



Walking on Kendall's Shape Space: Understanding Shape Spaces and Their Coordinate Systems

Christian Peter Klingenberg¹

Received: 30 May 2020 / Accepted: 10 August 2020 / Published online: 18 August 2020
© The Author(s) 2020

Abstract

More and more analyses of biological shapes are using the techniques of geometric morphometrics based on configurations of landmarks in two or three dimensions. A fundamental concept at the core of these analyses is Kendall's shape space and local approximations to it by shape tangent spaces. Kendall's shape space is complex because it is a curved surface and, for configurations with more than three landmarks, multidimensional. This paper uses the shape space for triangles, which is the surface of a sphere, to explore and visualize some properties of shape spaces and the respective tangent spaces. Considerations about the dimensionality of shape spaces are an important step in understanding them, and can offer a coordinate system that can translate between positions in the shape space and the corresponding landmark configurations and vice versa. By simulation studies “walking” along that are great circles around the shape space, each of them corresponding to the repeated application of a particular shape change, it is possible to grasp intuitively why shape spaces are curved and closed surfaces. From these considerations and the available information on shape spaces for configurations with more than three landmarks, the conclusion emerges that the approach using a tangent space approximation in general is valid for biological datasets. The quality of approximation depends on the scale of variation in the data, but existing analyses suggest this should be satisfactory to excellent in most empirical datasets.

Keywords Geometric morphometrics · Kendall's shape space · Procrustes distance · Procrustes superimposition · Shape coordinates · Tangent space

Introduction

Geometric morphometrics is widely used to analyze biological shapes in a variety of research contexts, including systematics and evolutionary biology, anthropology, the biomedical sciences, and increasingly developmental biology and genetics (Klingenberg 2010; Zelditch et al. 2012). Recent developments in morphometrics and the statistics of shape have seen an increasing emphasis on explicit characterization of the multidimensional spaces that underlie geometric morphometric analyses (Small 1996; Kendall et al. 1999; Dryden and Mardia 2016; Srivastava and Klassen 2016). Shape spaces are abstract representations in which each point represents a specific shape, and in which every

possible shape is represented by a particular point. Distances between points in a shape space correspond to the magnitude of the difference between the respective shapes. Shape spaces are specific to particular classes of shapes, for instance, all possible shapes for a given number of landmarks in two or three dimensions (Small 1996; Kendall et al. 1999; Dryden and Mardia 2016; Srivastava and Klassen 2016). The overall structure of a shape space reflects the entire possible range of variation in the corresponding class of shapes, for instance, all possible relative arrangements of a given number of landmarks. The concept of shape spaces gives a firm mathematical basis to statistical maneuvers such as estimating average shapes or characterizing variation of shapes around those averages, which are otherwise not well understood (Small 1996; Kendall et al. 1999; Dryden and Mardia 2016; Srivastava and Klassen 2016). Yet, because statistical maneuvers like these are fundamental to the biological applications of geometric morphometrics, an understanding of shape spaces is critically important for biological shape analysis.

✉ Christian Peter Klingenberg
cpk@manchester.ac.uk

¹ School of Biological Sciences, University of Manchester, Michael Smith Building, Oxford Road, Manchester M13 9PT, UK

Shape spaces tend to be complex, multidimensional, non-linear spaces. It is therefore very difficult to gain an intuitive understanding of these spaces. Abstract mathematical reasoning provides rigorous and powerful inferences about shape spaces, but it tends to be very hard to grasp for biologists. Some attempts have been made to provide graphical visualizations of simple shape spaces such as Kendall's shape space for triangles, which happens to be the surface of a sphere (Small 1996; Kendall et al. 1999; Rohlf 2000; Slice 2001; Dryden and Mardia 2016; Klingenberg 2016). Shape spaces for configurations with more than three landmarks are considerably more complex and difficult to visualize. An important aspect of understanding shape spaces is the coordinate system in which a shape space is situated and the way it relates to the original data such as landmark coordinates. These considerations, in turn, relate to the procedure of the Procrustes superimposition and the associated projections, which are used to fit empirical data to shape spaces and the associated tangent spaces (Small 1996; Kendall et al. 1999; Rohlf 1999; Dryden and Mardia 2016).

This paper aims to provide a more intuitive appreciation for shape spaces by extensively using graphical visualizations of Kendall's shape space for triangles. By explicitly enumerating the shape and non-shape components of the coordinate system of landmark data, it also becomes clear how the Procrustes superimposition and shape tangent space relate to the landmark coordinates. Throughout the paper, the focus is on the shape analysis of triangles in two dimensions, but I also examine how the insights gained from triangles apply to more complex configurations of landmarks.

Shape Distances and Shape Spaces

The shape of a landmark configuration (or of another object) is defined as all its geometric features, except for its size, position and orientation (Dryden and Mardia 2016). Following this definition, there is a separation between aspects of a landmark configuration that are part of shape and other aspects that are not part of shape. The non-shape components of size, position and orientation are easy to understand. What constitutes shape is harder to grasp, but proportions, angles and the relative arrangement of parts are geometric features that belong into this category. Shape distances and shape spaces must reflect the same distinction and include only those aspects of landmark configurations that are part of shape. The shape distance between two configurations differing only in size, position and orientation must be zero because they have the same shape. Therefore, those two configurations will also fall onto the same point in shape space.

Shape spaces are multidimensional spaces. The number of dimensions relates to how many different ways there are to make a small change to any given shape. The dimensionality

of any particular shape space depends on the number of landmarks and whether 2D or 3D data are used. It can be calculated as the number of landmark coordinates minus the number of non-shape components, which are features of variation that are not part of the shape space. For 2D data, each landmark has two coordinates, so that the total number of coordinates is twice the number of landmarks. From this, one dimension must be subtracted for size, two for translation, and one for rotation. Therefore, the shape dimensions for 2D data is twice the number of landmarks minus four. For 3D data, each landmark has three coordinates, and one dimension has to be removed for size, three for translation, and three for rotation (because there are separate rotations around the three coordinate axes, corresponding to the nautical terms of pitch, roll and yaw). Accordingly, the dimensionality of the shape space for 3D data is three times the number of landmarks minus seven. As a consequence, shape dimension is high for configurations with more than three landmarks and the corresponding shape spaces are impossible to visualize directly.

In principle, it would be possible to define a wide range of different shape distances, resulting in different shape spaces (Rohlf 2000). In practice, however, the shape distance that is used by far the most widely is Procrustes distance. It is based on a procedure for separating shape from non-shape aspects of variation in configurations of landmarks, called Procrustes superimposition (Dryden and Mardia 2016). This section discusses this procedure as well as the shape distance and the shape spaces that follow from it.

Ordinary Procrustes Superimposition: Comparing Two Shapes

The basic idea for Procrustes superimposition was first presented by Boas (1905) and a detailed description of the algorithm was provided by Sneath (1967) and subsequently by many other authors (Rohlf and Slice 1990; Goodall 1991; Kent 1994; Bookstein 1996; Rohlf 1999; Zelditch et al. 2012; Dryden and Mardia 2016; Klingenberg 2016). The task is to compare the shapes of two landmark configurations and to quantify the shape difference between them. The method does this by changing the non-shape features of one configuration to fit it as closely as possible to the other configuration, which is often called the target configuration.

The procedure usually starts by scaling both configurations to unit size (Fig. 1a). The size measure most widely used for this purpose is centroid size, the square root of the sum of squared distances of each landmark from the centroid of all landmarks (Dryden and Mardia 2016). The x and y coordinates of the centroid (also called the center of gravity) are obtained by averaging, respectively, the x and y coordinates of all the landmarks. Centroid size is best understood as a measure of spread of the landmarks around the center of

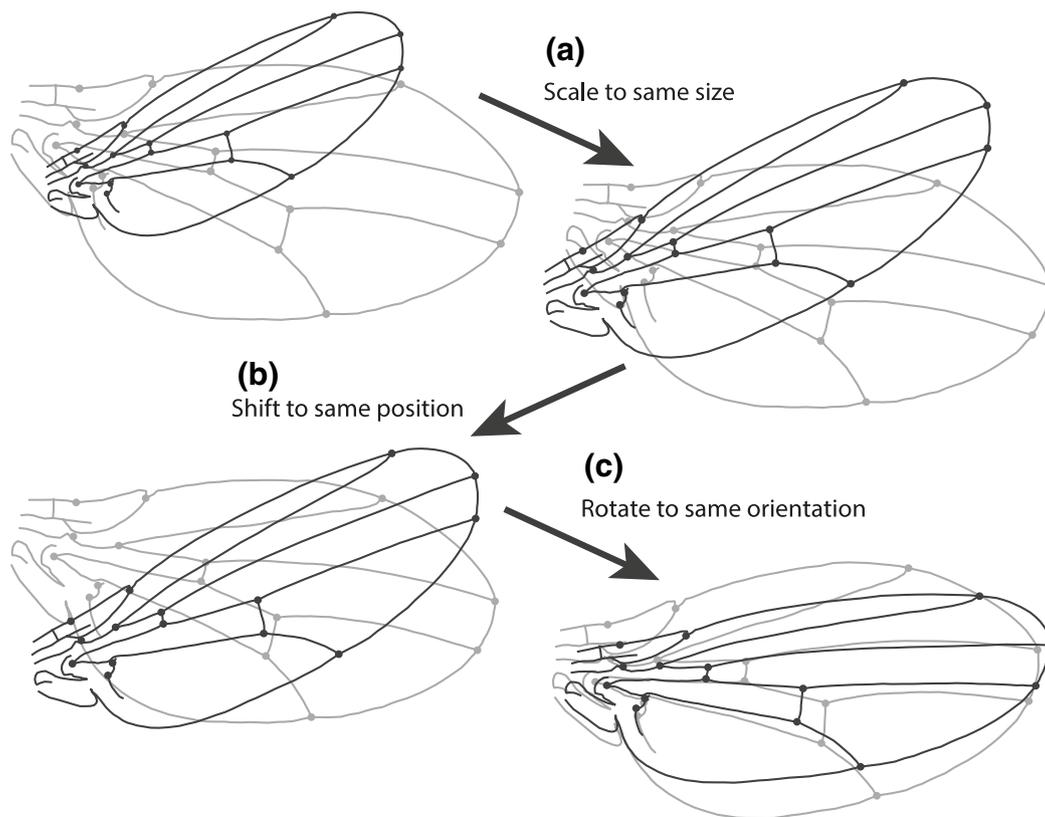


Fig. 1 Ordinary Procrustes superimposition of two landmark configurations (from Klingenberg 2015). The superimposition, here shown for two fly wings with 15 landmarks each, proceeds in three steps. **a** First, both configurations are scaled to a standard centroid size of 1.0.

b Second, both landmark configurations are shifted so that their centroids are at the origin of the coordinate system. **c** Finally, one of the configurations is rotated so that the positions of its landmarks match those in the target shape as closely as possible

the configuration. A landmark configuration is scaled to unit centroid size by dividing all the landmark coordinates by the centroid size of that configuration; after this step, the centroid size of the scaled configuration is 1.0. By scaling both landmark configurations to centroid size 1.0 in this manner, the procedure eliminates any size differences between them.

As the second step, differences in position are removed by shifting the scaled landmark configurations so that their centroids are in the same point (Fig. 1b). This can be achieved simply by subtracting the x and y coordinates of the centroid from the x and y coordinates of each landmark of the respective configuration. After this step, the two configurations are in the same overall position.

In the third and final step, one configuration is rotated around the common centroid until its landmarks are as closely matched to those of the target configuration as it is possible (Fig. 1c). After this rotation, the two landmark configurations have the same orientation. The criterion for the matching of landmarks is the sum of the squared distances between corresponding landmarks in the two configurations. The choice of this criterion also explains the matching of positions by centroids in the second step of the procedure.

These three steps successively remove the non-shape components of variation from the landmark configurations. Therefore, only shape variation remains. Because none of these steps changes shape, it also follows that the complete shape variation remains. Therefore, the coordinates of superimposed landmark configurations can be used for further analyses of shape.

The procedure outlined above, where both landmark configurations are scaled to unit centroid size, has been called the partial Procrustes superimposition (Dryden and Mardia 2016). This method can be modified slightly because the sum of squared distances between corresponding landmarks can be further reduced, and therefore the fit between configurations further improved, by scaling the configuration that is matched to the target to a slightly smaller centroid size (Fig. 2). The scaling factor depends on the magnitude of the shape difference between the landmark configurations (as can be seen in Fig. 2, the minimum is achieved for point F, where the line between the target T and fitted configuration F is perpendicular to the line from the origin O to the fitted configuration before rescaling,

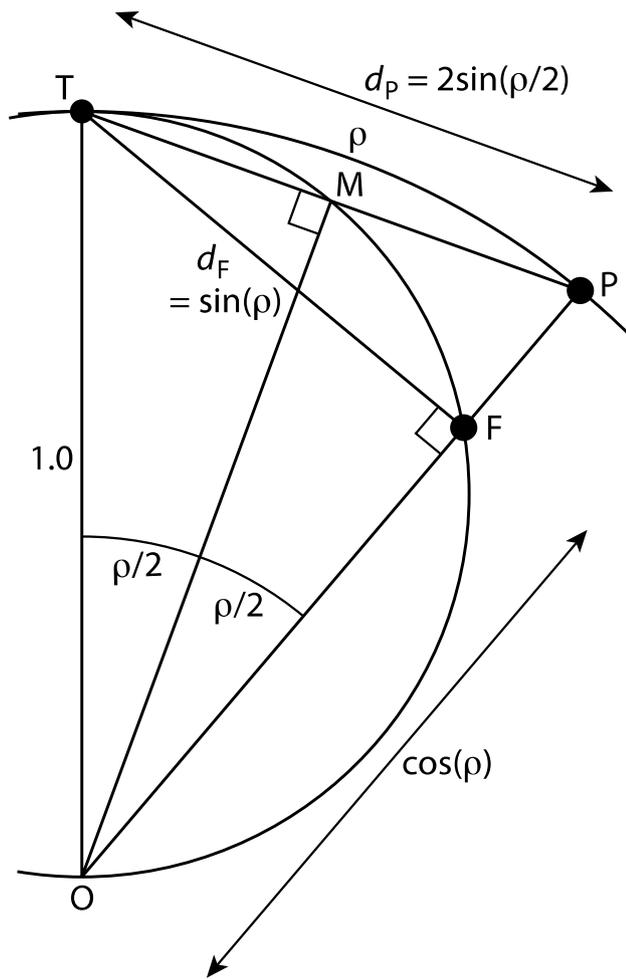


Fig. 2 Relationships between full and partial Procrustes distance and Riemannian distance between two triangle shapes (modified after Dryden and Mardia 2016). The point O is the origin of the coordinate system and T is the target shape. The part-circle through the points T, F, and O is a cross-section through Kendall's shape space, and the arc through points T and P is a part of a section through the space of landmark configurations aligned to shape T by partial Procrustes fit. Points P and F represent the shape fitted to T by partial and full Procrustes superimposition, respectively. They differ only in scaling: P is scaled to centroid size 1.0 (therefore it is on a circle of radius 1.0 around O, along with point T), whereas F is scaled to centroid size $\cos(\rho)$

P). This method, where only the target configuration is scaled to unit centroid size, but the other configuration is scaled so that it matches the target as closely as possible, is called a full Procrustes superimposition (Dryden and Mardia 2016). The scaling step is the only difference between the full and partial Procrustes superimposition; the translation and rotation steps are exactly the same (Dryden and Mardia 2016). In practice, the difference between these two variants of Procrustes superimposition is usually very small for biological data.

Full and Partial Procrustes Distance

In conjunction with Procrustes superimposition, it seems the most straightforward measure of shape difference between two configurations would be the sum of the squared distances between corresponding landmarks of the superimposed configurations. This is the criterion used for the superimposition, and therefore it reflects exactly those differences that cannot be removed by scaling, translating and rotating one configuration against the other. Because this is a sum of squared distances, however, it is not a direct measure of distance. Accordingly, the square root of the sum of squared distances between corresponding landmarks is used instead, and this measure of shape distance is called Procrustes distance (Dryden and Mardia 2016).

Depending on whether partial or full Procrustes superimposition has been used, the resulting Procrustes distances are different and are distinguished from each other by using different terms: full Procrustes distance (d_F) and partial Procrustes distance (d_P). Between shapes that are not identical, the full Procrustes distance is bound to be smaller than the partial Procrustes distance (Fig. 2), but this difference is usually small for biological data. For extreme shape differences, however, this difference can become noticeable, as can be appreciated from the theoretical maximum for the two distance measures: full Procrustes distance can range up to 1.0, whereas partial Procrustes distance can range up to the square root of 2, which is approximately 1.41 (Dryden and Mardia 2016).

Because all triangle shapes standardized to unit centroid size are bound to lie on an arc of radius 1.0 (Fig. 2), we can define yet another measure of distance between the shapes: the great circle distance (or geodesic distance) measured on that arc (the conventional symbol for this distance is ρ). This type of shape distance is known under several different names: Riemannian distance (Dryden and Mardia 2016), Procrustes distance (without any qualifiers such as "full" or "partial"; Dryden and Mardia 1998; Rohlf 1999), and often just its conventional symbol is used (Kendall 1984; Rohlf 1999; Dryden and Mardia 2016). This shape distance can range up to $\pi/2$, which is approximately 1.57 (Dryden and Mardia 2016). This measure of shape distance is particularly important in theoretical considerations.

The relationship between the different shape distances is most easily visible in Fig. 2. For a partial Procrustes superimposition of a shape (P) onto a target (T), both landmark configurations are scaled to centroid size 1.0, from which it follows that both these points are at a distance of 1.0 from the origin of the coordinate system (O). The connection between partial Procrustes distance and the Riemannian distance ρ is established via the two right triangles O–T–M and O–P–M, where M is the midpoint between P and T. For each of these triangles, the angle at the vertex O is $\rho/2$, and

because the distances between O and both P and T are 1.0, it follows that the distances T–M and P–M are both $\sin(\varrho / 2)$, so that the partial Procrustes distance between P and T is $2 \sin(\varrho / 2)$. For the full Procrustes superimposition, the shape is scaled so that the line from the shape being fitted to the target (F–T) is at a right angle to the line to the origin (F–O). By considering the right triangle O–F–T and the definitions of trigonometric functions, it follows that the full Procrustes distance is $\sin(\varrho)$ and also that the optimal scaling factor is $\cos(\varrho)$. It also is clear from Fig. 2 that $\varrho \geq d_p \geq d_F$, with equality holding if $\varrho = 0$. All of the relationships among the shape distances are fairly simple and completely deterministic, which means that it is quite easy to convert a shape distance from one of these measures to a different one.

Procrustes distances depend on the number and arrangement of landmarks in the configurations. These relationships are relatively complex, because the number of landmarks counts both for the number of squared distances that are summed up and for the number of landmarks that go into the calculation of centroid size for the size standardization. In practice, it is therefore preferable to compare Procrustes distances only within the context of a single analysis, but not between different analyses.

Kendall's Shape Space

To obtain a geometric representation of all possible shapes with a particular number of landmarks, it is possible to build a shape space. Every point in a shape space represents a shape and the distances between points represent the magnitudes of shape differences between the respective shapes. Regardless of whether we use partial or full Procrustes distance for this purpose, the result is a curved multidimensional space. Accordingly, shape spaces are rather complex.

Because shape spaces are nonlinear, they have some properties that may appear strange and not even counting dimensions is quite intuitive (in shape analysis, the topological dimension is used for shape spaces, not geometric dimension). For Euclidean (linear) spaces, the dimensionality corresponds to the number of coordinates we need to specify a point in the space or in how many orthogonal directions points can move in the space. Let's consider a circle, probably the simplest nonlinear space, and ask how many dimensions it has. It is tempting to see the Cartesian x and y coordinates of the plane in which the circle is embedded, suggesting there are two dimensions. But there is a strong relation between the x and y coordinates (the squared deviations from the center always add up to the squared radius of the circle) and every point on the circle can only slide in one direction (along the circle, but not in a radial direction). Also, it is sufficient to give a single angle to specify the position of a point on the circle (e.g., that it is in a 3 o'clock position). Therefore, from this perspective, it clearly appears

that a circle has only a single dimension. Similarly, the surface of a sphere has two dimensions (a longitude and latitude are sufficient to specify a point on a globe). This becomes particularly relevant when considering shape spaces for different numbers of landmarks or for 2D versus 3D data.

First, I focus on the case of triangles, landmark configurations with just three landmarks in two dimensions, which results in a relatively simple shape space that can be visualized directly. In this case, the shape space is two-dimensional: there are three landmarks with two coordinates each, and we need to subtract four dimensions for the degrees of freedom lost due to the standardization for size, position, and orientation, so that two dimensions remain.

Imagine we have a large sample of triangles of all possible shapes. Using the ordinary Procrustes superimposition introduced above, it is then possible to calculate a table with all pairwise distances between triangle shapes. From this table of distances, we can try to construct geometric representation where points represent the triangles, and the distances between points correspond to distances between the respective triangles. It is possible to do this by principal coordinate analysis (Gower 1966) or, equivalently, metric multidimensional scaling (Mardia et al. 1979). If full Procrustes distances between the triangle shapes are used, the points representing the triangles all lie exactly on the surface of a sphere of radius 0.5. This sphere is Kendall's shape space for triangles in two dimensions (Kendall 1984; Kendall et al. 1999; Rohlf 1999; Dryden and Mardia 2016). In practice, using principal coordinate analysis for this purpose is not very convenient because the direction in which the shape space is aligned against the resulting coordinate system is more or less arbitrary, depending entirely on the distribution of triangle shapes in the sample. For this reason, I present some considerations below that lead to a more convenient coordinate system, one that facilitates understanding the geometric aspects of shape variation.

If partial Procrustes distances are used to construct the geometric representation, no consistent representation results for the complete table of pairwise shape distances. If only an incomplete table is used, with only the pairwise distances among similar triangles included but without the distances between very different triangles, the geometric representation of triangles again is similar to Kendall's shape space. A similar problem applies for Riemannian shape distance. There is a degree of distortion in either of these cases, so that the principal coordinate analysis is not able to provide an exact solution—how much distortion there is depends on the maximum pairwise distances among shapes included in the analysis.

The arrangement of triangle shapes on Kendall's shape space is very specific and can be characterized by comparison to a globe (Fig. 3). If an equilateral triangle is chosen as the north pole, then a number of regularities emerge. The

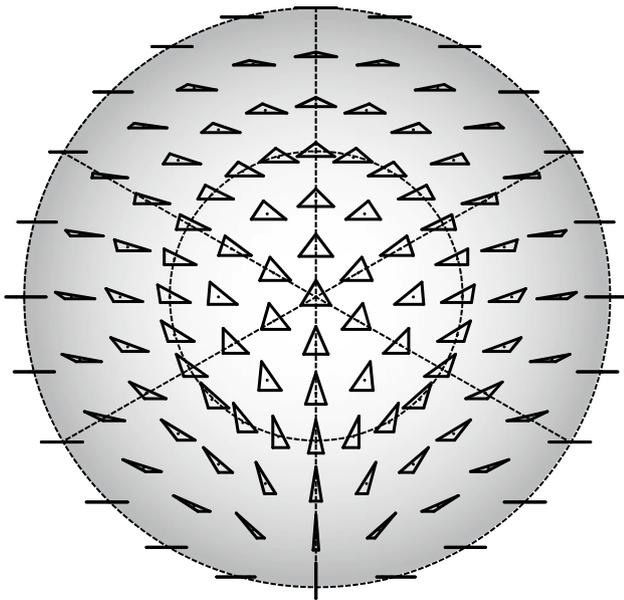


Fig. 3 Kendall's shape space for triangles in two dimensions (Klingenberg 2016). The diagram shows the view from the "north pole" corresponding to an equilateral triangle (center) onto one hemisphere. The outer dashed circle corresponds to the equator and contains the collinear triangles, where all three vertices lie on a straight line. The six meridians shown as straight dashed lines contain isosceles triangles

south pole corresponds to another equilateral triangle that is the mirror image to the one corresponding to the north pole (for one, the vertices are labelled clockwise and for the other, counter-clockwise). The equator contains collinear triangles: triangles where all three vertices lie on a single straight line (this circle also corresponds to the shape space for triangles in a single dimension; Small 1996; Kendall et al. 1999). Finally, six meridians correspond to isosceles triangles (Fig. 3). For each triangle shape on the northern hemisphere, the mirror image shape is located at the same longitude and corresponding latitude on the southern hemisphere.

Tangent Space

Because Kendall's shape space is a non-linear space, computations in it are rather cumbersome and most studies of shape variation therefore use only a local approximation of the nonlinear space by a flat, linear tangent space (Fig. 4). Shape tangent spaces have the same dimensionality as the shape spaces they approximate. In the case of triangles, therefore, each tangent space has two dimensions and is a plane. This tangent approximation is essentially the same as the approximation of the curved surface of the Earth by a flat map. The tangent space can provide a close linear approximation of a limited region of the shape space, but if there is a relatively large range of different shapes,

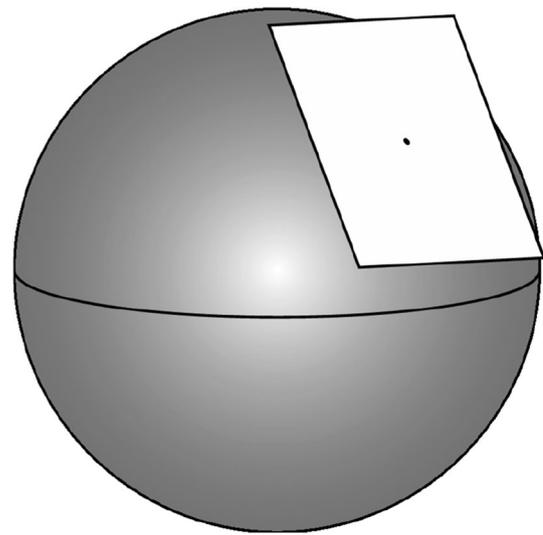


Fig. 4 The tangent approximation to Kendall's shape space for triangles (Klingenberg 2016). A flat space is found that touches the shape space at a reference point, such as the mean shape in a dataset, and provides a linear approximation to the shape space in the surroundings of this point

distortions are inevitable. Due to the same effect, world maps usually show Greenland and Antarctica with grossly distorted shapes, because the map projection is usually optimized for the equator and produces distortions for high latitudes.

Choosing the tangent point, where the tangent space touches the shape space, is an important step in any analysis using a tangent space. The mean shape in a sample is usually a good choice for the tangent point because it is usually near the center of the distribution, so that distortion tends to be relatively small in all directions, and no observations in the sample are expected to suffer disproportionate distortions. Note that the tangent projection also entails a slight expansion of centroid size, which in turn depends on the magnitude of the shape difference from the shape at the tangent point (the points F' and P' in Fig. 5 are slightly farther from point O than F and P , respectively).

Usually an orthogonal projection is used to project observations into the tangent space (Dryden and Mardia 2016). This means that observations are projected in a direction that is perpendicular to the shape tangent space. This direction is also parallel to the vector from the origin to the tangent point, which is usually the average shape in sample. An alternative method is stereographic projection, where each observation is projected along a straight line from the origin (Rohlf 1999), but this method leads to worse distortions for shapes that are relatively far from the tangent point.

Generalized Procrustes Superimposition: More than Two Shapes

The preceding discussion has mentioned average shapes in samples of landmark configurations, but has not provided an explanation. Therefore, this section presents the generalized Procrustes superimposition (or generalized Procrustes analysis, GPA), which is the method used for computing such average shapes in samples of landmark configurations, (Gower 1975; Rohlf and Slice 1990; Goodall 1991; Dryden and Mardia 2016). I particularly emphasize how this procedure relates to shape spaces and shape distances. Contrary to most previous explanations of this method, it is most helpful to view generalized Procrustes superimposition not primarily as a procedure for aligning landmark configurations relative to one another, but as a method for fitting empirical data to Kendall's shape space.

Generalized Procrustes superimposition is an iterative procedure that starts with a sample of landmark configurations. In the first round, one of these configurations is chosen arbitrarily as the target, for instance the first one in the dataset, and every other landmark configuration is fitted to that target configuration in turn by an ordinary Procrustes superimposition as described above. When this is complete, a consensus configuration is computed by averaging the landmark coordinates across all the aligned configurations (including the configuration chosen as the target for the first round) and rescaling the resulting configuration to unit centroid size. This consensus is then taken as the new target configuration and all the landmark configurations in the sample are fitted to it in a new round of ordinary Procrustes superimpositions. The coordinates of the aligned landmark configurations are then averaged again, this time without the target configuration (which was the consensus from the previous round, not one of the landmark configurations from the input data). The resulting average is used as a target configuration in the next round of ordinary Procrustes superimpositions, and this procedure is repeated until the changes between successive rounds become negligible. This is usually the case after just two to three rounds—the algorithm tends to converge fast for most datasets.

This procedure can use either the full or the partial Procrustes superimposition procedure as described above. Therefore, there are two slightly different methods, full GPA and partial GPA, that produce slightly different estimates of the mean shape in a sample. In full GPA, the procedure of rescaling size is used at every step of the iterative fitting by ordinary Procrustes superposition. The rescaling is done as above, as a function of the shape difference between the target and each configuration that is being fitted to it (using the factor $\cos(\varrho)$; Fig. 2). As a consequence of this, configurations with a shape far from the target tend to be weighted less than shapes closer to the target in the calculation of

the average shape. In partial GPA, by contrast, every landmark configuration is included with the unit centroid size and is therefore weighted equally. As a result, the two GPA methods may have slightly different behavior, particularly concerning the treatment of outliers.

When the GPA algorithm converges, the consensus configuration provides an estimate of the mean shape and all the landmark configurations in the sample are aligned optimally to that consensus shape in the 2D or 3D space of the original landmark configurations. From the perspective of multivariate statistics, however, we can also think of a different, multidimensional space in which the variables define the coordinate axes and thus the number of dimensions (note that there may or may not be variation in every direction in this space). In the context of shape analysis, the landmark coordinates serve as the variables, and we therefore can envision a multivariate space in which every coordinate of every landmark defines a coordinate axis and thus a separate dimension. The dimensionality of this space is twice the number of landmarks for 2D data and three times the number of landmarks for 3D data. To calculate Euclidean distances in multidimensional spaces, we can use the multivariate generalization of the Pythagorean theorem: the distance between two points in the space is the square root of the sum of squared differences between the two points along every coordinate axis, summed up over all the coordinate axes that make up the space. If we apply this calculation to the multivariate space defined by the landmark coordinates, it turns out that it is exactly the same as in the definition of Procrustes distance given above. Accordingly, the distance of every aligned configuration from the average shape in this multivariate space is precisely the Procrustes distance—full Procrustes distance if a full GPA was used or, alternatively, partial Procrustes distance if partial GPA was used. The pairwise distances among points in the space corresponding to the different configurations in the sample approximate the Procrustes distances among the respective shapes (as long as shape variation is small, i.e., as long as the shapes in the dataset only occupy a relatively small patch of Kendall's shape space). Accordingly, it emerges that the arrangement of the points representing the shapes in the multidimensional space approximates a limited portion of Kendall's shape space: there are single points representing landmark configurations in the multidimensional space and the pairwise distances between them at least approximate Procrustes distances between the respective shapes.

To understand the situation when shape variation is greater, that is, when the shapes in the sample are not all similar to each other, it is necessary to consider the global structure of the spaces of Procrustes-superimposed shapes (Fig. 5). Because the last step in a GPA includes an ordinary Procrustes fit of every landmark configuration in the sample to the mean shape, the target shape (point T in Fig. 5) should

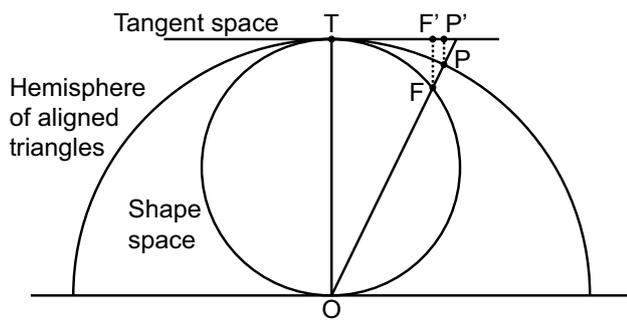


Fig. 5 A cross-section through Kendall’s shape space for triangles, the hemisphere of landmark configurations aligned by partial generalized Procrustes superimposition, and the tangent space (modified after Rohlf 1999; Klingenberg 2016). The shape corresponding to point T corresponds to the mean shape in the sample; it serves as the tangent point and also as the target configuration in the last iteration of the generalized Procrustes superposition. The points F’ and P’ are the tangent projections of F and P, respectively. Otherwise, points are labelled as in Fig. 2

be interpreted as the mean shape in this context. Because landmark configurations aligned by partial Procrustes superimposition all have centroid size 1.0, points representing these configurations all must lie on a spherical surface with the radius 1.0. Actually, triangles aligned by partial GPA lie on hemisphere of radius 1.0 (Rohlf 1999; Slice 2001). By contrast, landmark configurations superimposed by full GPA must lie inside this hemisphere because they have been scaled to smaller centroid sizes (unless they have the same shape as the target configuration). Triangles aligned by full GPA are actually in Kendall’s shape space (Fig. 5). The two spaces touch at the point corresponding to the shape that was used as the target in the last iteration of the cross the superposition. Because this point conventionally is used as the tangent point for the projection to the tangent space, it is the point where all three spaces intersect (point T in Fig. 5).

The shape tangent space for triangles is a plane (appearing as a straight line in the cross-section in Fig. 5). The tangent space has the same dimensionality as Kendall’s shape space (two for triangles—the surface of a sphere), but it is flat and not curved. The projections of landmark configurations into tangent space are different depending on whether a full or partial Procrustes superimposition is used (points F’ vs. P’ in Fig. 5), and these differences again depend on the amount of shape variation in the sample (more specifically, on the magnitude of the shape difference of each landmark configuration from the consensus shape that is used as the tangent point). In the vicinity of the tangent point (T in Fig. 5), the tangent space, the hemisphere of triangles aligned by partial Procrustes superimposition, and Kendall’s shape space are all close together, and the projections onto the tangent space will produce only negligible distortions. For shape data with limited amounts of variation, as for most

biological data, the tangent projection therefore provides a very good or excellent approximation of both Kendall’s shape space and the hemisphere of configurations aligned by partial Procrustes superimposition (Marcus et al. 2000).

Specifying the Coordinate System

Now that we understand the relationship between the different spaces (Fig. 5), we can use this relationship to explore the different components of shape and non-shape variation. This section does this first by setting up a coordinate system, then by examining how to change between different representations of shapes. Again, this section uses triangles as the simplest examples of shapes. Extending the reasoning to configurations with more than three landmarks is conceptually straightforward, but may be somewhat tedious (some discussion will follow in a later section).

Enumerating the Dimensions of Shape and Non-shape Variation

To set up a coordinate system for triangle shapes in two dimensions, we consider the arrangement in Fig. 5. It is easiest to start with the non-shape components of variation: translation, rotation and size. Each of these effects can be characterized as a vector $(\Delta x_1, \Delta y_1, \Delta x_2, \Delta y_2, \Delta x_3, \Delta y_3)$ describing changes of the x and y coordinates of the three landmarks from the target shape (the point T in Fig. 5). For convenience, we standardize the resulting vectors to unit length (i.e., so that the squared coefficients sum up to 1.0).

These non-shape changes include translations along the x and the y axis, that is, left–right and up–down shifts (Fig. 6a,b). The two corresponding vectors are:

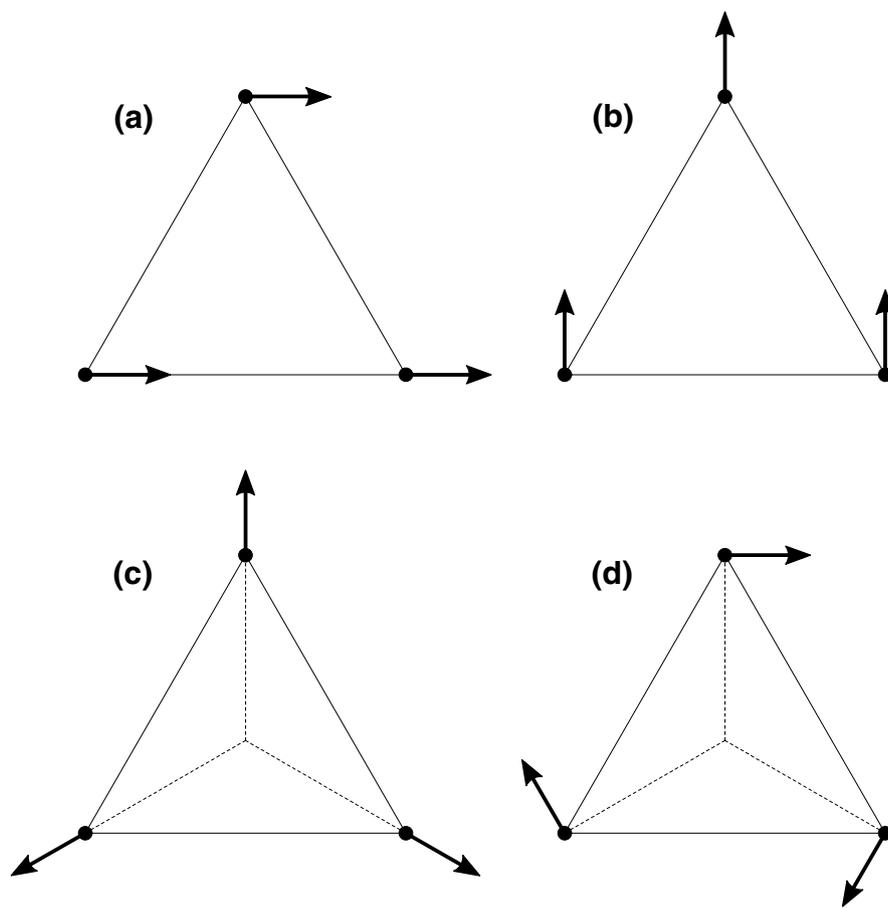
$$\left(\sqrt{\frac{1}{3}}, 0, \sqrt{\frac{1}{3}}, 0, \sqrt{\frac{1}{3}}, 0 \right)$$

$$\left(0, \sqrt{\frac{1}{3}}, 0, \sqrt{\frac{1}{3}}, 0, \sqrt{\frac{1}{3}} \right)$$

In order to standardize position, and to ensure that a landmark configuration has zero values for both these components, we can use centered landmark coordinates: subtracting the x and y coordinates of the centroid (the averages of the x and y coordinates of all landmarks) from the respective coordinates of every landmark, which results in shifting the configuration so that its centroid is at the origin of the coordinate system.

For the vector that stands for size variation, a little more thought is required. In the arrangement of Fig. 5, size

Fig. 6 Non-shape changes of a landmark configuration, in this example an equilateral triangle. **a** Translation along the x axis. **b** Translation along the y axis. **c** An increase in size (landmarks moving away from the centroid). **d** The linear approximation for a rotation around the centroid of the configuration. This approximation is not precise (the landmarks would move on circular arcs for an actual rotation; here, size also increases), but it is sufficient for rotations by a small angle. In all four diagrams, the changes are at the same, arbitrary scale (the corresponding vectors of shape change have a length of clearly less than one unit of Procrustes distance). In panels (c) and (d), all landmarks move equally far because of the symmetry of the reference shape (the equilateral triangle in this example). In general, the landmark displacements due to size change and rotation are proportional to the distances of the respective landmarks from the centroid in the reference shape



variation near the tangent point T involves shifts from that point up toward a larger size or down toward a smaller size, thus increasing or decreasing the distance from the origin O . This direction corresponds to that of the vector from the origin O to the tangent point T . The same conclusion also can be reached by reasoning that, for increasing the size slightly from point T without any other changes, all landmark coordinates need to be multiplied by the same constant slightly greater than 1.0, or for reducing the size, all coordinates need to be multiplied by the same constant slightly smaller than 1.0. Therefore, the vector that stands for the pure size changes is proportional to the vector of the centered landmark coordinates for the reference shape at the point T . As a consequence, in such a size change, the landmarks move away from the centroid or towards it (Fig. 6c). Because this vector represents a shape, the coordinates are standardized to centroid size 1.0, and thus the squared coordinates sum up to 1.0, so that no further change in scale is necessary. For the purposes of this paper, the entire shape space is of equal interest and the reference shape can therefore be chosen freely. Here I have selected an equilateral triangle as the reference shape, so the coordinate system specifies the same position of the shape space as in Fig. 3, with the equilateral triangle as the tangent point. The resulting vector is this:

$$\left(0, \sqrt{\frac{1}{3}}, 0.5, -\sqrt{\frac{1}{12}}, -0.5, -\sqrt{\frac{1}{12}} \right)$$

Finally, we need to find the vector for rotation. For a rotation around the centroid, each landmark moves along a circular arc around the centroid. For a rotation by a small angle, this can be approximated by a shift in a tangential direction by a distance that is proportional to the angle and to the distance of the landmark from the centroid (Fig. 6d). This can be obtained by taking the vector for the centered reference shape (i.e., the same vector as for size variation) and performing a rotation by 90° (note that the arrows in Fig. 6d are at right angles to the dashed lines running from the centroid to the three landmarks). This rotation can be done simply by swapping the x and y coordinates and then changing the sign of one coordinate (e.g. the y) for every landmark. The resulting vector for the equilateral triangle (again, scaled to unit length) is this:

$$\left(\sqrt{\frac{1}{3}}, 0, -\sqrt{\frac{1}{12}}, -0.5, -\sqrt{\frac{1}{12}}, 0.5 \right)$$

These four vectors together span the space of non-shape variation in the landmark coordinates of triangles in the vicinity of the reference shape (point T in Fig. 5). Each of these four vectors is orthogonal to the shape tangent space. Projecting out these four dimensions is a way to remove the non-shape variation from landmark data and thus a possible alternative to Procrustes superimposition and tangent projection in some situations, such as simulation studies (again, provided that variation is limited, so that nonlinear effects are relatively minor—Procrustes superimposition will be better able to handle those).

For the shape components, there is a choice for the directions of the coordinate axes. For triangles, there are two shape dimensions and therefore two coordinate axes need to be chosen. In principle, any pair of perpendicular vectors in shape tangent space will be equally suitable as coordinate axes and therefore could be used for this purpose. Here, I select the axes of symmetric and asymmetric shape variation about the vertical axis of the triangle, which are known to be orthogonal components in shape space and have clear geometric interpretations (Mardia et al. 2000; Kent and Mardia 2001; Klingenberg et al. 2002; Klingenberg 2015). This is possible because the reference shape, the equilateral triangle, is itself symmetric about the vertical axis. The symmetric component of variation corresponds to the vertical direction in Fig. 3, from which it is visible that this component encompasses isosceles triangles ranging from broad and flat to tall and narrow (note that all these triangles are symmetric about the vertical axis). The vector for the corresponding shape change stands for a triangle getting taller and narrower in a symmetric way, with the landmark at the top moving up (landmark 1) and the two landmarks of the base moving towards each other by equal amounts along the x axis:

$$\left(0, \sqrt{\frac{1}{3}}, -0.5, -\sqrt{\frac{1}{12}}, 0.5, -\sqrt{\frac{1}{12}} \right)$$

This vector needs to obey a number of constraints because it must be orthogonal to all preceding vectors representing the non-shape components of variation (e.g., maintaining unit centroid size, etc.), and some additional conditions to fulfil the symmetry requirements. Finally, the shape change for the asymmetry components corresponds to horizontal movement off the vertical meridian in Fig. 3. Again, several constraints apply that stand for constant size, position, orientation and no variation in the symmetric component of shape. The resulting shape change features a horizontal shift of the landmark at the top and opposite shifts of the two landmarks of the baseline:

$$\left(-\sqrt{\frac{1}{3}}, 0, \sqrt{\frac{1}{12}}, -0.5, \sqrt{\frac{1}{12}}, 0.5 \right)$$

Together, these two vectors span the shape tangent space. All six vectors, including the four non-shape axes,

provide a coordinate system that can characterize the local variation of landmark coordinates in the neighborhood of the target shape. At a larger scale, these vectors also provide a coordinate system for all possible triangle shapes, where each axis has a particular geometrical meaning. This aspect will be used in the subsequent parts of this paper.

From Landmark Coordinates to Position in Kendall Shape Space

The standard tool for aligning landmark configurations to Kendall's shape space is generalized Procrustes superimposition, as discussed above. For the purposes of exploring the entire shape space, we can simplify this procedure by choosing a target shape arbitrarily rather than estimating the mean shape in a sample. In particular, we can choose an especially suitable target shape such as an equilateral triangle, which results in a convenient alignment of the coordinate system and the shape space (Fig. 3). The superimposition procedure therefore consists simply of aligning every shape in a dataset of simulated triangle shapes to the chosen target shape, equivalent to the series of ordinary Procrustes alignments of landmark configurations to the target shape that is part of the last iteration of the generalized Procrustes superimposition algorithm. If we want to align landmark configurations to Kendall's shape space, we use full Procrustes superimposition. Alternatively, if we want to fit configurations to the hemisphere of aligned preshapes (Fig. 5), we use partial Procrustes superimposition. The differences of the triangle shapes used in this analysis to the target shape can be arbitrarily big; accordingly, this approach can be used to explore the entire Kendall shape space.

The coordinates of the aligned triangles will be expressed as the x and y coordinates of all three landmarks, so that the result will be a vector of six coordinate values for each triangle. Yet, from the previous considerations, it is clear that points representing the triangles are all embedded in just three dimensions (Fig. 5): the two dimensions of the tangent space and the additional vertical dimension from the origin of the coordinate system (point O) to the target shape (point T), which corresponds to the size of vector from the preceding section. It is possible to obtain the more compact notation in terms of these three coordinate axes by multiplying the vector of six landmark coordinates representing each aligned triangle by a matrix composed of the two vectors that together span the shape tangent space and the size vector (with the vectors taken as a column vectors and assembled side by side into a 6-by-3 matrix). The result is the vector of three elements, which can be used in ordinary 3D visualizations and each have geometrically meaningful interpretations (Figs. 2 and 5).

From Position in Kendall Shape Space to Landmark Coordinates

For exploring Kendall's shape space, an equally important task is to start with coordinates of points in the shape space and to obtain the shapes of the corresponding triangles. For doing this, the same coordinate system as above can be used: a set of three Cartesian coordinates with the origin at point O of Fig. 5, consisting of the two coordinate axes of the shape tangent space and the axis from the origin of the coordinate system to the point T representing the target shape. Every point on Kendall's shape space, and therefore every possible triangle shape, is included in this three-dimensional space. The distance of each point in Kendall's shape space (point F in Fig. 5) from the origin O represents the centroid size of the triangle after full Procrustes superimposition, reflecting the scaling by the factor $\cos(\varrho)$ according to the difference from the target shape T (Fig. 3).

The transition from the three-dimensional coordinate system to the six coordinates of the landmarks in the triangles again can be achieved by multiplying the vector of the three coordinates of a point in Kendall's shape space with a matrix composed of the two vectors that span the shape tangent space and the size vector (this time, the vectors are taken as row vectors and stacked on top of each other, resulting in a 3-by-6 matrix). Conventionally, shapes are visualized with a centroid size of 1.0, corresponding to a scaling to point P in Fig. 5. This scaling to unit centroid size needs to be done jointly for all six landmark coordinates (i.e., so that the vector of landmark coordinates, which is automatically centered if it is computed as described here, has a length of 1.0).

The only exception to this is the antipode to the target shape in Kendall's shape space, the point O itself. This point does represent a triangle shape, but in the arrangement of Fig. 5, this point is the origin of the coordinate system and has coordinates (0, 0, 0). Therefore, this point needs special treatment to be transformed into the appropriate shape. In the situation explained here, this would be the equilateral triangle that is the mirror image of the target shape.

With these methods, it is possible to go back and forth between triangle shapes and positions in Kendall's shape space. We can make use of this to explore the structure of the shape space in more detail.

Walking on Kendall's Shape Space for Triangles

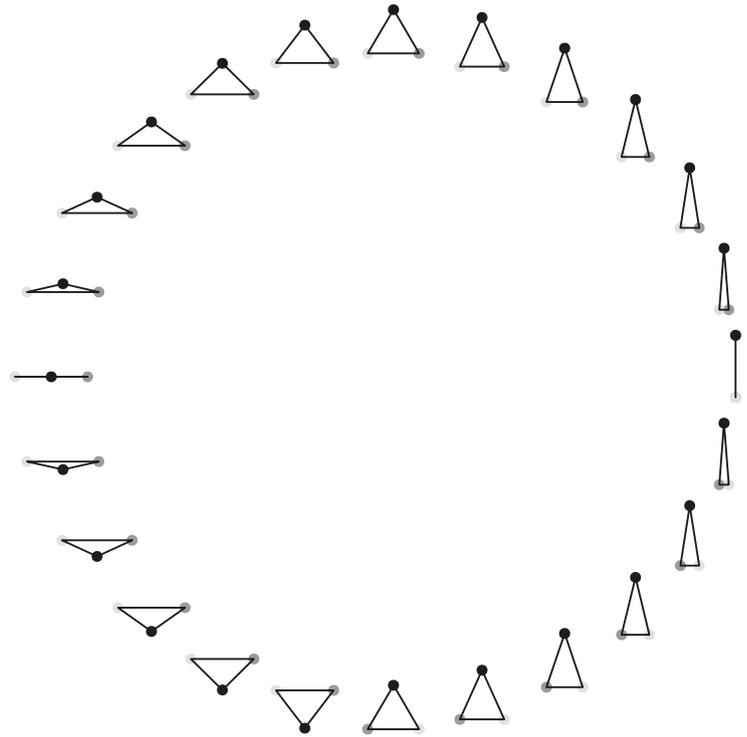
To obtain some intuitive understanding of Kendall's shape space for triangles, and therefore also more generally of the nature of shape variation at a large scale, a useful exercise is to imagine taking a walk around the shape space. If we choose a starting shape and repeatedly apply the same small

shape change to it, the resulting path is a straight line on the surface of the sphere or, more exactly, a great circle. This reasoning leads to the surprising conclusion that, by applying the same shape change to a starting triangle shape over and over, the same shape should eventually result again. This is equivalent to the thought experiment of taking an airplane and flying around the Earth in an exactly straight line—in the end, due to the curvature of the Earth, the airplane should return to the starting position. Examining the sequence of triangle shapes along such a route of travel on a great circle in Kendall's shape space can help to develop an intuitive grasp of shape variation at a large scale.

Walking on a Meridian

The first walk on Kendall's shape space is on the great circle that appears as the vertical meridian in Fig. 3, starting at the position of the equilateral triangle corresponding to the north pole, which also serves as the target shape for the Procrustes superimposition. Figure 7 shows the sequence of triangle shapes along this great circle (the view is from the left side in Fig. 3). Because this particular great circle corresponds to the component of shape variation symmetric about a vertical axis, all shapes along this path are isosceles triangles. In clockwise direction from the start at the equilateral triangle (Fig. 7, top), triangles become taller and narrower as the vertices on the baseline move towards each other: the baseline vertex on the right side (dark grey) moves toward the left and the vertex on the left (light grey) side moves toward the right (in clockwise direction in Fig. 7). On the equator of Kendall's shape space (3 o'clock in Fig. 7), the two baseline points coincide, so that the triangle is collinear. As the path continues beyond this point, the baseline vertices cross over as each of them maintains its respective direction of movement, so that the one that started out on the right side (dark grey) is now on the left, and vice versa. As a result, the triangles at corresponding distances above and below the equator are mirror images of each other. Also, as a second consequence, the triangles become broader and flatter again with increasing distance from the equator. This trend continues for a half-circle (to 9 o'clock in Fig. 7). At the position of the equilateral triangle that is the antipode and mirror image of the target shape (6 o'clock in Fig. 7), the orientation of the triangles changes by a rotation of 180° due to the Procrustes superimposition to the triangle at the north pole (actually, the triangle at 6 o'clock is drawn in an arbitrary orientation because its orientation is not defined by the Procrustes superimposition). The trend of triangles broadening, with the baseline points diverging relative to the height of the triangle, continues to the equator where the triangle is collinear because all three vertices lie on a single straight line. At this point, we cannot continue to describe the trend in this way because the triangle is already maximally broad

Fig. 7 Walk on a meridian in Kendall's shape space for triangles. The meridian chosen here is the one that appears as a vertical line in Fig. 3. The triangle shapes are shown in the orientations according to a Procrustes superimposition using the equilateral triangle at the 12 o'clock position as the target configuration, except for the mirror-image equilateral triangle at the 6 o'clock position, which is shown in an arbitrary orientation (its orientation is undefined; see the main text for further explanation)



and flat. Accordingly, we need to find a different characterization of the same shape change. We can observe that, as part of the trend of triangles flattening, the vertex opposite to the baseline (black) started from a position some distance below the baseline (just past 6 o'clock in Fig. 7) and has gradually been moving up towards the baseline. This aspect of the shape change is continuing across the equator of Kendall's shape space up to the equilateral triangle that is the target shape and was used as the starting point for this walk around the shape space (top of Fig. 7).

The reasoning above has repeatedly applied a shape change that led along a great circle all the way around Kendall's shape space. The change of the description at the second crossing of the Equator, however, raises the question whether it really was a single shape change or whether the shape change itself changed (at the 9 o'clock position in Fig. 7). Here is helpful again to use the analogy of the airplane. If the airplane departs at any point on the Earth and flies due north, it can keep doing so only until it reaches the north pole, but at that point it can no longer fly further north. Instead, the description needs to be changed: the airplane must cross the north pole in a straight line and afterwards fly due south. This new description works until the airplane reaches the south pole, where the description again needs to change. The airplane must cross the south pole in a straight line and then fly north to the point of departure. Changing the description of the course the airplane is taking over the Earth does not mean it is deviating from a great circle. In just the same way, switching the description of shape

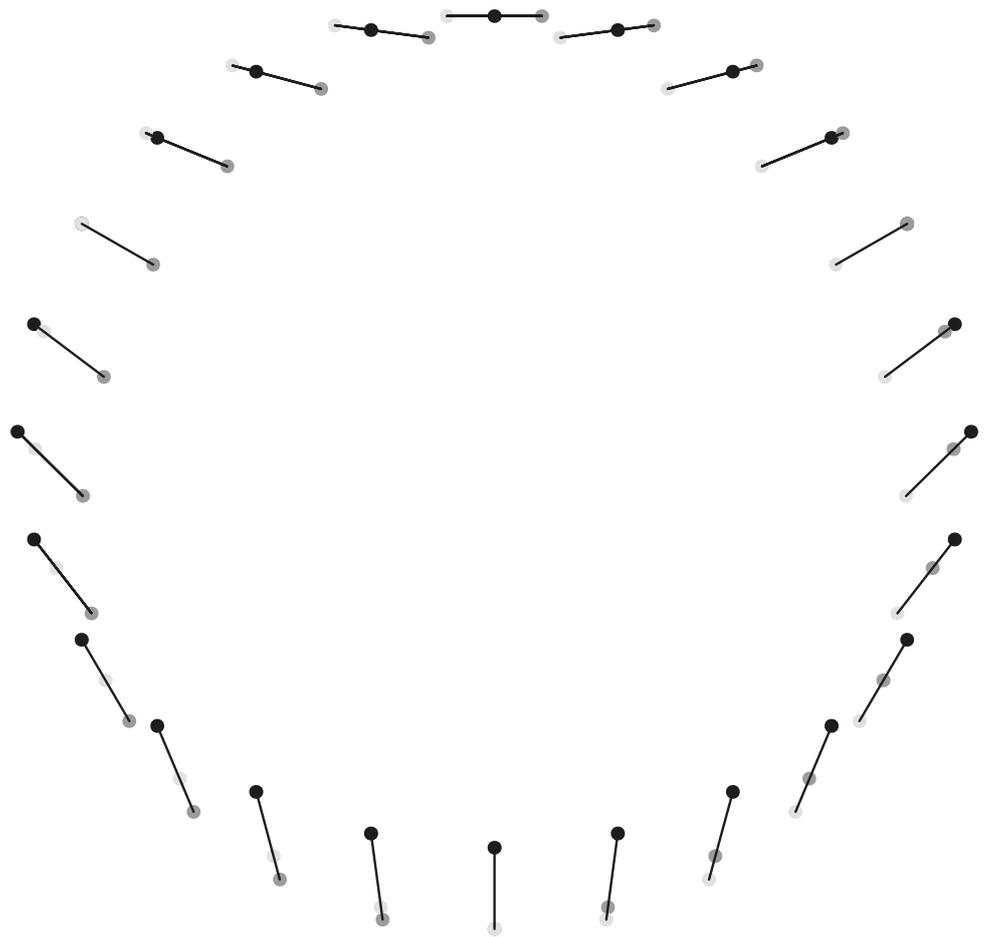
changes does not deviate from a great circle on Kendall's shape space (as long as the descriptions of the shape change, each focusing on different aspects of the change, are correct in themselves).

Walking Around the Equator

The second example is a walk around the equator of Kendall's shape space for triangles (Fig. 8), the great circle that is at equal distances between the two equilateral triangles that correspond to the poles (the equator appears as the outermost circle in the view of Fig. 3). The equator is somewhat special because it contains only collinear triangles: triangles with all three vertices on a straight line. As far as they apply to collinear triangles, other aspects of the structure of the shape space still hold. For instance, the symmetric component about the left-right axis corresponds to the vertical meridian in the view of Fig. 3, which intersects the equator in two points. These correspond to those two triangles that are both collinear and symmetric about the vertical axis: a completely flat triangle with one vertex in the middle between the two others (12 o'clock position in Fig. 8, black vertex in the center) and a triangle where the baseline has vanished so that the two vertices of the baseline are on top of each other (6 o'clock position in Fig. 8, with the light and dark grey vertices coinciding).

If we start at 12 o'clock in Fig. 8 and move clockwise, the black vertex moves toward the dark grey one, and goes past it at the 2 o'clock position. The order of vertices has

Fig. 8 Walk around the equator of Kendall's shape space for triangles. Accordingly, the triangles are all collinear, with all three vertices on a single straight line. The arrangement is essentially the same as on the peripheral circle of Fig. 3, with the configurations aligned by Procrustes superposition to the equilateral triangle (the resulting differences in orientation are irrelevant in the context of shape analysis)



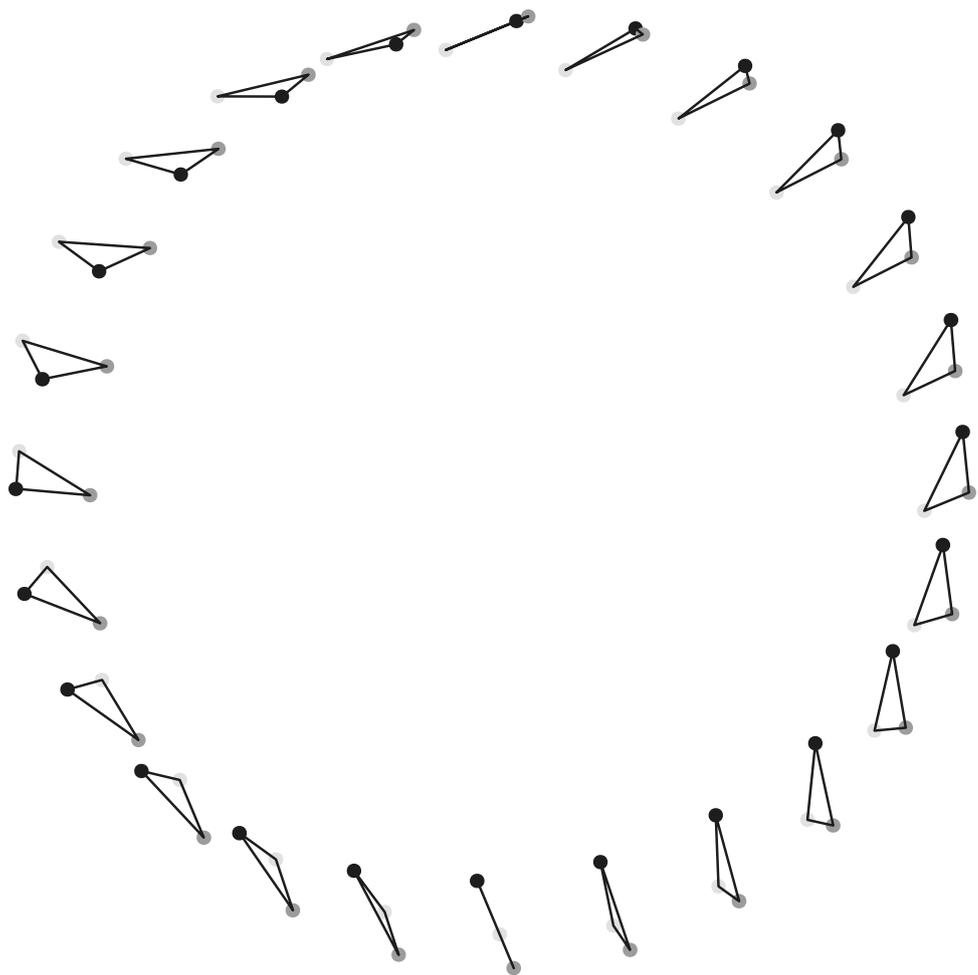
therefore switched from light grey—black—dark grey to light grey—dark grey—black. The black vertex keeps moving to a more and more extreme position relative to the other two vertices, until there is a limit at the 6 o'clock position, where the black vertex is at one end of the landmark configuration and the light and dark grey vertices are in the same position on the opposite end. At this position, therefore, we need to change the description of the shape change. An alternative description for the change from the 2 o'clock position on is that the dark grey vertex shifted its relative position away from the black vertex and toward the light grey one. Under this description, the dark grey vertex goes past the light grey one at the 6 o'clock position and continues to a more and more extreme relative position, until it reaches the limit at the 10 o'clock position. At that position, we can return to the original description of the shape change, that the black vertex shifts toward the dark grey one. At the 10 o'clock position, the black vertex goes past the light grey one and then moves gradually further away from it, reaching the midpoint at the 12 o'clock position. This completes the walk around the equator of the shape space.

Walking on a General Great Circle

The preceding two walks followed great circles that were aligned in special ways with the coordinate system adopted here for Kendall's shape space: a meridian going through the two poles of the equilateral triangles and the equator of collinear triangles. The next walk is not on such a special route, but is a great circle that is at oblique angles to the equator and meridians. Accordingly, it must have two points where it intersects the equator and also an intersection with every meridian. As a consequence, we expect two collinear triangles and a total of six isosceles triangles along this great circle, but the majority of triangles are not expected to be special in any such way (Fig. 9).

The great circle that this final walk follows is inclined by 45° relative to plane of the equator (and thus also makes a 45° angle with the axis between the two poles). It intersects the equator at two points defining a line that goes from the lower left to the upper right in Figs. 3 and 7 at an angle of 45° (i.e., the third diagram following the positions of 12 o'clock and 6 o'clock, when counting clockwise on the outer circle in Fig. 3 or in Fig. 8). These intersection points are

Fig. 9 Walk on a great circle on Kendall's shape space for triangles that is at oblique angles to the equator and meridians. The diagrams at the 12 o'clock and 6 o'clock positions show the shapes at the intersections of the great circle with the equator (these triangle shapes are therefore collinear). The 3 o'clock position is the point of the highest latitude (45° above the equator) and the 9 o'clock position is the point of lowest latitude (45° below the equator). The triangle shapes are shown in the orientations according to a Procrustes superimposition using the equilateral triangle as the target configuration



taken as the 12 o'clock and 6 o'clock positions in Fig. 9. The great circle rises above the equator toward the lower-right of Fig. 3 and dips below the equator in the upper-left side (note that shapes below the equator are not shown in Fig. 3, but they are the mirror images of the shapes above the equator). The point with the highest latitude, of 45° , is shown in Fig. 3 on the inner dashed circle, toward the lower right (the third diagram, counted either clockwise from 3 o'clock or counterclockwise from 6 o'clock). In Fig. 9, this point is in the 3 o'clock position. The point with the lowest latitude, where the great circle dips deepest to 45° below the equator, is in the upper-left direction in Fig. 3 (the corresponding shape is not shown in that figure, but its mirror image is shown on the inner dashed circle). In Fig. 9, this point is shown in the 9 o'clock position. None of the triangle shapes in Fig. 9 is exactly an isosceles triangle, because none of them is precisely at the intersection of the great circle with one of the respective meridians (Fig. 3). Because it does intersect with all six of those meridians, however, a total of six isosceles triangles exist on the great circle (this is the case for every great circle that is not itself a meridian, i.e., does not go through the two poles).

If we start the walk at the 12 o'clock position and proceed clockwise (Fig. 9), we therefore depart at the level of the equator and ascend to a latitude of 45° above the equator (3 o'clock), then descend back to the equator (6 o'clock), continue down to a latitude of 45° below the equator (9 o'clock), and finally return to the equator. The shape changes along the path do not have an obvious or simple interpretation, and several alternative descriptions could be found for the observed pattern. From the 12 o'clock position (Fig. 9), the dark grey vertex moves down and away from the line connecting the other two vertices and first toward, but eventually beyond the position of the light grey vertex. Simultaneously, the angle at the light grey vertex of the triangle expands, slowly up to about the 3 o'clock position and more rapidly after that, so that it reaches 180° at the 6 o'clock position (at the equator, corresponding to a collinear triangle). From that position, as we cross the equator, the order of vertex labels of the triangles reverses (this is related to the fact that mirror-image triangles are located in corresponding locations on opposite hemispheres). A shape change that continues past the 6 o'clock position is the shift of the light grey vertex away from the dark grey vertex toward the black vertex. In

addition, the light grey vertex starts to move away from the edge connecting the other two vertices and the angle at the light grey vertex starts to decrease again. Both of these features are reversals of the trends so far and both become more and more extreme toward the 12 o'clock position.

The shape changes along this walk around the shape space are inherently more complex and more difficult to understand than the ones in the previous two examples. The reason is that those concern a meridian of isosceles triangles, where all triangles are symmetric, or the equator, where there are only collinear triangles; because of those special conditions, the shape changes are relatively simple. But of course, the majority of triangle shapes do not belong to those special categories. Yet the strategy of understanding the changes throughout the sequence of shapes along a great circle by combining a chain of smaller-scale changes also works in this general case. The continuity of the great circle corresponds to a continuity in the changes among the triangle shapes that occur on it in Kendall's shape space. The walks on great circles in shape space show that repeatedly applying the same shape change eventually results in the triangle shape from which the walk started. At first, this may seem surprising and counterintuitive, but thinking through a few of these walks may develop a new sense of intuition on this kind of space that curves in on itself and is closed in this manner.

Shape Spaces for More than Three Landmarks

Shape spaces for configurations with more than three landmarks are more complex and have more dimensions than the shape space for triangles (Small 1996; Kendall et al. 1999). They are more complex because they can have features such as singularities, which can be viewed as the equivalent of ridges or spikes on two-dimensional surfaces. Because of the higher dimensionality, these shape spaces are also more difficult to visualize. This section is an attempt to provide some ideas of how insights from the shape space for triangles can be carried over to the more complex situations as they underlie the vast majority of morphometric analyses in biology.

Enumerating Dimensions

The first and most straightforward step when considering shape spaces is to set up the coordinate system, following the reasoning outlined above for triangles. For 2D landmark configurations with more than three landmarks, the vectors that characterize the non-shape components can be enumerated in the same way as for triangles. The vectors for translation contain equal shifts of either the x or the y coordinates of all landmarks. The vector for size variation is the centered

target shape (often the mean shape in a sample). Finally, the vector for rotation is the same as the target shape rotated by 90° , which can be obtained by taking the centered target shape vector, swapping the x and y coordinates of each landmark and changing the sign of one coordinate of each landmark (e.g., every y coordinate). The shape components are more complex for landmark configurations with more than three landmarks, and there may be many reasonable choices for coordinate systems. For a 2D configuration with k landmarks, there are $2k - 4$ dimensions in the shape space as well as in its tangent space.

For 3D data, the non-shape vectors can be constructed in an analogous manner, but there are seven non-shape vectors. There are three translation vectors, one each for the x , y and z coordinate. The size vector, as for 2D data, is the same as the centered target shape. There are also three rotation vectors, which are obtained from the target shape by pairwise swapping of coordinates and changing signs of one of them, for each of the combination of the x with y , x with z , and y with z coordinates (in each of these combinations, the third coordinate of every landmark is left unchanged, and is the axis of the respective rotation). For a 3D configuration with k landmarks, there are therefore $3k - 7$ shape dimensions.

All the shape coordinates need to be orthogonal to the vectors of non-shape variation described above. This can be achieved by generalized Procrustes superimposition and projection to the tangent space. Alternatively, the shape variation can be extracted by projecting out the non-shape vectors from the data using some tools from matrix algebra: $\mathbf{Y} = \mathbf{X}(\mathbf{I} - \mathbf{N}\mathbf{N}^T)$, where \mathbf{Y} is the matrix of shape coordinates from n specimens, \mathbf{X} is the matrix of landmark coordinates, \mathbf{I} is an identity matrix, \mathbf{N} is the matrix composed of the four or seven non-shape vectors above (written as column vectors and assembled side by side into a matrix), and the superscript "T" denotes the matrix transpose. \mathbf{X} and \mathbf{Y} are n -by- $2k$ matrices for 2D data and n -by- $3k$ matrices for 3D data, \mathbf{I} is $2k$ -by- $2k$ for 2D data and $3k$ -by- $3k$ for 3D data, and \mathbf{N} is $2k$ -by- 4 for 2D data and $3k$ -by- 7 for 3D data. Because the projection procedure is based on linear operations, it works only for data where the variation is sufficiently small (including rotations, etc.). In practice, therefore, generalized Procrustes superimposition is the more reliable approach.

An important special case are landmark configurations with object symmetry, that means, if a line or plane of symmetry runs through the configuration and divides it into left and right halves that are mirror images of each other. This applies to many biological structures (e.g. the heads of most animals, elements of the axial skeleton in vertebrates, and in plants many leaves and flowers) and adds further constraints to the possible variation of landmarks. The Procrustes procedure for configurations with object symmetry ensures that the overall mean shape, which is used to define the shape tangent space, is perfectly symmetric, and the shape space

can be subdivided into orthogonal components of symmetric shape variation and asymmetry (Mardia et al. 2000; Kent and Mardia 2001; Klingenberg et al. 2002; Klingenberg 2015). In this case, there are two kinds of landmarks: single landmarks in the median line or plane and landmarks that occur as pairs on the left and right sides. The two components of variation entail different constraints for these two types of landmarks. The single median landmarks can move within the line or plane of symmetry for the symmetric component, but only in a lateral direction (perpendicular to the median axis or plane) for the asymmetry component. For paired landmarks, the constraint is that the relative movements in landmarks of each pair are in the same anatomical direction (anterior/posterior, medial/lateral and superior/inferior) for the symmetric component and in opposite anatomical directions for the asymmetry component. For either component, the relative movements in the two landmarks of each pair are perfectly correlated. As a result of these constraints, it is possible to enumerate the degrees of freedom for each component to calculate the dimensionality of the corresponding subspace of Kendall's shape space (Klingenberg et al. 2002; Klingenberg 2015). For 2D data, both components have the same dimensionality. For 3D data, the dimensionalities depend on the numbers of paired and single median landmarks: paired landmarks contribute equally to both components, but for each median landmark, there are two degrees of freedom for the symmetric component (anterior/posterior and superior/inferior movements), but only one degree of freedom for the asymmetry component (lateral movement only; Klingenberg et al. 2002; Klingenberg 2015). With object symmetry, therefore, there is a structure of the shape space that can be used for distinguishing different aspects of shape variation. But with more than three landmarks, each of these components usually has multiple dimensions (not just one each as for triangles). For types of symmetry other than bilateral symmetry, such as radial symmetry, similar decompositions of the shape space exist, but are more complex (Savriama and Klingenberg 2011).

Apart from their partition into different components of variation due to object symmetry, there is no obvious way to characterize the structure of shape spaces for configurations with more than three landmarks. A possibility is the thin-plate spline decomposition of all possible shape changes from a reference shape into principal warps and partial warps and uniform components (Bookstein 1989, 1991; Rohlf 1993). If the reference shape for the thin-plate spline is set equal to the target shape for the Procrustes superimposition and the shape tangent space, the complete set of partial warps and uniform components of shape change will provide a set of coordinate axes for the shape tangent space. Nevertheless, it is not obvious how this decomposition relates to other geometric considerations, such as symmetry (it is plausible that there is a connection if a perfectly symmetric

reference shape is used for the thin-plate spline), nor is it clear what the biological relevance of such a purely geometric decomposition may be.

In practice, the Procrustes superimposition and tangent projection take care of separating the shape and non-shape aspects of variation. This provides a local linear approximation of the small region of shape space in the neighborhood occupied by the data at hand. If the landmark configurations have object symmetry, it is important to use analyses that take this into account and generate symmetric and asymmetry components of variation, depending on the context of the study (Klingenberg et al. 2002; Klingenberg 2015). For most contexts, when asymmetry is of no interest, the symmetric component of variation provides an optimal characterization of shape variation, with approximately half the dimensionality of the entire shape space. When asymmetry itself is the focus of attention, of course the asymmetry component is the appropriate component of variation to consider. Explicitly distinguishing these levels can provide considerable insight and analytical power for biological studies (Klingenberg et al. 2012; Klingenberg 2014). Apart from these choices, there is no need to specify a coordinate system for shape analyses in empirical morphometric studies. Multivariate analyses such as principal component analysis, multivariate regression, partial least squares analysis or canonical variate analysis will automatically provide coordinate systems for the shape tangent space that are optimal for the specific question each of these analyses aims to answer.

Global Structure of Shape Spaces

The global structure of shape spaces for configurations with more than three landmarks is difficult to grasp. Such spaces are multidimensional and non-linear spaces, and therefore very hard to visualize. Whereas much about the overall structure of shape spaces cannot be understood intuitively, it is possible and important to consider some questions, in particular as far as they relate to fitting a tangent space to the shape space.

The main question in this context is whether all shape spaces are like the shape space for triangles, which is a sphere, and therefore is locally smooth in the neighborhood of every possible triangle shape. Fortunately, this is also the case for all shape spaces for configurations with more than three landmarks in two dimensions (Le and Kendall 1993). This means that the tangent projection can be used to obtain a local linear approximation of the shape space, not just for triangles, but also generally for 2D data from configurations with more than three landmarks.

For 3D data, however, things are more complicated because shape spaces contain singularities (Le and Kendall 1993; Small 1996; Kendall et al. 1999), which are abrupt local changes: think of them as the equivalent of the sharp

ridges, spikes or edges we are familiar with from curved 2D surfaces. Singularities in Kendall shape spaces for 3D configurations occur in situations when all the landmarks in a configuration are aligned along a single straight line (Le and Kendall 1993; Small 1996). For these configurations, we can apply a rotation around the axis along which the landmarks are aligned, and regardless of the angle of the rotation, the landmark configuration is unchanged. These collinear configurations are also the only situation where singularities occur for 3D data, whereas the remainder of the shape spaces are non-singular (Le and Kendall 1993; Small 1996).

To get an intuitive idea of a singularity in a shape space, consider the space of triangles in 3D. Because we can flip over triangles (i.e., apply a rotation by 180° around an axis that lies in the plane of the triangle), the shape of each triangle and its mirror image is the same in 3D. This is fundamentally different from the considerations about in the preceding sections triangles in 2D, where the option of flipping over the triangle is not available (flipping uses a rotation through the third dimension). As a consequence, the shape space for triangles in 3D is a hemisphere and not a sphere, as for triangles in 2D. The entire shape space for triangles in 3D is the same as the hemisphere visible in Fig. 3 (which is half the shape space for triangles in 2D). The opposite hemisphere, corresponding to the mirror images of the triangles visible in Fig. 3, can be viewed as being pushed in and glued against the hemisphere that is visible in the figure, so that each triangle shape and its mirror image are in the same point. As a result, the equator of collinear triangles (which is the locus of all triangles that are identical with their mirror images) forms a sharp edge where the hemisphere ends or abruptly folds in from convex to concave. This sharp edge is a singularity. It is clear that, at this edge, a tangent plane is not fully determined (it can be rotated freely around the local direction of the edge at the tangent point of interest).

The conclusion is that the approach of tangent projections fails at singularities. Away from singularities, however, the tangent space approximation works also for 3D shape analyses (Small 1996).

Tangent Space Approximation to Shape Spaces

The main practical role of the shape space is to provide information on the variation of shapes through the arrangement of the data points in the neighborhood of the average shape. The shape tangent space provides a linear approximation and therefore makes it possible to use standard approaches from multivariate statistics in the context of shape analysis. The most important question concerning the global structure of shape spaces is therefore whether they consistently provide the conditions for using this strategy with actual biological data, or whether different approaches are required instead.

In essence, this boils down to the question whether shape spaces are sufficiently smooth in the neighborhood of the mean shape for the tangent projection to be a satisfactory approximation. The answer to this question is yes, as long as the shape space in the vicinity of the mean shape contains no singularity (Le and Kendall 1993; Small 1996; Kendall et al. 1999).

This raises the question whether singularities are to be expected in biological data. As outlined in the preceding section, singularities in shape spaces occur only under quite restrictive conditions: if all the landmarks are aligned along a single straight line, and only for 3D data. These conditions are rarely met for biological data and easy to diagnose, so that this should not be a serious problem in practice. Morphometric data usually do not consist of landmarks aligned along a single line. There is no need to worry about structures that are just long and slender: as long as the range of variation does not include the situation where the landmarks are actually aligned along a straight line, the conditions for a singularity are not met. If the data indeed include this “single file” arrangement of landmarks, then analyses in two or even a single dimension (Small 1996) avoid the problem of singularities in the shape space and are likely to fit the data at least reasonably well (if the landmarks are all on a straight line, they occupy just one dimension).

How well the tangent approximation works for actual biological data can be assessed by comparing Procrustes distances and tangent distances in empirical datasets. The problem here, if there is any, is not about singularities in the shape space, but about the scale of variation in the data. The expectation is that this will be unproblematic for data with a relatively small amount of shape variation, so tests that examine biological data at large taxonomic scales are particularly valuable. The study by Marcus et al. (2000) has done this for skull shape across all orders of mammals and found an excellent overall agreement of shape distances even at this very large scale, with correlations exceeding 0.999 for both full and partial Procrustes superimposition. Nevertheless, the tangent distances of specimens to the most outlying taxon, the common dolphin, were consistently slightly less than the partial Procrustes distances for the same pairwise comparisons, suggesting that there was some subtle effect of the distortion due to tangent projection for the extreme comparisons. In recent years, a number of studies have computed similar correlations between pairwise Procrustes and tangent distances among landmark configurations in diverse organisms and consistently found that these correlations were extremely close to 1.0 (Pretorius and Scholtz 2001; Polly 2002; Frost et al. 2003; Fontaneto et al. 2004; Lockwood et al. 2004; McNulty 2004; Viscosi and Cardini 2011; Bai et al. 2012; De Meulemeester et al. 2012; Renner 2012; Neustupa 2013; Siver et al. 2013; Ullmann et al. 2017; Manacorda and Asurmedi 2018). Rohlf (1999) commented that he was not aware of any reports of biological datasets

where the tangent projection did not provide a satisfactory approximation (with the exception of inadvertent inclusion of reflected specimens in the data). The same still appears to be the case from today's perspective.

Conclusions

This paper has explored Kendall's shape space for triangles by travelling around it on paths that are great circles on its spherical surface. The observation that taking an initial triangle shape and repeatedly applying the same shape change eventually results in a return to the starting shape highlights the fact that the shape space is closed and curving in on itself. Initially, this may appear counterintuitive, but it is an important insight for thinking about shape variation. The curved nature also extends to shape spaces for configurations with more than three landmarks.

Because the curvature of Kendall's shape space or the space of preshapes aligned by partial Procrustes superimposition can produce complications for subsequent multivariate analyses of shape, the data are usually projected into a shape tangent space that touches Kendall's shape space at the mean shape. The tangent projection provides a local linear approximation to Kendall's shape space. It appears that the tangent approximation is very good in most biological datasets where it has been examined, including datasets at large taxonomic scale. In principle, singularities of shape spaces for configurations with more than three landmarks could be problematic for the tangent space approximation or invalidate it altogether. Theoretical considerations of the conditions when singularities occur on shape spaces indicate that this is unlikely for biological shapes. Therefore, the method to approximate shape space in the vicinity of the mean shape by a linear tangent space is valid for the great majority of biological applications of geometric morphometrics. It is important, however, to keep in mind that certain constraints apply to the patterns of variation that are possible within the shape tangent space.

Acknowledgements The thought experiment of walks around Kendall's shape for triangles has long been a part of my explanations in workshops in have taught, and more usually than not I lost myself somewhere along the way. Hence, I have long had the idea of working out in detail and writing up a more formal account. I am very grateful to the participants in all those workshops for their understanding. I also thank Julien Claude for his helpful comments on an earlier version of this manuscript.

Conflict of interest The author has no conflict of interest.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated

otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Bai, M., Ahrens, D., Yang, X.-K., & Ren, D. (2012). New fossil evidence of the early diversification of scarabs: *Alloioscarabaeus cheni* (Coleoptera: Scarabaeoidea) from the Middle Jurassic of Inner Mongolia, China. *Insect Science*, *19*, 159–171.
- Boas, F. (1905). The horizontal plane of the skull and the general problem of the comparison of variable forms. *Science*, *21*, 862–863.
- Bookstein, F. L. (1989). Principal warps: thin-plate splines and the decomposition of deformations. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *11*, 567–585.
- Bookstein, F. L. (1991). *Morphometric tools for landmark data: Geometry and biology*. Cambridge: Cambridge University Press.
- Bookstein, F. L. (1996). Biometrics, biomathematics and the morphometric synthesis. *Bulletin of Mathematical Biology*, *58*, 313–365.
- De Meulemeester, T., Michez, D., Aytikin, A. M., & Danforth, B. N. (2012). Taxonomic affinity of halictid bee fossils (Hymenoptera: Anthophila) based on geometric morphometrics analyses of wing shape. *Journal of Systematic Palaeontology*, *10*(4), 755–764.
- Dryden, I. L., & Mardia, K. V. (1998). *Statistical shape analysis*. Chichester: Wiley.
- Dryden, I. L., & Mardia, K. V. (2016). *Statistical shape analysis, with applications in R* (2nd ed.). Chichester: Wiley.
- Fontaneto, D., Melone, G., & Cardini, A. (2004). Shape diversity in the trophi of different species of *Rotaria* (Rotifera, Bdelloidea): A geometric morphometric study. *Italian Journal of Zoology*, *71*, 63–72.
- Frost, S. R., Marcus, L. F., Bookstein, F. L., Reddy, D. P., & Delson, E. (2003). Cranial allometry, phylogeography, and systematics of large-bodied papionins (Primates: Cercopithecinae) inferred from geometric morphometric analysis of landmark data. *Anatomical Record*, *275A*, 1048–1072.
- Goodall, C. R. (1991). Procrustes methods in the statistical analysis of shape. *Journal of the Royal Statistical Society B*, *53*, 285–339.
- Gower, J. C. (1966). Some distance properties of latent root and vector methods used in multivariate analysis. *Biometrika*, *53*, 325–338.
- Gower, J. C. (1975). Generalized Procrustes analysis. *Psychometrika*, *40*, 33–51.
- Kendall, D. G. (1984). Shape manifolds, Procrustean metrics, and complex projective spaces. *Bulletin of the London Mathematical Society*, *16*, 81–121.
- Kendall, D. G., Barden, D., Carne, T. K., & Le, H. (1999). *Shape and shape theory*. Chichester: Wiley.
- Kent, J. T. (1994). The complex Bingham distribution and shape analysis. *Journal of the Royal Statistical Society B*, *56*, 285–299.
- Kent, J. T., & Mardia, K. V. (2001). Shape, Procrustes tangent projections and bilateral symmetry. *Biometrika*, *88*, 469–485.
- Klingenberg, C. P., Barluenga, M., & Meyer, A. (2002). Shape analysis of symmetric structures: Quantifying variation among individuals and asymmetry. *Evolution*, *56*, 1909–1920.
- Klingenberg, C. P. (2010). Evolution and development of shape: Integrating quantitative approaches. *Nature Reviews Genetics*, *11*, 623–635.
- Klingenberg, C. P., Duttke, S., Whelan, S., & Kim, M. (2012). Developmental plasticity, morphological variation and evolvability: A multilevel analysis of morphometric integration in the shape of compound leaves. *Journal of Evolutionary Biology*, *25*, 115–129.

- Klingenberg, C. P. (2014). Studying morphological integration and modularity at multiple levels: Concepts and analysis. *Philosophical Transactions of the Royal Society of London B Biological Sciences*, 369, 20130249.
- Klingenberg, C. P. (2015). Analyzing fluctuating asymmetry with geometric morphometrics: Concepts, methods, and applications. *Symmetry*, 7(2), 843–934.
- Klingenberg, C. P. (2016). Size, shape, and form: Concepts of allometry in geometric morphometrics. *Development Genes and Evolution*, 226(3), 113–137.
- Le, H., & Kendall, D. G. (1993). The Riemannian structure of Euclidean shape spaces: A novel environment for statistics. *Annals of Statistics*, 21(3), 1225–1271.
- Lockwood, C. A., Kimbel, W. H., & Lynch, J. M. (2004). Morphometrics and hominoid phylogeny: Support for a chimpanzee–human clade and differentiation among great ape subspecies. *Proceedings of the National Academy of Sciences of the USA*, 101, 4356–4360.
- Manacorda, C. A., & Asurmedi, S. (2018). Arabidopsis phenotyping through geometric morphometrics. *GigaScience*, 7(7), 1–20.
- Marcus, L. F., Hingst-Zaher, E., & Zaher, H. (2000). Application of landmark morphometrics to skulls representing the orders of living mammals. *Hystrix-Italian Journal of Mammalogy*, 11, 27–47.
- Mardia, K. V., Kent, J. T., & Bibby, J. M. (1979). *Multivariate analysis*. London: Academic Press.
- Mardia, K. V., Bookstein, F. L., & Moreton, I. J. (2000). Statistical assessment of bilateral symmetry of shapes. *Biometrika*, 87, 285–300.
- McNulty, K. P. (2004). A geometric morphometric assessment of hominoid crania: Conservative African apes and their liberal implications. *Annals of Anatomy*, 186, 429–433.
- Neustupa, J. (2013). Patterns of symmetric and asymmetric morphological variation in unicellular green microalgae of the genus *Micrasterias* (Desmidiaceae, Viridiplantae). *Fottea*, 13(1), 53–63.
- Polly, P. D. (2002). Phylogenetic tests for differences in shape and the importance of divergence times: Eldredge’s enigma explored. In N. MacLeod & P. L. Forey (Eds.), *Morphology, shape and phylogeny* (pp. 220–246). London: Taylor & Francis.
- Pretorius, E., & Scholtz, C. H. (2001). Geometric morphometrics and the analysis of higher taxa: A case study based on the metendos-ternite of the Scarabaeoidea (Coleoptera). *Biological Journal of the Linnean Society*, 74, 35–50.
- Renner, M. A. M. (2012). Further insight into lobule teeth homology in Lejeuneaceae subf. Lejeuneoideae from *Cheilelejeunea oscilla*, a new species from Australia. *Bryologist*, 115(4), 536–556.
- Rohlf, F. J., & Slice, D. E. (1990). Extensions of the Procrustes method for the optimal superimposition of landmarks. *Systematic Zoology*, 39, 40–59.
- Rohlf, F. J. (1993). Relative warp analysis and an example of its application to mosquito wings. In L. F. Marcus, E. Bello & A. García-Valdecasas (Eds.), *Contributions to morphometrics* (pp. 131–159). Madrid: Museo Nacional de Ciencias Naturales.
- Rohlf, F. J. (1999). Shape statistics: Procrustes superimpositions and tangent spaces. *Journal of Classification*, 16, 197–223.
- Rohlf, F. J. (2000). On the use of shape spaces to compare morphometric methods. *Hystrix-Italian Journal of Mammalogy*, 11, 9–25.
- Savriama, Y., & Klingenberg, C. P. (2011). Beyond bilateral symmetry: Geometric morphometric methods for any type of symmetry. *BMC Evolutionary Biology*, 11, 280.
- Siver, P. A., Wolfe, A. P., Rohlf, F. J., Shin, W., & Jo, B. Y. (2013). Combining geometric morphometrics, molecular phylogeny, and micropaleontology to assess evolutionary patterns in Mallomonas (Synurophyceae: Heterokontophyta). *Geobiology*, 11, 127–138.
- Slice, D. E. (2001). Landmark coordinates aligned by Procrustes analysis do not lie in Kendall’s shape space. *Systematic Biology*, 50, 141–149.
- Small, C. G. (1996). *The statistical theory of shape*. New York: Springer-Verlag.
- Sneath, P. H. A. (1967). Trend-surface analysis of transformation grids. *Journal of Zoology*, 151, 65–122.
- Srivastava, A., & Klassen, E. P. (2016). *Functional and shape data analysis*. New York: Springer-Verlag.
- Ullmann, P. V., Bonnan, M. F., & Lacovara, K. J. (2017). Characterizing the evolution of wide-gauge features in stylopodial limb elements of titanosauriform sauropods via geometric morphometrics. *Anatomical Record*, 300(9), 1618–1635.
- Viscosi, V., & Cardini, A. (2011). Leaf morphology, taxonomy and geometric morphometrics: A simplified protocol for beginners. *PLOS ONE*, 6, e25630.
- Zelditch, M. L., Swiderski, D. L., & Sheets, H. D. (2012). *Geometric morphometrics for biologists: A primer* (2nd ed.). Amsterdam: Elsevier.

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.