V = V \Lambda, \quad V^T V = I.$$

At any solution to these equations, the objective function is  $d_p(X, S_N)^2 = \text{tr}(V^T X X^T V) = \text{tr}(V^T V \Lambda) = \text{tr}(\Lambda)$ . Given any solution  $(V, \Lambda)$  to these equations, orthogonally diagonalize  $\Lambda = Z \Omega Z^T$  where  $Z^T Z = I$  and  $\Omega$  is diagonal. Then  $(WZ, \Omega)$  is also a solution. The value of the objective function,  $\text{tr} \Lambda = \text{tr} \Omega$ , is the same for both solutions. Therefore, we can take  $\Lambda$  to be diagonal. Therefore, the stationary points are those for which the columns of  $V$  are eigenvectors of  $X X^T$  (that is, singular vectors of  $X^T$ ) and the diagonal entries of  $\Lambda$  are the associated eigenvalues of  $X X^T$  (that is, squares of the singular values of  $X^T$ ). The minimum value of the objective function is obtained by taking the  $m$  smallest singular values.  $\square$

Note that, in the sense of Euclidean PCA, if we regard the data as a set of points in  $\mathbb{R}^{n+1}$ , the best approximating  $(n+1-m)$ -subspace is just the span of the singular vectors associated with the  $n+1-m$  *largest* singular values of  $X^T$ . That subspace is the orthogonal complement of the span of the singular vectors associated with the  $m$  smallest singular values found in the proposition. Thus in this case, the two approximations coincide (after intersecting with  $S^n$ ).

The linearization of the distance function, considered here, reduces the calculation to linear algebra, produces unique best approximations, and also provides the nesting property shared by Euclidean PCA: the best  $S^p$  lies inside the best  $S^k$  for  $p < k$ :

**Corollary 1.** *The principal symmetric space approximation of  $X$  with respect to  $S^n$  in the projection metric is the unbranched tree  $S^n \supset S^{n-1} \supset \dots \supset S^1$ , where each  $S^p$  is as determined in Proposition 1.*

As stated, we have restricted the dimension of the subspheres in Proposition 1 and Corollary 1 to be positive. If they are applied with  $m = n$  to yield 0-dimensional approximations, they yield the *pair* of antipodal points that best approximates the data in the projective metric, because  $S_N$  is then disconnected. Depending on the application, this may be what is wanted. Even if the best *single* point is wanted, the best such  $S_N$  may still be a usefully good approximation if the data is, in fact, strongly clustered around a single point. If the data is not strongly clustered, and the best single point is wanted, then it may be necessary to switch to another metric (e.g., the Riemannian metric) and calculate the best point within each  $S^m$  in Cor. 1, creating a branched tree of approximations.

**Example 2.** We present a sequence of 3 examples of synthetic datasets on  $S^3$ . Each contains 20 data points. In the first dataset, each  $x_i$  is the projection of a point in  $\mathcal{N}(0, \text{diag}(1, 1, 0.1, 0.05)^2)$  to  $S^3$ , where  $\mathcal{N}(\mu, \Sigma)$  is the normal distribution

with mean  $\mu$  and covariance  $\Sigma$ . The data lies close to the 2-sphere  $x_4 = 0$  and even closer to the great circle  $x_3 = x_4 = 0$ . The singular values of  $X^\top$  were found to be  $(0.3486, 0.4095, 3.0571, 3.2195)$ . Thus the residual of the best  $S^2$  approximation to the data is 0.3486, and the residual of the best  $S^1$  approximation (shown in Fig. 3, left) is 0.5378. The residual of the best  $S^0$  approximation is 3.1040 and is clearly found to be not relevant. Likewise, the intrinsic mean is not relevant for this dataset.

In the second dataset, each  $x_i$  is the projection of a point in  $\mathcal{N}(0, \text{diag}(1, 0.3, 0.1, 0.05)^2)$  to  $S^3$ . Thus the data are more strongly clustered around the 0-sphere  $x_1 = \pm 1$ ,  $x_2 = x_3 = x_4 = 0$ . This is revealed in the singular values  $(0.3339, 0.9568, 1.2380, 4.1762)$ . The residual of the best  $S^2$ ,  $S^1$ , and  $S^0$  approximations are 0.3339, 1.0134, and 1.5998, respectively. The nested approximations are shown in Fig. 3 (centre). The intrinsic mean is not relevant for this dataset.

In the third dataset, each  $x_i$  is the projection of  $\mathcal{N}([1, 0, 0, 0], \text{diag}(0, 0.4, 0.1, 0.05)^2)$  to  $S^3$  and is thus clustered around the point  $(1, 0, 0, 0)$ . The singular values are  $(0.1712, 0.3685, 1.5513, 4.1747)$ . The nested approximations are shown in Fig. 3 (right). The intrinsic mean (which here coincides with the best  $S^0$ , in the appropriate metric) is relevant for this dataset.

**4. Grassmannians.** The Grassmannian  $G(k, n)$  of  $k$ -dimensional subspaces (or  $k$ -planes) of  $\mathbb{R}^n$  is a symmetric space (see Example 1). Data comprising subspaces may arise if we wish to track the eigenspace decomposition of symmetric matrices such as diffusion tensors, or if we collect a sequence of low-dimensional approximating subspaces to Euclidean data using Euclidean PCA as some parameter (e.g., time) evolves. Related applications occur in computer vision and signal processing [23].

The classification of the totally geodesic submanifolds of Grassmannians is surprisingly complicated [15]. Here we restrict our attention to a specific type of totally geodesic submanifold, namely the space of  $k$ -planes orthogonal to a given subspace  $W$  of  $\mathbb{R}^n$ .

**Lemma 4.1.** *The space of  $k$ -planes in  $\mathbb{R}^n$  orthogonal to a given  $(n-m)$ -dimensional subspace  $W$  of  $\mathbb{R}^n$  is a totally geodesic submanifold of  $G(k, n)$  and is diffeomorphic to  $G(k, m)$ .*

*Proof.* The symmetric algebra has canonical decomposition  $\mathfrak{g} = \mathfrak{h} + \mathfrak{m}$ , where [16, Ex. XI.10.3]

$$\begin{aligned} \mathfrak{g} &= \mathfrak{o}(n), \\ \mathfrak{h} &= \mathfrak{o}(k) + \mathfrak{o}(n-k) \\ &= \left\{ h(U, V) := \begin{pmatrix} U & 0 \\ 0 & V \end{pmatrix}, U \in \mathfrak{o}(k), V \in \mathfrak{o}(n-k) \right\}, \end{aligned}$$

and

$$\mathfrak{m} = \left\{ m(X) := \begin{pmatrix} 0 & -X^\top \\ X & 0 \end{pmatrix}, X \in \mathbb{R}^{(n-k) \times k} \right\}.$$

A geodesic connecting two points on a Grassmannian may be characterized as a linear interpolation of each principal angle. Fix an orthogonal basis  $(e_1, \dots, e_k)$  of the subspace at the origin, and extend this to an orthogonal basis  $\{e_i\}$  of  $\mathbb{R}^n$ . Then

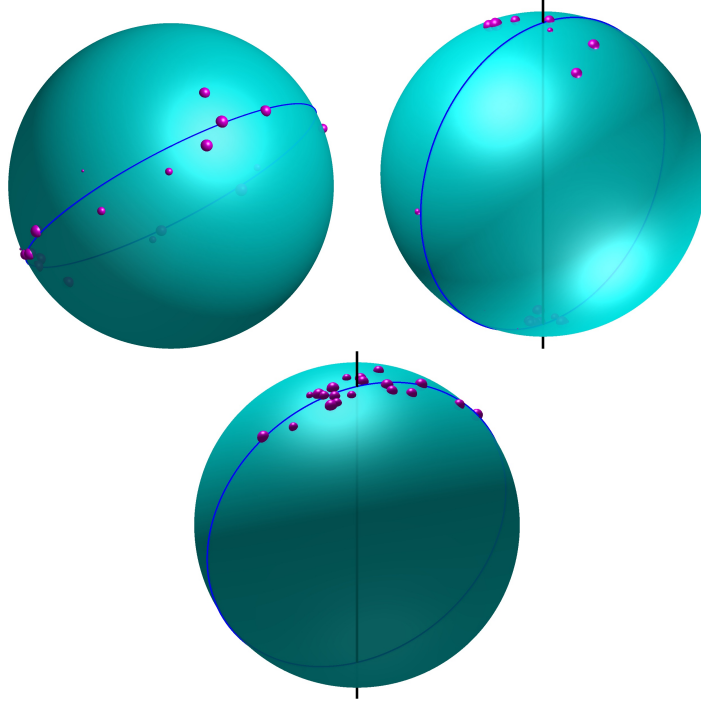


FIGURE 3. Results for datasets 1–3 of Example 2. In each case, the best subspheres that approximate a set of 20 points on  $S^3$  is shown. Data points further from the best  $S^2$  are shown smaller. The best  $S^1$  is shown in blue, lying on the best  $S^2$  in teal. In datasets 2 and 3, the axis of the best  $S^0$  (which consists of two antipodal points) is shown in black. In dataset 3, this also coincides with a standard mean of the data.

any set of  $k$  orthogonal vectors orthogonal to  $(e_1, \dots, e_k)$  may be written as

$$\begin{pmatrix} 0_{k \times k} \\ X \end{pmatrix}$$

with respect to the basis  $\{e_i\}$ , where  $X \in \mathbb{R}^{(n-k) \times k}$ . The geodesic connecting  $X$  to the origin, intersecting  $X$  at time  $t = 1$ , is  $\exp(tm(X)).o$ . In particular, suppose  $W$  is an  $(n - m)$ -dimensional subspace of  $\mathbb{R}^n$  orthogonal to  $(e_1, \dots, e_k)$ . Then any orthogonal basis  $(w_1, \dots, w_{n-m})$  for  $W$  may be written as

$$\begin{pmatrix} 0_{k \times (n-m)} \\ R \end{pmatrix}$$

with respect to the basis  $\{e_i\}$ , where  $R \in \mathbb{R}^{(n-k) \times (n-m)}$ . Then the geodesic  $\{\exp(tm(X)).o, t \in \mathbb{R}\}$  consists of subspaces orthogonal to  $W$  if and only if  $X^T R = 0$ . It therefore suffices to show that for any  $R$ , the subspace

$$\mathfrak{m}' = \left\{ m(X), X \in \mathbb{R}^{(n-k) \times k}, X^T R = 0 \right\}$$

defines a Lie triple system. A calculation shows that

$$[[m(X), m(Y)], m(Z)] = Z^T Y X^T - Z^T X Y^T - Y^T X Z^T + X^T Y Z^T$$

and we are done as  $X^T R = Y^T R = Z^T R = 0$  implies  $(Z^T Y X^T - Z^T X Y^T - Y^T X Z^T + X^T Y Z^T) R = 0$ .

Fixing an orthogonal basis of  $W$ , extending this to an orthogonal basis of  $\mathbb{R}^n$ , and expressing subspaces orthogonal to  $W$  in terms of this basis gives the required diffeomorphism.  $\square$

Let the columns of the matrix  $W \in \mathbb{R}^{n \times (n-m)}$  be an orthonormal basis for the subspace  $W$ . Let  $X, Y \in \mathbb{R}^{n \times k}$  be orthonormal bases for two elements  $X, Y$  of  $G(k, n)$ . The relationship between  $X$  and  $Y$  is measured by their *principal angles*  $\theta = (\theta_1, \dots, \theta_k)$ , defined by  $\cos \theta_k = \sigma_k(X^T Y)$ . The geodesic distance between  $X$  and  $Y$  in the Riemannian symmetric space  $G(k, n)$  is  $\|\theta\|_2$ . Another popular measure of distance between subspaces is  $\max_k \theta_k$  [10, p. 584]. However, like Conway et al. [3], we find that it is far easier and more natural to use the “chordal distance”  $\|\sin \theta\|_2$  (so named because when equipped with this metric,  $G(k, n)$  isometrically embeds in a Euclidean sphere). The chordal and geodesic metrics are compatible and thus have the same totally geodesic subspaces.

In the present context, the advantage of the chordal distance is that it linearizes the calculation of the distance from a  $k$ -plane to a totally geodesic submanifold.

**Lemma 4.2.** *The chordal distance between two subspaces  $X, Y \in G(k, n)$  is given by  $\|X^T Y^\perp\|_F$ , where  $Y^\perp$  is the orthogonal complement of  $Y$ .*

*Proof.* The squared chordal distance  $d_c(X, Y)^2$  is

$$\sum_{i=1}^k \sin^2 \theta_i = k - \sum_{i=1}^k \cos^2 \theta_i = k - \sum_{i=1}^k \sigma_i^2(X^T Y) = k - \|X^T Y\|_F^2.$$

The  $n \times n$  matrix  $Q = [Y|Y^\perp]$  is orthogonal, so we have

$$\|X^T Q\|_F^2 = \text{tr}(Q^T X X^T Q) = \text{tr}(X X^T) = k$$

and also

$$\|X^T Q\|_F^2 = \|X^T [Y|Y^\perp]\|_F^2 = \|X^T Y\|_F^2 + \|X^T Y^\perp\|_F^2.$$

Therefore

$$d_c(X, Y)^2 = k - \|X^T Y\|_F^2 = \|X^T Y^\perp\|_F^2.$$

$\square$

An immediate consequence is the following:

**Proposition 2.** *The chordal distance from a subspace  $X \in G(k, n)$  to the set  $G(k, m)$  of  $k$ -planes orthogonal to  $W \in \mathbb{R}^{n, n-m}$  is  $\|X^T W\|_F^2$ .*

*Proof.* We consider the cases  $k = m$  and  $k < m$  separately. If  $k = m$  then  $G(k, m)$  is a single point,  $Y^\perp = W$ , and we are done. If  $k < m$ , then an orthogonal basis for the orthogonal complement  $Y^\perp$  of any  $Y$  orthogonal to  $W$  may be written as  $[W|U]$  for some  $U \in \mathbb{R}^{n, m-k}$  that satisfies  $U^T W = 0$ . We have:

$$d_c(X, Y)^2 = \|X^T Y^\perp\|_F^2 = \|X^T W\|_F^2 + \|X^T U\|_F^2.$$

This is to be minimized over all choices of orthogonal  $U$  that are also orthogonal to  $W$ . Any  $U$  that is orthogonal to both  $W$  and  $X$  achieves  $\|X^T U\|_F = 0$ , giving the result.

Note that we can choose such a  $U$ , since:

$$\begin{aligned} \dim(X \cup W)^\perp &= n - (\dim X + \dim W - \dim(X \cap W)) \\ &= n - (k + (n - m) - \dim(X \cap W)) \\ &= m - k + \dim(X \cap W) \\ &> 0. \end{aligned}$$

□

As in the case of spheres, the best approximating Grassmannians can now be read off from the SVD of a matrix representing from the dataset.

**Proposition 3.** *Let  $X_1, \dots, X_d$  be a set of  $d$   $k$ -planes with orthogonal bases  $X_1, \dots, X_d$ . Then the matrix  $W$  minimizing the sum of squared chordal distances of the  $X_i$  to  $W$  is precisely the matrix of singular vectors of  $X^\top$  corresponding to its  $p$  smallest singular values, where  $X = [X_1, \dots, X_d] \in \mathbb{R}^{n \times kd}$  is the matrix obtained by concatenating the  $X_i$ s. The chordal distance of the  $X_i$  to  $W$  is the 2-norm of the  $p$  smallest singular values of  $X^\top$ . The principal symmetric space approximations are nested, in that the best  $G(k, p)$  lies in the best  $G(k, q)$  for  $p \leq q$ .*

*Proof.* The sum of the squared chordal distances is

$$d(X, W)^2 = \sum_{i=1}^d d(X_i, W)^2 = \sum_{i=1}^d \|X_i^\top W\|_F^2 = \|X^\top W\|_F^2.$$

This expression is formally identical to that studied in Proposition 1, hence the result follows as in Proposition 1. □

## 5. Tori.

**5.1. Products of symmetric spaces.** Given two symmetric spaces  $(G, H, \sigma)$  and  $(G', H', \sigma')$ , the direct product  $(G \times G', H \times H', \sigma \times \sigma')$  is also a symmetric space [16, p. 228]. The simplest case to consider is that of products of spheres,  $S^{a_1} \times S^{a_2} \times \dots$ . Here we focus on products of  $n$  circles giving the  $n$ -torus. The following section considers products of  $n$  2-spheres (Sec. 6).

In the previous examples, of spheres and Grassmannians, the action of the symmetry group  $G$  (e.g.,  $G = SO(3)$  acting on points or great circles in  $S^2$  by rotations) was transitive. Our task was limited to selecting, from the single group orbit available, the best point (or points). On products of spheres, the action of the symmetry group is not transitive; there are many (even infinitely many) distinct orbits. In fitting models with both continuous and discrete parameters, one common approach is to consider each value of the discrete parameters as specifying a different model; which value is chosen then corresponds to a model selection problem. This is the approach adopted here. We note that in the Bayesian paradigm model selection arises naturally through the choice of prior; however we will not pursue this further.

Two examples illustrate the complexity of the situation.

First, consider data consisting of  $n$  angles, i.e.  $x \in \mathbb{T}^n$ . Totally geodesic submanifolds are subtori described by resonance relations of the form  $a \cdot x = c$ ,  $a \in \mathbb{Z}^n$ ,  $c \in \mathbb{R}$ . Each fixed  $a$  specifies a different resonance relation, while the continuous parameter  $c$  selects the best model for a given discrete parameter  $a$ .

Second, consider data consisting of  $n$  points on  $S^2$ , i.e. spherical polygons. Totally geodesic manifolds are products of copies of  $S^1$  and  $S^2$ . An example is  $x_1$  lying on a great circle;  $x_2$  lying on a second great circle, and obeying a resonance

relation with  $x_1$ ;  $x_3$  arbitrary; and  $x_4, \dots, x_n$  being rotations of a fixed spherical polygon.

**5.2. Classification of totally geodesic submanifolds of tori.** The product  $(S^1)^n$  is a symmetric space that we identify with the flat torus  $\mathbb{T}^n := (\mathbb{R}/\mathbb{Z})^n$  with standard coordinates  $x \in [0, 1)^n$ . The connected totally geodesic submanifolds of  $\mathbb{R}^n$  are the affine subspaces; taking their translations by  $\mathbb{Z}^n$  and passing to the quotient gives the connected totally geodesic submanifolds of  $\mathbb{T}^n$ . Amongst these, we wish to select those that are regular submanifolds. We will describe them by the resonance relations that they satisfy. The group  $\mathbb{T}^n$  acts by translations on  $\mathbb{T}^n$  and leaves the resonance relations invariant; we regard the submanifolds that satisfy different resonance relations as belonging to different models. Thus, the problem of finding the principal symmetric space approximations to given data involves first fixing the resonance relation and then determining the best fitting submanifold that obeys that resonance relation.

We will show in Proposition 4 that the regular connected totally geodesic submanifolds of  $\mathbb{T}^n$  are all tori. Up to translations, they are parameterized by unimodular matrices  $A \in \mathbb{Z}^{k \times n}$ , i.e., matrices with integer entries whose  $k \times k$  minors do not have a common factor (i.e., their greatest common divisor is equal to 1). Specifically, they have the form

$$T := \{x \in \mathbb{T}^n : Ax = c\} \quad (3)$$

for some  $c \in [0, 1)^n$ .

**Example 3.** The case of geodesics in  $\mathbb{T}^2$  gives a feel for the requirement that the submanifold be represented by an unimodular matrix  $A$ . (i) The subset  $x_1 + \sqrt{2}x_2 = 0$  of  $\mathbb{T}^2$ , associated with  $A = [1, \sqrt{2}]$ , is totally geodesic, but it is an irregular submanifold. It is questionable for data fitting as it passes arbitrarily close to every point of the torus. (ii) The subset  $2x_1 = 0$  of  $\mathbb{T}^2$ , associated with  $A = [2, 0]$ , consists of the two vertical lines  $(0, y)$  and  $(\frac{1}{2}, y)$  for  $0 \leq y < 1$ . This set is a regular totally geodesic submanifold, but it is not connected—and  $A$  is not unimodular. (iii) The subset  $2x_1 + 5x_2 = c$ , associated with the unimodular matrix  $A = [2, 5]$ , is a regular, connected, totally geodesic submanifold of  $\mathbb{T}^2$ . We will give an example of fitting such a geodesic below.

**Proposition 4.** [17] *Every regular connected codimension- $k$  totally geodesic submanifold of  $\mathbb{T}^n$  is a subtorus given by Eq. (3) for some  $c \in [0, 1)^k$  and unimodular  $A \in \mathbb{Z}^{k \times n}$ .*

*Proof.* First let  $A$  be unimodular. We will show that  $T$  in Eq. (3) is a regular connected codimension- $k$  totally geodesic submanifold and is a subtorus. Rows can be added to  $A$  to create a matrix,  $C$ , of determinant 1 [21]. The linear map

$$\phi: \mathbb{R}^n \rightarrow \mathbb{R}^n, \quad \hat{x} \mapsto C\hat{x}$$

is invertible, therefore the map  $\hat{x} \mapsto A\hat{x}$  is surjective. Let  $x \in T$  and let  $\hat{x}$  be any point in  $\mathbb{R}^n$  such that  $\hat{x} \bmod 1 = x$ . We are given that  $A\hat{x} = c + m$  for some  $m \in \mathbb{Z}^k$ . From the surjectivity of  $\hat{x} \mapsto A\hat{x}$ , there is a  $p \in \mathbb{Z}^n$  such that  $Ap = m$ . Therefore  $A(\hat{x} - p) = c$ . That is, some integer translation of  $\hat{x}$  lies on the affine subspace  $\{\hat{x} \in \mathbb{R}^n : A\hat{x} = c\}$ , which is the cover of a connected totally geodesic submanifold of  $\mathbb{T}^n$ . Hence  $T$  is a totally geodesic submanifold of  $\mathbb{T}^n$ .

The map  $\phi$  descends to an automorphism of  $\mathbb{T}^n$ ; it provides a change of coordinates on  $\mathbb{T}^n$ . In coordinates  $y = Cx$ , the submanifold is given by  $y_1 = c_1, \dots, y_k =$

$c_k$ , with  $y_{k+1}, \dots, y_n \in [0, 1)^{n-k}$ . This submanifold is a connected regular submanifold of  $\mathbb{T}^n$  and is a subtorus.

To show the converse, let  $T$  be any regular connected totally geodesic submanifold of  $\mathbb{T}^n$ . Its translation  $U$  to the origin is a subgroup, hence a subtorus of  $\mathbb{T}^n$ . The kernel of the exponential map of the Lie algebra of  $U$  is a lattice in  $\mathbb{Z}^n$ . Form a matrix whose rows are a basis of this lattice. The null space of this matrix has a unimodular integer basis whose entries are the resonance relations satisfied by elements of  $U$  and  $T$ . These form the rows of  $A$ .  $\square$

The matrix  $A$  describes the resonance relations satisfied by the subtorus. If  $a_i \cdot x \bmod 1 = c_i$  for all  $i$ , then for any  $m_i \in \mathbb{Z}$  we have  $(\sum_{i=1}^k m_i a_i) \cdot x \bmod 1 = c_i$  as well. That is, the set of resonance relations forms a  $k$ -dimensional lattice

$$L := \left\{ \sum_{i=1}^k m_i a_i, \quad m_i \in \mathbb{Z} \right\} \subset \mathbb{Z}^n$$

with the rows of  $A$  as a basis. Two matrices  $A, A'$  describe the same lattice, and the same family of subtori, if there is a matrix  $Z \in GL(k, \mathbb{Z})$  such that  $A' = ZA$ . Recall that the *dual lattice*  $L^*$  is defined by

$$L^* := \{y \in \mathbb{R}^n : y \in \text{span}(L), \quad y \cdot x \in \mathbb{Z} \quad \forall x \in L\}$$

which has a basis given by the columns of  $A^\top (AA^\top)^{-1}$ ; see [25] for an introduction to lattices in  $\mathbb{Z}^n$ .

We first describe the geometry of the subtorus  $T$  defined by Eq. (3); see Figures 4–6 for examples.

**Proposition 5.** *The lifted subtorus  $\{x \in \mathbb{R}^n : x \bmod 1 \in T\}$  is a translation of the direct product of the subspace  $\text{null}(A)$  of  $\mathbb{R}^n$  and the dual lattice  $L^*$ , where  $L^*$  and  $\text{null}(A)$  are orthogonal.*

*Proof.* Let  $x_0 \in \mathbb{R}^n$  be any solution to  $Ax_0 = c$  so that the lifted subtorus is given by  $\{x_0 + y : Ay \in \mathbb{Z}^k\}$ . Decompose  $y$  into its components  $y_1 \in \text{null}(A)$  and  $y_2 \in \text{null}(A)^\perp$ . The component  $y_1$  is arbitrary. For  $y_2$ , note that  $\text{null}(A)^\perp = \text{im}(A^\top)$  so that  $y_2 = A^\top c$  for some  $c \in \mathbb{R}^k$ . Therefore  $Ay_2 = AA^\top c \in \mathbb{Z}^k$ . Since  $L = \{m^\top A : m \in \mathbb{Z}^k\}$ , a comparison with the definition of the dual lattice gives the result.  $\square$

**Example 4.** Let  $n = 3$  and  $k = 2$ . We consider the 1-dimensional subtorus (i.e. geodesic) that passes through 0 in direction  $d = [1, 2, 3]^\top$ . The transposed null space of  $d$  is

$$A := \begin{bmatrix} -3 & 0 & 1 \\ -2 & 1 & 0 \end{bmatrix}.$$

That is, the points on the subtorus are those that satisfy the resonance relations  $-3x + z = 0$  and  $-2x + y = 0$ . A basis for the dual lattice is given by

$$B := A^\top (AA^\top)^{-1} = \begin{bmatrix} -\frac{3}{14} & -\frac{1}{7} \\ -\frac{6}{14} & \frac{5}{7} \\ \frac{5}{14} & -\frac{3}{7} \end{bmatrix}.$$

The lattice generated by the columns of  $B$  shows the intersection between the subtorus and a lifted plane orthogonal to it (see Figure 4).

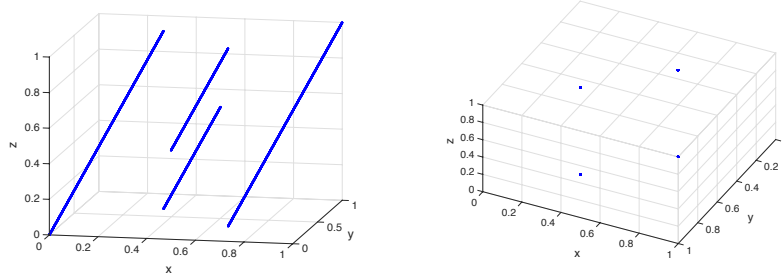


FIGURE 4. The geodesic in  $\mathbb{T}^3$  through the origin in direction  $d = [1, 2, 3]^T$ , seen from two different viewpoints. Viewing in direction  $d$  (right) shows the lattice formed in  $\mathbb{R}^3$  by the intersection of the lifted geodesic with an orthogonal plane.

**Example 5.** Let  $n = 2$ ,  $k = 1$ ,  $c = 0$ , and  $A = [2, 5]$ . A basis for the dual lattice is  $A^T(AA^T)^{-1} = [\frac{2}{29}; \frac{5}{29}]$ ; this vector is orthogonal to the tangent space of the geodesic and gives the spacing between its successive winds, which are spaced a distance  $1/\sqrt{29} \approx 0.19$  apart (see Figure 5). The fractional part of  $2x_1 + 5x_2$  measures the angular distance from a point  $x$  to the geodesic.

**5.3. Finding the best subtorus with given resonance relation.** We now consider the problem of computing the distance from a datapoint to a subtorus. Consider the example shown in Figure 4. To compute the Euclidean distance, it is necessary to (i) lift the datapoint to  $\mathbb{R}^3$ ; (ii) project to a plane orthogonal to the tangent space of the subtorus; and (iii) find the nearest point in the dual lattice  $L^*$ ; and (iv) compute the distance to this point. The difficult step is (iii), an instance of the Closest Vector Problem (CVP) in the dual lattice  $L^*$  [18, 25]. However, this is a difficult problem in high dimensions and the degree of complexity it entails seems unnecessary here, as this step can be avoided by modifying the metric suitably.

As we are working with angular distances, we replace the standard angular distance  $d(x, y) = 2\pi|x - y| \leq \pi$ ,  $x, y \in \mathbb{T}$ , by the chordal distance  $d_c(x, y) = \frac{1}{2} \sin \pi|x - y| \leq \frac{1}{2}$ . The intrinsic mean of angles  $x_i \in [0, 1)$  is easily calculated in the chordal metric as the circular mean (defined componentwise):

$$\bar{x} = \text{atan2} \left( \frac{1}{d} \sum_{i=1}^d \sin(2\pi x_i), \frac{1}{d} \sum_{i=1}^d \cos(2\pi x_i) \right). \quad (4)$$

We now introduce a further modification of the metric that is adapted to the chosen family of subtori.

**Definition 5.1.** Let  $C \in GL(n, \mathbb{Z})$ . Then  $d_C(x, y) := d_c(Cx, Cy)$ .

**Proposition 6.** Let  $A$  be the first  $k$  rows of  $C \in GL(n, \mathbb{Z})$ . Amongst the subtori with resonance relation  $A$ , the subtorus of best fit in the metric  $d_C$  to the data  $x_1, \dots, x_d \in \mathbb{T}^n$  is  $\{x \in \mathbb{T}^n : Ax = c\}$ , where  $c$  is the circular mean of  $Ax_1, \dots, Ax_d$ .

*Proof.* In coordinates  $y = Cx$ , the subtorus is given by

$$\hat{T} = \{y \in \mathbb{T}^n : y_i = c_i, i = 1, \dots, k\},$$

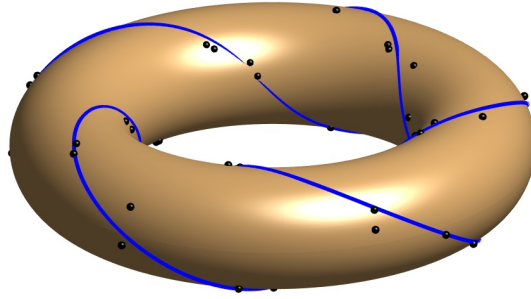
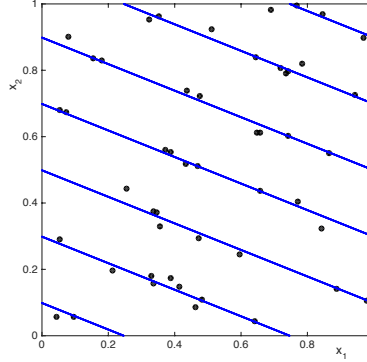


FIGURE 5. Fitting data on a torus. Here the closed geodesic of best fit is computed to a set of 50 data points on  $S^1 \times S^1$ . The data set is synthetic and has been chosen to lie near the geodesic with resonance relation  $2x_1 + 5x_2 = \text{const.}$ ; each data point has normal random noise of standard deviation  $0.1/(2\pi)$  in each angle.

and the distance from  $y$  to the subtorus  $T = \{x: Ax = c\}$  is determined by the angular displacement  $\hat{y} - c \in \mathbb{T}^k$  where  $\hat{y} = (y_1, \dots, y_k)$ . The distance  $d_C(x, T) = d_c(\hat{y}, c)$  is minimized at  $c = \hat{y}$ .  $\square$

Note that although  $d_C$  depends on the whole matrix  $C$ , the best subtorus only depends on its first  $k$  rows,  $A$ .

If the rows of  $A$  are pairwise orthogonal and all have the same length, then  $d_C(x, T) = d_c(x, T)$ , but in general the two metrics are not the same. Different bases  $A$  of  $L$  lead to different distance measures and different best tori. Most lattices have no orthogonal bases. This suggests choosing a basis for the resonance relations in which the relations are as nearly orthogonal as possible. This is another standard problem in lattice theory, one that can be solved exactly in low dimensions, and approximately (by the LLL algorithm [18]) in high dimensions.

**Example 3 (ctd.)** The angle between the two basis vectors in Example 1 is  $32^\circ$ . A more nearly orthogonal basis is

$$\begin{bmatrix} -1 & -2 \\ -1 & 1 \\ 1 & 0 \end{bmatrix}$$

in which the angle between the basis vectors is  $75^\circ$ .

The best fit of a 1-torus with fixed resonance relation to data in  $\mathbb{T}^2$  is shown in Figure 5, and the best fits of 1-tori and 2-tori to data in  $\mathbb{T}^3$  with fixed resonance relations is shown in Figure 6.

**5.4. Model selection for tori.** In finding the best subtorus amongst those with *fixed* resonance relations, the overall scaling of the metric is irrelevant. It becomes relevant during the model selection phase, when fitted subtori with different resonance relations are compared. Here we illustrate one possible approach to this issue using (i) the unscaled circular means (Eqn. (4)); and (ii) the ‘leave one out’ model selection method. Item (i) means that the maximum distance of any point to a subtorus, in each coordinate, is 0.5, regardless of the resonance relations or winding density of the subtorus. While the metric could be scaled down, to make it more closely approximate the original Riemannian distance in  $\mathbb{T}^n$ , doing so would strongly favour models with very dense windings, as they pass close to every point in  $\mathbb{T}^n$ . Therefore we stick with the unscaled metric  $d_C$  defined above. In this metric a geodesic with twice the length must have less than half the (Euclidean) residual in order to be considered a better fit.

Item (ii) means that for each data point  $x_i$ , the subtorus of best fit to the data set omitting point  $x_i$  is calculated, from which the prediction error of this fit to  $x_i$  can be calculated. In the scaled chordal metric we are using, this is  $e_i := \|\frac{1}{2} \sin \pi(Ax_i - c)\|_2$ . This lies in the interval  $[0, \frac{1}{2}\sqrt{k}]$ , taking the value 0 if the omitted data point lies on the geodesic, and taking the value  $\frac{1}{2}\sqrt{k}$  if it lies midway between two winds of the geodesic in each of the  $k$  directions. The mean prediction error  $\|e_i\|_2$  is taken as a measure of the goodness of fit of the model with resonance relations  $A$ . The resonance relation with minimum  $\|e_i\|_2$  is chosen.

The method is illustrated on a synthetic data set of 50 points that lie near the geodesic  $2x_1 + 5x_2 = \text{const.}$ ; see Figure 5. All resonance relations with  $\|A\|_\infty < 10$  are tested. The leave-one-out method selects the ‘correct’  $A = [2, 5]$  for this dataset.

**5.5. Nested approximations.** So far we have presented a method for finding the best subtorus of a given dimension. However, note that the same method naturally produces a nested sequence of approximations of subtori of different dimensions.

**Proposition 7.** *Let  $A \in GL(n, \mathbb{Z})$ , let  $A_k$  be the first  $k$  rows of  $A$ , and let  $x_1, \dots, x_d$  be data in  $\mathbb{T}^n$ . For each  $k = 1, 2, \dots, n-2$ , the  $(n-k)$ -dimensional subtorus with resonance relations  $A_k$  of best fit contains the  $(n-k-1)$ -dimensional subtorus of resonance relations  $A_{k+1}$  of best fit.*

*Proof.* The subtori are  $A_k x = c$ , where  $A_k$  is the first  $k$  rows of  $A$  and the entries in  $c \in \mathbb{R}^k$  are the circular means of  $Ax_i$ . Adding another resonance relation, i.e. increasing  $k$  by 1, does not change the first  $k$  entries of  $c$ .  $\square$

Thus to each  $A \in GL(n, \mathbb{Z})$  we get a nested sequence of subtori of dimension 1 to  $n-1$  and a residual associated to each subtorus. If the rows of  $A$  are nearly orthogonal, this is a close analogue of standard PCA.

**Example 3 (ctd.)** We take

$$A = \begin{bmatrix} -1 & -2 & 0 \\ -1 & 1 & 1 \\ 1 & 0 & -1 \end{bmatrix}.$$

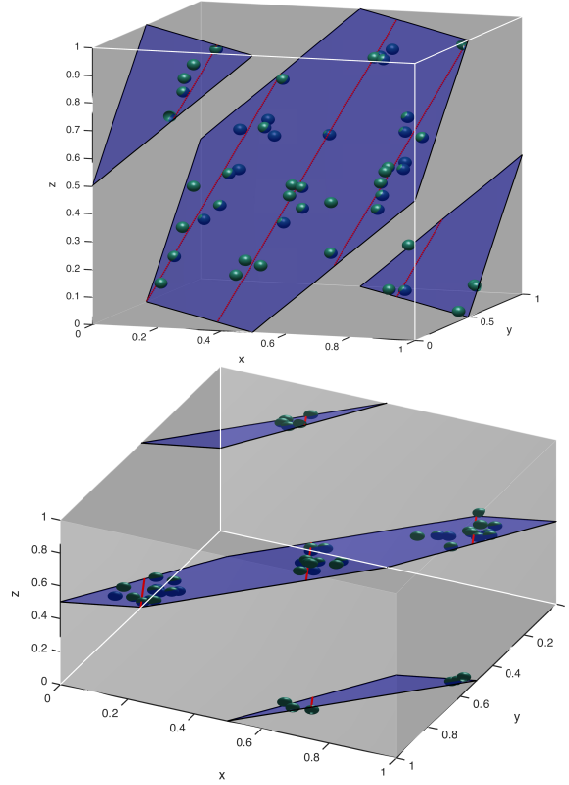


FIGURE 6. Fitting data on a torus (see Example 4). The best 1-torus and 2-torus approximating 50 data points on  $\mathbb{T}^3$ , are shown from two different viewing directions. The subtori have been chosen from those with fixed resonance relations, i.e., only their translations fitted.

We take a synthetic data set of 50 points on  $\mathbb{T}^3$  (see Figure 6). When  $k = 1$  we are seeking the best 2-torus of the form  $-x - y - z = \text{const.}$ ; it has mean residual 0.049. When  $k = 2$  we are seeking the best 1-torus of the form  $-x - y - z = \text{const.}$ ,  $-2x + y = \text{const.}$ , i.e., the best geodesic parallel to  $[1, 2, 3]$ . It has mean residual 0.049 orthogonal to the previously found 2-torus, i.e. in the direction  $[-1, -1, 1]$ , and mean residual 0.169 in the direction  $[-2, 1, 0]$ ; its mean residual is  $\sqrt{0.049^2 + 0.169^2} = 0.176$ . These residuals are scaled so that the distance between winds is 1, i.e., the distance between winds of the blue 2-torus is 1 and the distance, measured within the blue 2-torus, between winds of the red 1-torus is 1.

**6. Polyspheres.** The polyspheres  $S^2 \times \cdots \times S^2$  arise frequently in practical applications, for example in joint data. We begin by considering the case  $S^2 \times S^2$ , as the arguments are analogous in higher dimensions. We must classify the totally geodesic submanifolds of  $S^2 \times S^2$ , for which purpose we recall that the symmetric algebra of  $S^2$  is  $\mathfrak{o}(3) = \mathfrak{h} + \mathfrak{m}$ , where

$$\mathfrak{h} = \left\{ h(A) := \begin{pmatrix} 0 & 0 \\ 0 & A \end{pmatrix}, A \in \mathfrak{o}(2) \right\}, \quad \mathfrak{m} = \left\{ m(\xi) := \begin{pmatrix} 0 & -\xi^\top \\ \xi & 0 \end{pmatrix}, \xi \in \mathbb{R}^2 \right\}.$$

A straightforward calculation shows that  $[m(\xi), m(\zeta)] = h([\xi, \zeta])$ , where  $[\xi, \zeta] := \zeta\xi^T - \xi\zeta^T$ , and that  $[h(A), m(\xi)] = m(A\xi)$ , which will be used frequently in the coming calculations. The symmetric algebra of  $S^2 \times S^2$  is  $(\mathfrak{h} + \mathfrak{h}) + (\mathfrak{m} + \mathfrak{m})$ , and the totally geodesic submanifolds (and hence symmetric subspaces) take the form

$$\{\exp(v) \cdot x \mid v \in \mathfrak{m}'\},$$

where  $x \in S^2 \times S^2$  is an arbitrary point, and  $\mathfrak{m}' \subset \mathfrak{m} + \mathfrak{m}$  is a Lie triple system.

We now prove the classification results for totally geodesic submanifolds of polyspheres. We begin by stating some useful lemmas.

**Lemma 6.1.** *The two-dimensional vector subspaces of  $\mathbb{R}^2 \oplus \mathbb{R}^2$  take one of the forms (up to a reordering of basis elements):*

1.

$$\{(\xi, A\xi) \mid \xi \in \mathbb{R}^2\}, \quad A \in \text{Mat}(2 \times 2)$$

2.

$$\{(t_1\zeta_1, t_2\zeta_2) \mid t_i \in \mathbb{R}\}, \quad \zeta_i \in \mathbb{R}^2$$

*The three-dimensional vector subspaces of  $\mathbb{R}^2 \oplus \mathbb{R}^2$  take the form (up to a reordering of basis elements):*

$$\{(\xi, A\xi + t\zeta) \mid \xi \in \mathbb{R}^2, t \in \mathbb{R}\}, \quad \zeta \in \mathbb{R}^2, A \in \text{Mat}(2 \times 2)$$

*Proof.* Suppose the subspace  $V$  is spanned by two vectors,  $(u_1, v_1), (u_2, v_2)$ . We first consider the case where  $u_1$  and  $u_2$  are linearly independent. Then if  $(x, y) \in V$  we have  $x = au_1 + bu_2$ , where

$$\begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} u_1 & u_2 \end{pmatrix}^{-1} x,$$

and hence  $y = av_1 + bv_2$ , i.e.

$$y = \begin{pmatrix} v_1 & v_2 \end{pmatrix} \begin{pmatrix} u_1 & u_2 \end{pmatrix}^{-1} x,$$

and the subspace takes the form  $(x, Bx)$  for some matrix  $B$ . If the  $u_i$  are linearly dependent, but the  $v_i$  are linearly independent, the argument is similar. The remaining case to consider is where both the  $u_i$  and  $v_i$  are linearly dependent, in which case the subspace takes the form  $(t_1u, t_2v)$ , for some fixed vectors  $u$  and  $v$ .

Now suppose the subspace  $V$  is generated by  $(u_1, v_1), (u_2, v_2), (u_3, v_3)$ . Then either the  $(u_i)$  or  $(v_i)$  span  $\mathbb{R}^2$ , assume that  $(u_i)$  do. For any fixed  $t$ , if  $(x, y) \in V$  then  $x = au_1 + bu_2 + tu_3$  in a unique manner, indeed

$$\begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} u_1 & u_2 \end{pmatrix}^{-1} (x - tu_3).$$

Then

$$y = \begin{pmatrix} v_1 & v_2 \end{pmatrix} \begin{pmatrix} u_1 & u_2 \end{pmatrix}^{-1} x + t \left( v_3 - \begin{pmatrix} v_1 & v_2 \end{pmatrix} \begin{pmatrix} u_1 & u_2 \end{pmatrix}^{-1} u_3 \right)$$

and we see that  $V$  takes the form  $(x, Ax + tv)$  for some fixed  $A$  and  $v$ . □

We will also make use of the following elementary results.

**Lemma 6.2.** *Let  $A \in \mathfrak{so}(2)$  be non-zero, and let  $B$  be a  $2 \times 2$  matrix. Suppose that  $BAB^T Bz = BAz$  for all  $z \in \mathbb{R}^2$ . Then either  $B \in O(2)$  or  $B = 0$ .*

*Proof.* Firstly note that the condition  $BAB^T Bz = BAz$  for all  $z \in \mathbb{R}^2$  is equivalent to  $BAB^T B = BA$ , by for instance the bijective correspondence between linear transformations and matrices. Let  $C = B^T B$ ; multiplying both sides by  $B^T$  results in the relation  $CAC = CA$ . Note that  $C$  is symmetric, whilst  $A$  is antisymmetric, and hence  $CAC$  is also antisymmetric; as  $\mathfrak{so}(2)$  is one-dimensional we have  $CAC = kA$  for some  $k \in \mathbb{R}$ . This gives also  $CA = kA$ , and taking transposes  $-AC = -kA$ . Then  $CAC = C(AC) = kCA = k^2A$ . It follows that  $k^2A = kA$ , and hence  $k = 0$  or  $1$ . Moreover, as  $A$  is invertible, we conclude that  $C = kI$  and the result follows.  $\square$

**Lemma 6.3.** *Let  $V = \{m(\xi, A\xi) \mid \xi \in \mathbb{R}^2\}$ , where  $A \in O(2)$ . Then the submanifold  $\{\exp(v) \cdot x \mid v \in V\}$  takes the form  $\{(x, Rx) \in P\} \simeq S^2$ , for some  $R \in O(3)$*

*Proof.* Without loss of generality we consider a basis such that  $x = (e_z, e_z)$ , in which the second component of  $\exp(v) \cdot x$  is given by

$$\exp(m(A\xi)) \cdot e_z$$

for some  $A \in O(2)$ , where  $e_z = (0, 0, 1)^T$ . Note that  $m(A\xi) = Bm(\xi)B^T$ , where  $B \in O(3)$  takes the form

$$B = \begin{pmatrix} A & 0 \\ 0 & 1 \end{pmatrix}$$

Then

$$\exp(m(A\xi)) = \exp(Bm(\xi)B^T) = B \exp(m(\xi)) B^T,$$

and as  $B^T e_z = e_z$ , we have that

$$\exp(m(A\xi)) \cdot e_z = B \exp(m(\xi)) \cdot e_z,$$

as  $\{\exp(m(\xi)) \cdot e_z\}$  spans  $S^2$  the result follows immediately.  $\square$

**Lemma 6.4.** *Suppose that  $D \in O(2)$ , and that  $D^T xy^T = xy^T D$  for all  $x, y \in \mathbb{R}^2$ . Then  $D = I$ .*

*Proof.* Let  $C_i = x_i y_i^T$ , with  $x_1 = y_1 = (1, 0)^T, x_2 = y_2 = (0, 1)^T, x_3 = (1, 0)^T, y_3 = (0, 1)^T$ . Writing out

$$D = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

we see that the equations  $D^T C_i = C_i D$  give respectively  $b = 0, c = 0$  and  $a = d$ , from which the result follows.  $\square$

The totally geodesic submanifolds of  $S^2 \times S^2$  are classified in the following theorem.

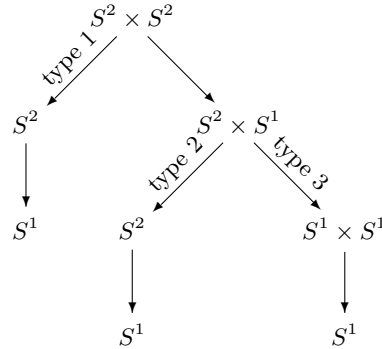
**Theorem 6.5.** *The 1-dimensional connected totally geodesic submanifolds of  $S^2 \times S^2$  are the submanifolds  $\{(\exp(r_1 t)x, \exp(r_2 t)y) : t \in \mathbb{R}\}$ , where  $r_i \in \mathfrak{o}(3)$  and  $x, y \in S^2$ .*

*The 2-dimensional connected totally geodesic submanifolds of  $S^2 \times S^2$  are of the following types:*

1.  $\{(x, Rx) : x \in S^2\} \simeq S^2$ , where  $R \in O(3)$  is fixed.
2.  $\{(x, y) : x \in S^2\} \simeq S^2$  where  $y \in S^2$  is fixed, or  $\{(x, y) : y \in S^2\} \simeq S^2$ , where  $x \in S^2$  is fixed.
3.  $\{(\exp(r_1 t_1)x, \exp(r_2 t_2)y) : t_1, t_2 \in \mathbb{R}\} \simeq S^1 \times S^1$ , where  $r_i \in \mathfrak{o}(3)$  and  $x, y \in S^2$  are fixed.

The 3-dimensional connected totally geodesic submanifolds of  $S^2 \times S^2$  are precisely the submanifolds  $\{(x, \exp(rt)y) : x \in S^2, t \in \mathbb{R}\} \simeq S^2 \times S^1$ , for some fixed  $y \in S^2$  and  $r \in \mathfrak{o}(3)$ .

The principal symmetric space decompositions of  $S^2 \times S^2$  can be summarized by the following diagram:



*Proof.* We must search for and exponentiate Lie triple systems  $\mathfrak{m}' \subset \mathfrak{m} + \mathfrak{m}$ ; these take the form  $m(V)$ , where  $V \subset \mathbb{R}^2 \oplus \mathbb{R}^2$  take the forms described in Lemma 6.1. Amongst the two-dimensional cases, we begin by those of the form  $V = \{(\xi, B\xi) \mid \xi \in \mathbb{R}^2\}$ ,  $B \in \text{Mat}(2 \times 2)$ , whereupon we compute

$$[m(x, Bx), m(y, By)], m(z, Bz)] = m([x, y]z, B[x, y]B^T Bz).$$

As  $[x, y] \in \mathfrak{o}(2)$ , Lemma 6.2 shows that we obtain a Lie triple system if either  $B^T B = I$ , or  $B = 0$ . We see by Lemma 6.3 that taking an orthogonal  $B$  results in a subspace of the first kind listed, whilst taking  $B = 0$  trivially results in the second kind. It remains to check subspaces  $\{(t_1 \zeta_1, t_2 \zeta_2) \mid t_i \in \mathbb{R}\}$ ,  $\zeta_i \in \mathbb{R}^2$ , but these are trivially totally geodesic, as then  $[\mathfrak{m}', \mathfrak{m}'] = 0$ . Exponentiating the resulting subspace gives the third case of the lemma.

We then consider the three-dimensional submanifolds, where  $V = (\xi, A\xi + t\zeta) \mid \xi \in \mathbb{R}^2, t \in \mathbb{R}\}$ ,  $\zeta \in \mathbb{R}^2, A \in \text{Mat}(2 \times 2)$ . We first compute

$$[m(Bx + sv), m(By + tv)] = h(B[x, y])B^T + v(tx^T - sy^T)B^T + B(sy - tx)v^T$$

It follows that

$$\begin{aligned} [[m(Bx + sv), m(By + tv)], m(Bz + rv)] = \\ m(B[x, y]B^T Bz + v(tx^T - sy^T)B^T Bz + B(sy - tx)v^T Bz \\ + rB[x, y]B^T v + rv(tx^T - sy^T)A^T v + rA(sy - tx)v^T v) \end{aligned}$$

To obtain a Lie triple system, we require that the right hand side equal  $m(B[x, y]z + pv)$ , for some scalar  $p$ , for all  $x, y, z$ . This is only possible if  $B = 0$ . Indeed, setting  $s, t = 0$  results in the equation

$$B[x, y]B^T Bz + rB[x, y]B^T v = B[x, y]z + pv,$$

fix  $x, y$  so that  $[x, y] = A$  is an arbitrary (non-zero) matrix in  $\mathfrak{o}(2)$ . We are left with two lines in  $\mathbb{R}^2$  (as we vary  $r, p$ ), it follows that we require  $v$  to be an eigenvector of  $BAB^T$ , however as  $BAB^T \in \mathfrak{o}(2)$  it has no real eigenvectors unless  $B = 0$ . The results then follows immediately upon exponentiating  $m(V)$ .  $\square$

**Theorem 6.6.** *The totally geodesic submanifolds of  $(S^2)^n$  are isomorphic to a product of copies of  $S^2$  and  $S^1$ . The rooted tree structure of the principal symmetric space approximations  $\mathcal{PSSA}(X, (S^2)^n)$  is characterized as follows: every edge of the tree corresponds to either*

1. *A reduction to an  $m$ -torus inside an  $n$ -torus ( $m < n$ )*
2. *A 2-dimensional reduction arising from a coupling of two spheres after situation 1 of Lemma 6.5.*
3. *One of the following one-dimensional reductions: the restriction to a great circle  $S^1 \subset S^2$ , or to a trivial submanifold  $x_0 \subset S^1$ .*

*Proof.* We begin by noting that the case  $S^2 \times S^2$  follows this pattern: the edge 1 is the two-dimensional reduction resulting from a coupling of spheres, the edge 2 comes from the inclusion  $x_0 \subset S^1$ , and the edge 3 is an inclusion  $S^1 \subset S^2$ . The remaining edges are clearly also of this pattern.

We now sketch a proof that the reductions of Lemma 6.5 are the only possible ones, even for higher polyspheres. The main point is that result and proof of Lemma 6.1 is generic, indeed the vector subspaces of  $\mathbb{R}^2 \oplus \dots \oplus \mathbb{R}^2$  take the form

$$\{(\xi_1, \dots, \xi_m, \dots, A_1^d \xi_1 + \dots + A_m^d \xi_m + t_d v_d) \mid \xi_1, \dots, \xi_m \in \mathbb{R}^2, t_1, \dots, t_d \in \mathbb{R}\}$$

The argument is essentially the same as before, roughly we proceed by letting  $\{u^1, \dots, u^k\}$  be a basis for the subspace, where we can write each vector  $u^i = (u_1^i, \dots, u_n^i)$ , each  $u_j^i \in \mathbb{R}^2$  being a 2d column vector. Form the block matrix  $U = [u]_i^j$ , a rank  $k$  matrix of size  $2n \times k$ . Form a  $k \times k$  submatrix of full rank by discarding rows. Consider a vector  $(x_1, \dots, x_n)$ ; the form of  $x_i$  depends on the discards in the rows of block  $i$ : Where no rows are discarded, we have a free  $\xi_i$ , if both rows are discarded we  $A_j^i \xi_j$  where we discarded both rows, we have in addition the  $t_j v_j$  where we discarded only one.

Consider then the possible Lie triple systems  $m(V)$ , where  $V$  takes the form above. Pick two terms from  $V$ ; these must reduce to one of the cases described in Lemma 6.5 if we set the other free terms  $\xi_j, t_j$  to zero. This proves that the  $A_j^i$  must be orthogonal or zero, and must be zero if paired with a  $tv$  term; it remains to show that for any given  $i$  only one  $A_j^i$  can be non-zero. For this purpose we compute

$$\begin{aligned} [[m(Ax_1 + By_1), m(Ax_2 + By_2)], m(Ax_3 + By_3)] = \\ \left( A[[x_1, x_2]]A^T + A(x_1 y_2^T - x_2 y_1^T)B^T \right. \\ \left. + B(x_2 y_1^T - x_1 y_2^T)A^T + B[[y_1, y_2]]B^T \right) (Ax_3 + By_3), \end{aligned}$$

which must equal  $A[[x_1, x_2]]x_3 + B[[y_1, y_2]]y_3$  if we are to obtain a Lie triple system. Following our argument we assume that  $A, B$  are both orthogonal. Now setting  $x_1 = -x_2 = x$ , and  $y_1 = y_2 = y$  results in

$$2(ACB^T - BCA^T)(Ax_3 + By_3) = 0,$$

where  $C = xy^T$ . As  $(Ax_3 + By_3)$  spans  $\mathbb{R}^2$  we are left with the relation  $ACB^T = BCA^T$ , and as  $A, B$  are orthogonal it follows that  $D^T C = CD$ , where  $D = A^T B$ . Lemma 6.4 then shows that  $D = I$  and hence  $A = B$ . The main equation becomes

$$A([x_1, x_2] + [y_1, y_2])(x_3 + y_3) = A([x_1, x_2]x_3 + [y_1, y_2]y_3),$$

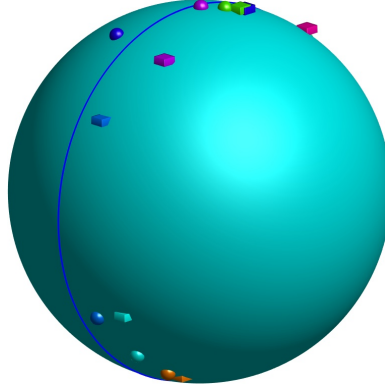


FIGURE 7. Fitting data on a polysphere (see Example 6). The rotations of points  $x_i$  are plotted as cubes, whilst the points  $y_i$  are plotted as circles; a different colour is chosen for each  $i$ . The great circle shows the best subspace  $S^1 \subset S^2 \subset S^2 \times S^2$ .

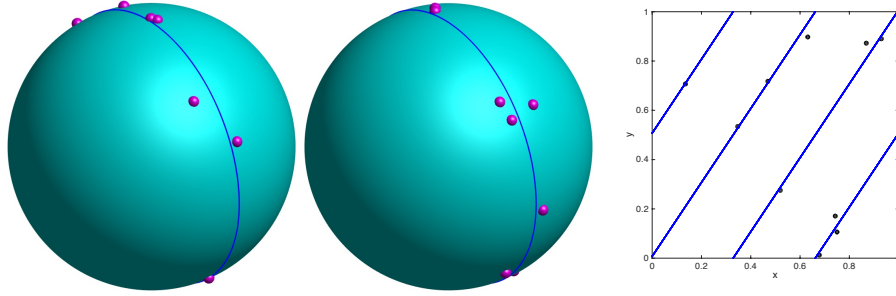


FIGURE 8. Fitting data on a polysphere (see Example 6). The best approximating torus  $S^1 \times S^1 \subset S^2 \times S^2$  is shown with two great circles inside the two spheres. Due to the difficulty of plotting the nested approximation  $S^1 \subset S^2 \times S^2$  we have plotted (right) the projection of the points in  $S^2 \times S^2$  to the approximating torus  $S^1 \times S^1$  and shown the best approximating  $S^1$  as a subset of this.

and hence

$$A([x_1, x_2]y_3 + [y_1, y_2]x_3) = 0,$$

which cannot hold for all  $x_i, y_i$  as  $A$  is invertible.  $\square$

The complexity of rooted tree arising from principal symmetric space approximations of polyspheres shows that a model selection problem cannot be avoided; see the remarks in the introduction and §5.4.

**Example 6.** We illustrate the behaviour with two synthetic datasets on  $S^2 \times S^2$ . The data of Figure 7 illustrates the middle branch of the tree, whilst the rightmost branch of the tree is shown in Figure 8.

**7. Discussion.** The principal symmetric space analysis (PSSA) proposed here has a number of appealing features.

1. The approximating submanifolds are geometrically the simplest possible, with zero exterior curvature and the highest available degree of symmetry;
2. they are global and do not rely on a reference point;
3. they are a special case of the submanifolds geodesic at a point used in PGA (but note that symmetric spaces can also contain submanifolds geodesic at a point that are *not* totally geodesic);
4. they can deal with significant curvature and diverse topology of both the ambient and the submanifold (e.g., tori);
5. in many cases the distance between a point and a submanifold can be computed analytically, without optimization or the numerical solution of differential equations; and
6. in many cases there exists a compatible metric that reduces the computation of the PSSA to linear algebra, and reduces or eliminates the branching of the tree of nested approximations.

On the other hand, its application is limited to data in symmetric spaces and to data sets that are structured in the proposed way. Exploring the latter condition in real data sets and in statistical models is an important area for future research. (Note that other techniques, such as Principal Nested Spheres [13] and geodesic splines, use approximating submanifolds that are not totally geodesic. Total geodesicity will not be suitable for all data sets.) On spheres and tori, the totally geodesic subspaces and the exponential barycentric subspaces coincide, which suggests studying the exterior geometry of the latter. Finally, we note that some manifolds have enormous numbers of types of totally geodesic submanifolds; on  $\mathbb{T}^n$  these are parameterized by  $SL(n, \mathbb{Z})$ . This complexity raises questions about model selection and effective computation, because for large  $n$  it is not practical to simply check each type separately.

**Acknowledgments.** RM was supported in part by the Royal Society Te Apārangi and would like to thank the Isaac Newton Institute for Mathematical Sciences, Cambridge, for support and hospitality during the programme Geometry, Compatibility, and Structure Grant number EP/R014604/1 where work on this paper was undertaken. CC was supported by the EU Horizon 2020 MSCA-RISE project Challenges in Preservation of Structure (CHiPS) in the form of a short term secondment to Massey University.

## REFERENCES

- [1] J. Berndt and C. Olmos, [Maximal totally geodesic submanifolds and index of symmetric spaces](#), *J. Differential Geom.*, **104** (2016), 187–217.
- [2] B. Y. Chen and T. Nagano, [Totally geodesic submanifolds of symmetric spaces, II](#), *Duke Math J.*, **45** (1978), 405–425.
- [3] J. H. Conway, R. H. Hardin and N. J. Sloane, [Packing lines, planes, etc.: Packings in Grassmannian spaces](#), *Experiment. Math.*, **5** (1996), 139–159.
- [4] J. Damon and J. S. Marron, [Backwards principal component analysis and principal nested relations](#), *J. Math. Imaging Vision*, **50** (2014), 107–114.
- [5] B. Eltzner, S. Jung and S. Huckermann, [Dimension reduction on polyspheres with application to skeletal representations](#), in *Geometric Science of Information*, Lecture Notes in Comput. Sci., 9389, Springer, Cham, 2015, 22–29.
- [6] P. T. Fletcher, C. Lu, S. M. Pizer and S. C. Joshi, [Principal geodesic analysis for the study of nonlinear statistics of shape](#), *IEEE Transactions on Medical Imaging*, **23** (2004), 995–1005.
- [7] P. T. Fletcher and S. C. Joshi, [Principal geodesic analysis on symmetric spaces: Statistics of diffusion tensors](#), in *Computer Vision and Mathematical Methods in Medical and Biomedical Image Analysis*, LNCS, 3117, Springer, Berlin, 2004, 87–98.

- [8] T. Fletcher, Geodesic regression on Riemannian manifolds, in *Proceedings of the Third International Workshop on Mathematical Foundations of Computational Anatomy-Geometrical and Statistical Methods for Modelling Biological Shape Variability*, 2001, 75–86.
- [9] C. G. Gebhardt, M. C. Steinbach and R. Rolfes, [Understanding the nonlinear dynamics of beam structures: A principal geodesic analysis approach](#), *Thin-Walled Structures*, **140** (2019), 357–372.
- [10] G. H. Golub and C. F. Van Loan, *Matrix Computations*, Johns Hopkins Series in the Mathematical Sciences, 3, Johns Hopkins University Press, Baltimore, MD, 1989.
- [11] S. Huckemann, T. Hotz and A. Munk, Intrinsic shape analysis: Geodesic PCA for Riemannian manifolds modulo isometric Lie group actions, *Statist. Sinica*, **20** (2010), 1–58.
- [12] I. Jolliffe, *Principal Component Analysis*, Springer Series in Statistics, Springer-Verlag, New York, 1986.
- [13] S. Jung, I. L. Dryden and J. S. Marron, [Analysis of principal nested spheres](#), *Biometrika*, **99** (2012), 551–568.
- [14] K. Kenobi, I. L. Dryden and H. Le, [Shape curves and geodesic modelling](#), *Biometrika*, **97** (2010), 567–584.
- [15] S. Klein, [Totally geodesic submanifolds in Riemannian symmetric spaces](#), in *Differential Geometry*, World Sci. Publ., Hackensack, NJ, 2009, 136–145.
- [16] S. Kobayashi and K. Nomizu, *Foundations of Differential Geometry*, John Wiley & Sons, Inc., New York, 1996.
- [17] J. Lawrence, [Enumeration in torus arrangements](#), *European J. Combin.*, **32** (2011), 870–881.
- [18] A. K. Lenstra, H. W. Lenstra Jr. and L. Lovász, [Factoring polynomials with rational coefficients](#), *Math. Ann.*, **261** (1982), 515–534.
- [19] X. Pennec, [Barycentric subspace analysis on manifolds](#), *Ann. Statist.*, **46** (2018), 2711–2746.
- [20] Q. Rentmeesters, [A gradient method for geodesic data fitting on some symmetric Riemannian manifolds](#), in *2011 50<sup>th</sup> IEEE Conference on Decision and Control and European Control Conference*, 2011, 7141–7146.
- [21] H. J. S. Smith, On systems of linear indeterminate equations and congruences, *Phil. Trans. Roy. Soc. London*, **151** (1861), 293–326.
- [22] S. Sommer, F. Lauze and M. Nielsen, [Optimization over geodesics for exact principal geodesic analysis](#), *Adv. Comput. Math.*, **40** (2014), 283–313.
- [23] W. Utschick, [Tracking of signal subspace projectors](#), *IEEE Trans. Signal Process.*, **50** (2002), 769–778.
- [24] J. A. Wolf, [Elliptic spaces in Grassmann manifolds](#), *Illinois J. Math.*, **7** (1963), 447–462.
- [25] D. Wubben, D. Seethaler, J. Jalden and G. Matz, [Lattice reduction](#), *IEEE Signal Processing Magazine*, **28** (2011), 70–91.

Received April 2019; revised October 2019.

E-mail address: [charles.curry@ntnu.no](mailto:charles.curry@ntnu.no)

E-mail address: [stephen.marsland@vuw.ac.nz](mailto:stephen.marsland@vuw.ac.nz)

E-mail address: [r.mclachlan@massey.ac.nz](mailto:r.mclachlan@massey.ac.nz)