Journal of Classification 20:181-220 (2003)
DOI: 10.1007/s00357-003-0012-6

The Geometry of Triadic Distances

Mark de Rooij

Leiden University

John C. Gower

The Open University

Abstract: Triadic distances t defined as functions of the Euclidean (dyadic) distances a_1, a_2, a_3 between three points are studied. Special attention is paid to the contours of all points giving the same value of t when a_3 is kept constant. These isocontours allow some general comments to be made about the suitability, or not, for practical investigations of certain definitions of triadic distance. We are especially interested in those definitions of triadic distance, designated as canonical, that have optimal properties. An appendix gives some results we have found useful.

Keywords: Triadic distance, Dyadic distance, Isocontours, Triangle geometry, Multidimensional scaling, Parametric models.

Authors' addresses: Mark de Rooij, Department of Psychology, Leiden University, P.O. Box 9555, 2300 RB Leiden, The Netherlands; John C. Gower, Department of Statistics, Walton Hall, The Open University, Milton Keynes, MK7 6AA, U.K.

1. Introduction

Originating with Hayashi (1972), interest in triadic distance models has grown during the last decade with developments by Cox, Cox and Branco (1991), Pan and Harris (1991), Joly and Le Calvé (1995), Daws (1996), Heiser and Bennani (1997), De Rooij and Heiser (2000), De Rooij (2001), De Rooij (2002), and Gower and De Rooij (2003). There has been little interest in the underlying geometry of triadic distances. In this paper we discuss Euclidean properties of triadic distances, some proposed in the literature and others new.

Before examining triadic distance, we briefly consider the place of dyadic terms in data analysis. In general, these terms may express distances or they may be of bilinear product form. Distance-terms are primarily used for multidimensional scaling or (hierarchical) classification, where the results of analysis are often presented as graphical displays, whose interpretation relies heavily on the metric and distance properties of the measure. Bilinear terms are often used as parametric interaction terms in statistical models. Then, geometrical considerations may be irrelevant but the links between bilinearity, inner-products and distance are sometimes exploited to give visualizations similar to those used in multidimensional scaling (see e.g. Kempton, 1984 and Denis and Gower, 1996). Furthermore, although rare, direct distance parameters could be, and perhaps more frequently should be, included as terms in statistical models. Examples of such approaches can be found in Carroll and Pruzansky (1980, 1983, 1986) who proposed 'hybrid' models for the analysis of proximity data; Takane, Bozdogan and Shibayama (1987) who described discriminant analysis using Euclidean distances; and Takane (1987) and De Rooij and Heiser (2003a, 2003b) who use Euclidean distances in the analysis of contingency tables.

Thus, distance interpretations are important not only in multidimensional scaling but also in some applications of statistical models; the geometry required is well-understood by all. Similar considerations apply to triadic terms, whether they are set in a multidimensional scaling context (Gower and De Rooij, 2003) or used as parameters that model three-way interactions, but the required geometry is not well-understood. Our objective is to provide some tools that will contribute to the understanding of triadic geometry and hence to the interpretation of the analysis of triadic distances. In this paper we discuss triadic terms that are all based on dyadic distances for which the triangle inequality holds.

To fix notation see Figure 1, which shows three points A_1, A_2, and A_3. The dyadic distance between A_2 and A_3 is denoted by a_1 (the side opposite

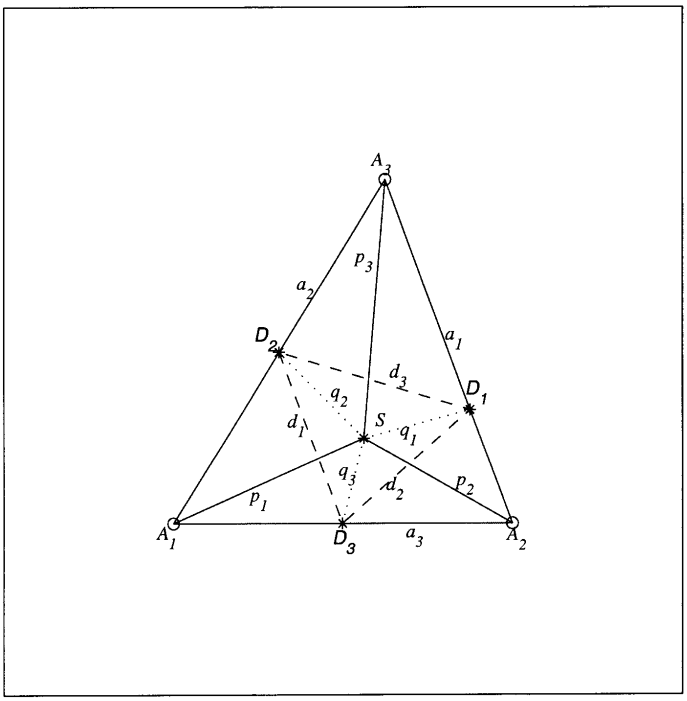

Figure 1. Three points A_1, A_2 and A_3, with dyadic distances a_1, a_2 and a_3, distances from the vertices to a center S, p_1, p_2 and p_3, distances from a center S to points D_1, D_2, and D_3 on the sides of the triangle A_1 A_2 A_3 denoted by q_1, q_2, and q_3, and distances between D_1, D_2, and D_3 denoted by d_1, d_2, and d_3.

A_1), the distance between A_1 and A_3 by a_2, and the distance between A_1 and A_2 by a_3. Also shown is a "center" S whose distances from A_1, A_2, A_3 are p_1, p_2 and p_3. Points D_1, D_2 and D_3 are shown on the sides of the triangle and the lengths SD_1, SD_2 and SD_3 are denoted by q_1, q_2 and q_3; and the lengths D_2D_3, D_1D_3 and D_1D_2 by d_1, d_2 and d_3 respectively. We refer to $D_1D_2D_3$ as an *interior triangle* and when, as is often the case, D_1, D_2 and D_3 are orthogonal projections of S, then $D_1D_2D_3$ is termed the *pedal triangle* of S. Appendix A3.3 gives conditions for an interior triangle to be a pedal triangle.

Triadic distances (t) are symmetric functions of dyadic distances, with the forms:

$$t = f(a_1,a_2,a_3),\ t = f(d_1,d_2,d_3),\ t = f(p_1,p_2,p_3),\ \text{or}\ t = f(q_1,q_2,q_3).^1$$

[1] Of course $f(.)$ is not the same function in all cases, though sometimes it is.

Analogous to the symmetry of dyadic distance, the functions $f(.)$ are chosen to be symmetric in the sense that the parameters a_1,a_2,a_3 (equally d_1,d_2,d_3 etc.) may be permuted without affecting t:

i.e. $t = f(a_1,a_2,a_3)=f(a_1,a_3,a_2)=f(a_2,a_1,a_3)=f(a_2,a_3,a_1)=f(a_3,a_1,a_2)=f(a_3,a_2,a_1)$.

Although some triadic distances may be expressed most compactly in terms of (d_1,d_2,d_3) or of (p_1,p_2,p_3) or of (q_1,q_2,q_3), all those considered below remain symmetric functions of a_1,a_2,a_3, i.e. all triadic distances in the sequel can be rewritten as symmetric functions of a_1, a_2, and a_3.

The most popular functions used in the literature are the L_r-transforms:

$$t = \left(a_1^r + a_2^r + a_3^r\right)^{1/r}.$$

When $r = 1$ the triadic distance is often called the perimeter distance, when $r = 2$ the generalized Euclidean distance, and when $r = \infty$ the generalized dominance distance, with $t = max(a_1,a_2,a_3)$. In this paper we also explore the properties of many other symmetric functions. Our intention is partly to bring to attention many interesting possibilities for modeling triadic distances but also to eliminate some definitions that are shown to have undesirable properties. To this end we are particularly interested in triadic distances that express some interesting interpretable optimal properties of $A_1A_2A_3$. Often, optimal properties are associated with special definitions of the center S, including the centroid (G), the mediancenter (M), the incenter (I), the circumcenter (C), and the orthocenter (H) as well as others derived in the following.

For dyadic Euclidean distances the circle center A_1, defines a set of points A_2 with equal dyadic Euclidean distance from A_1: this set is named an isocontour. Similarly, for given points A_1 and A_2, we define the triadic isocontour to be a set of points A_3, having equal triadic distances with A_1 and A_2. Specific points in this set will be denoted by A_3. The horizontal dimension will be denoted by x, and the vertical by y. Coordinates for A_3 will be denoted by $[x,y]$ (see Appendix A3.4), so the equation of the isocontour of the set A_3 is a function we denote by $h(x,y)$, which may be shown in two-dimensions. Just as the diamond, the circle, and the square are isocontours for the city-block, Euclidean, and dominance dyadic distances respectively, for different triadic distances the isocontours have varying forms. Without knowledge of the isocontours, a triadic distance graph is not interpretable. The contour lines are essential for deciding which points have smaller or large values of t. Current work on triadic distances has paid little attention to their geometric interpretation and more specifically to their isocontours.

All the triadic distances we study satisfy the property:

$$f(la_1, la_2, la_3) = g(l)f(a_1,a_2,a_3)$$

where, $g(l)=l$, $g(l)=l^2$, or $g(l)=l^3$. With this property, the values of t for similar triangles are proportional up to a scaling factor, enabling us to choose a scale in which $a_3 = 1$. Hence, to study the contour lines, we will fix A_1 to have coordinates [-½,0], and A_2 to have coordinates [½,0]. We can rotate and isotropically scale any three points to satisfy this scaling. Because [x,*y*], [-x,*y*], [x,-*y*] and [-x,-*y*] generate the same dyadic distances (and thus the same triadic distance) with A_1 and A_2, it follows that for any isocontour $h(x,y)$, $h(x,y) = h(-x,y) = h(x,-y) = h(-x,-y)$, i.e. we have two axes of symmetry, A_1A_2 and the line through the center orthogonal to A_1A_2.

Section 2 discusses the properties of triadic distances defined directly by the form $f(a_1,a_2,a_3)$ and section 3 discusses optimal properties of center-based triadic distances. Section 4 mentions some sub-optimal triadic distances and Section 5 discusses the relevance of our results and makes some recommendations.

2. Symmetric Functions of a_1, a_2, and a_3

Two symmetric functions that play an important part in the following are:

$$s = a_1^2 + a_2^2 + a_3^2$$

and the area Δ (Heron's form, see e.g. Apostol, 1967):

$$\Delta = \sqrt{r(r-a_1)(r-a_2)(r-a_3)}$$

where $r = ½(a_1+a_2+a_3)$, the semi-perimeter, or equivalently:

$$16\Delta^2 = (-a_1^4 - a_2^4 - a_3^4 + 2a_2^2a_3^2 + 2a_1^2a_3^2 + 2a_1^2a_2^2)\,.$$

We first discuss some basic triadic distances, that have geometric appeal but not necessarily any optimal property. We start with the L_r-transforms and then discuss some other functions of a_1, a_2, and a_3.

2.1 The Perimeter Distance

Perimeter distance is defined as the sum of the dyadic distances, that is:

$$t = a_1 + a_2 + a_3.$$

For all points on A_1A_2 between A_1 and A_2, $t = 2a_3$. This is the minimum distance for the perimeter model. For one-dimensional representations this

property is important, since then the triadic distance is completely determined by the two furthest points of the triple. In our representation, where A_1 and A_2 are fixed at unit distance, the perimeter model has minimum 2 and:

$$a_1 + a_2 = t - 1.$$

That $a_1 + a_2$ is a constant is well known as defining an ellipse with foci A_1 and A_2 (e.g. Wells, 1991, p. 64). The ellipse crosses the x-axis at a point X given by $A_1X = t/2$ and hence X has coordinates $((t-1)/2, 0)$ and the ellipse has a major axis of length $(t-1)/2$. It crosses the y-axis at Y where $a_1 = a_2 = (t-1)/2$ and hence $OY^2 = \left((t-1)/2\right)^2 - \frac{1}{4} = t(t-2)/4$, giving the square of the minor axis. Thus, the ellipse has eccentricity

$$e = \frac{\sqrt{t(t-2)}}{t-1} \text{ and equation: } \frac{4x^2}{(t-1)^2} + \frac{4y^2}{t(t-2)} = 1.$$

The contours shown in Figure 2, are nicely behaved, i.e., they are convex and smooth. As $t \to \infty$, $e \to 1$, showing that the elliptic contours tend to circularity as t increases.

2.2 Generalized Euclidean Distance

For the generalized Euclidean distance:

$$t = \sqrt{a_1^2 + a_2^2 + a_3^2}$$

and thus in our parameterization

$$a_1^2 + a_2^2 = t^2 - 1.$$

From Appendix A3.4, we have that the equation for the isocontour is:

$$x^2 + y^2 = \tfrac{1}{2}t^2 - \tfrac{3}{4},$$

which is a circle, centered at the origin, of squared radius $\frac{1}{2}t^2 - \frac{3}{4}$, shown in Figure 3. As with standard dyadic Euclidean distance, the contour lines are concentric circles. When $t = \sqrt{\frac{3}{2}}$ we achieve the minimal triadic distance and the circle collapses to a single point at the origin. Unlike perimeter distance, generalized Euclidean distance varies for points between A_1 and A_2.

2.3 The Generalized Dominance Distance

The generalized dominance distance is defined by $t = max(a_1,a_2,a_3)$. With our normalization $a_3 = 1$, we have to consider the two situations: (i) $t >$

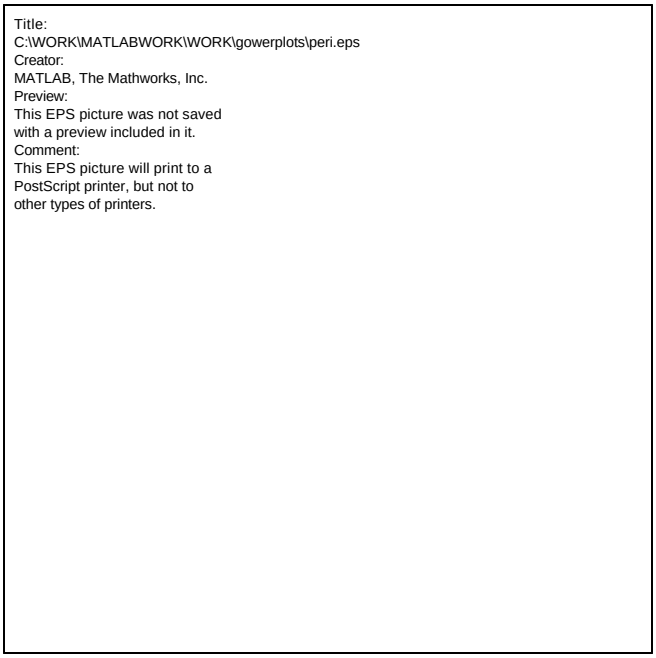

Figure 2. Elliptical contours for perimeter model.

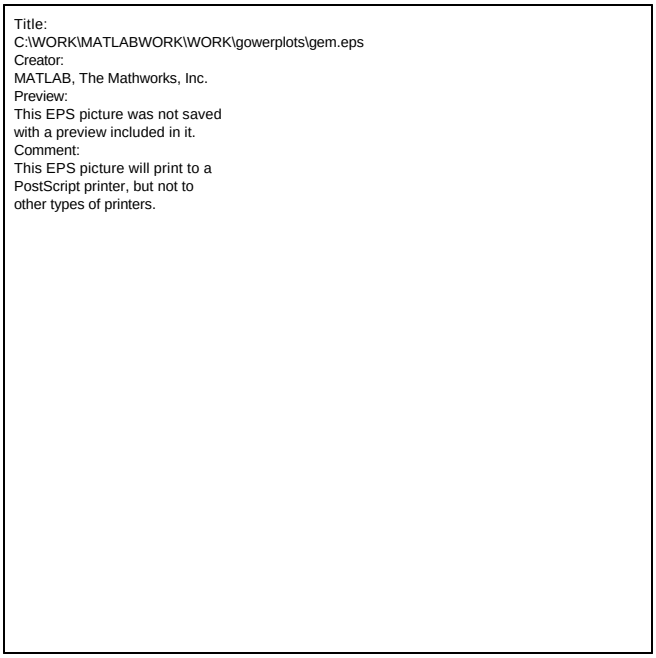

Figure 3. Circular contours for generalized Euclidean model.

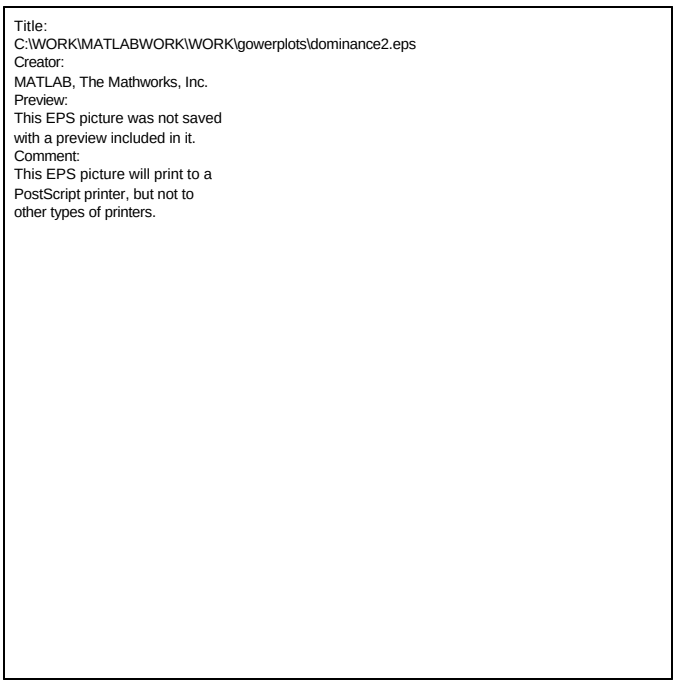

Figure 4. Intersecting circular contours for generalized dominance distance.

$a_3 = 1$ and (ii) $t = a_3 = 1$. Inspection of Figure 4 shows that in case (i) the contour for constant t is given by the intersection of the two circles centered at A_1 and A_2, each with radius t. When A_3 lies within the shaded area of Figure 4, a_3 is the longest side, so that $t = 1$, which is case (ii); the whole of the shaded area corresponds to $t = 1$.

2.4 The Variance Function

Variance is defined by

$$t^2 = var(a_1,a_2,a_3) = (a_1^2 + a_2^2 + a_3^2) - \tfrac{1}{3}(a_1 + a_2 + a_3)^2$$

and may be considered as giving a measure of departure from an equilateral triangle (which has zero variance). Figure 5 shows the contours plotted for values of $3t^2$. The equations and shapes of the contours are quite complicated (see Appendix A1). For $3t^2 < ½$ we have closed pairs of egg-shaped contours; for $½ < 3t^2 < 1½$ the two parts fuse while for $1½ < 3t^2 < 2$ the contours form disjoint sections, including small roughly circular contours around A_1 and A_2; when $3t^2 > 2$ the contours again become continuous, tending to circularity as t increases.

Figure 5. Disjoint and double contour lines for triadic variance distance. For values of $3x^2$ exceeding 1.75 the contours approach an elliptical shape.

2.5 Area

Writing Δ for the area of the triangle $A_1A_2A_3$, we have (from Heron's formula):

$$16\Delta^2 = (a_1+a_2+a_3)(-a_1+a_2+a_3)(a_1-a_2+a_3)(a_1+a_2-a_3).$$

The area is an interesting property of a triangle in its own right; we shall see that it occurs in several of the functions considered in Section 3. With A_1 and A_2 fixed at [-½, 0] and [½ ,0] the area equals $\frac{1}{2}y$, so the isocontours are lines parallel to A_1A_2, together with their reflections. Thus the isocontours are sets of lines parallel to the horizontal; a diagram is unnecessary. When A_3 is above (below) A_1A_2 we may adopt the convention that the area is positive (negative).

2.6 The Product Model

The product model $a_1a_2a_3$ might be considered, partly because of its relationship with loglinear models and also because it occurs as a component

Figure 6. Disjoint and elliptical shaped contours for product distance.

of some of the functions considered in Section 3. We have $t = a_1 a_2$ for which the equation of the isocontour (see Appendix A3.4) is:

$$t^2 = (x^2 + y^2 + \tfrac{1}{4})^2 - x^2 .$$

Figure 6 shows the isocontours (Cassinian ovals, see Wells, 1991, p25). Noteworthy are the figures of eight close to the origin and the disjoint ovals around A_1 and A_2 For larger t the isocontours become elliptical.

3. Triadic Distances Based on Centers

For the generalized Euclidean model it is well known that when the center S is the centroid G, then:

$$t^2 = a_1^2 + a_2^2 + a_3^2 = 3(p_1^2 + p_2^2 + p_3^2)$$

This relationship links the triadic distances defined in terms of a_1, a_2, a_3, and those defined in terms of S and p_1, p_2, p_3. Moreover, we know that G is the center that minimizes $p_1^2 + p_2^2 + p_3^2$.

The emphasis in this section is on the optimal properties of the centers themselves. Of course, $p_1^2 + p_2^2 + p_3^2$ may be evaluated for any center S but only for G does it have the optimal property. We term $t = p_1^2 + p_2^2 + p_3^2$ the *canonical triadic distance* for the center G. Here we have replaced t^2 by t, see section 5 for comment on the consequences of this change. Other centers induce their own canonical triadic distances. Thus, the mediancenter M minimizes $p_1 + p_2 + p_3$ and the circumcenter C minimizes $max(p_1,p_2,p_3)$ The incenter I and the orthocentre H also induce canonical triadic distances: for I, the inradius r minimizes $max(q_1,q_2,q_3)$ the distance from the sides of the original triangle; for H, we have the remarkable property that the pedal triangle has smallest perimeter of all interior triangles, provided the original triangle is acute-angled (see e.g. Wells, 1991, under the entry Billiard Ball Paths in Polygons, p. 14). In general, the canonical triadic distances are defined as minimal values of some salient geometric property of the triangle $A_1A_2A_3$ and we shall require not the individual values of p_1, p_2, p_3 but only the optimal value, t, of each criterion. Like the triadic distances discussed in Section 2 these may be expressed as symmetric functions of a_1, a_2, a_3. These functions can be algebraically complicated but our emphasis is on the geometry, which remains simple. In this section, we concentrate on the canonical triadic distances and content ourselves with some brief remarks in Section 4 on the enormous number of possible non-canonical forms.

The above remarks mainly refer to measures based on p_1, p_2, p_3 principally the minimization of the L_2-norm, the L_1-norm and the minimax criterion, i.e.:

$$p_1^2 + p_2^2 + p_3^2$$

$$p_1 + p_2 + p_3$$

$$max(p_1, p_2, p_3)$$

respectively, but similar functions of q_1, q_2, q_3 and of d_1, d_2, d_3 are also considered.

3.1 Distances Based on p_i

In this section attention is paid to functions of p_i, the distances from a center to the vertices.

3.1.1 The L_2-norm and the Centroid

As already stated, the L_2-norm is minimized by taking the centroid G as center. This case is covered by the generalized Euclidean distance discussed in Section 2.2.

Figure 7. Elliptic shaped contour lines for sum of distances around mediancenter. See Figure A2 for further details.

3.1.2 The L_1-norm and the Mediancenter

For triangles with no angle exceeding 120°, the mediancenter, M, (Haldane, 1948; Gower, 1974) minimizes $p_1 + p_2 + p_3$ with $min(p_1 + p_2 + p_3) = k$ where $k^2 = \frac{1}{2}s + 2\sqrt{3}\,\Delta$ (See appendix A2). When the angle at A_2, say, exceeds 120°, M coincides with A_2 and then $p_1 + p_2 + p_3 = a_1 + a_3$.

From Figure 7 we see that the isocontours for $(p_1 + p_2 + p_3)$ are, at least to the eye, of elliptical shape. Indeed, they are approximated by ellipses with major axis t-½ and minor axis t-$\frac{\sqrt{3}}{2}$. Appendix A2 shows that the reality is more complex and the contour is made up of several sections that join together smoothly. The four main regions are separated by the 120° lines shown in Figure 7. When $t \geq \frac{2}{\sqrt{3}}$ the four sections are all circular, but when $t < \frac{2}{\sqrt{3}}$ there are eight sections, part of the central circular arcs now being replaced by elliptical arcs. These elliptical arcs closely follow the

circular arcs they replace but now the "minor axis" becomes $\sqrt{t^2-\frac{1}{4}}$. The minimal permissible value of t is $t = 1$, when the isocontour collapses to the line A_1A_2 and we attain the maximal difference of $1-\frac{\sqrt{3}}{2} = 0.134$ between the "ellipse" (now a line) and the circle it replaces. Despite their inherent complexity, these contours are very regular with no disconcerting features; they may be compared with the elliptical contours of perimeter distance (section 2.1).

3.1.3 Minimax and the Circumcenter

It is clear that when a solution $p_1 = p_2 = p_3$ is admissible then this must minimize $max(p_1,p_2,p_3)$. Otherwise, it would be possible to shorten the largest of p_1,p_2,p_3 without exceeding either of the shorter lengths. Such a solution is well-known to be possible by choosing S to be the circumcenter C when $p_1 = p_2 = p_3 = R$, the circumradius, where:

$$R = \frac{a_1a_2a_3}{4?}.$$

From Appendix A3.4, the equation of an isocontour, which by definition is circular, is:

$$4R^2y^2 = \left(y^2+x^2+\tfrac{1}{4}\right)^2 - x^2,$$

which factorizes into:

$$\left(x^2+y^2-2y\sqrt{(R^2-\tfrac{1}{4})}-\tfrac{1}{4}\right)\left(x^2+y^2+2y\sqrt{(R^2-\tfrac{1}{4})}-\tfrac{1}{4}\right)=0,$$

representing the upper and lower circumcircles through A_1 and A_2 of radius R. Figure 8 shows these circular isocontours.

Circular contours are easy to interpret but we have to be aware that there are two circles for each value of t. Furthermore, in the area of their intersection, there is very little to distinguish between different values of t (see Figure 8). Circumcenters may be exterior to $A_1A_2A_3$ but this is no problem. Flat triangles will have very large values of R.

3.2 Distances Based on q_i

In this section attention is paid to functions of q_i, the distances from a center toward the sides of the triangle.

Figure 8. Double circular contours for minimax (p_i) and circumcenter.

3.2.1 The L_2-norm and the *N*-center

The minimization of $q_1^2 + q_2^2 + q_3^2$ leads to a new center that we have labeled *N*. Equation (A3.6) of Appendix 3.3 gives:

$$q_1^2 + q_2^2 + q_3^2 = 4\Delta^2\left(\frac{s_1^2}{a_1^2} + \frac{s_2^2}{a_2^2} + \frac{s_3^2}{a_3^2}\right)$$

which on minimization under the constraint $s_1 + s_2 + s_3 = 1$ leads immediately to $s_k = a_k^2/s$. The main characteristic of *N* is that it is at the centroid of the pedal triangle $D_1D_2D_3$; this is shown geometrically in Section 3.3.1. The minimal value of $q_1^2 + q_2^2 + q_3^2$ is $t = 4\Delta^2/s$ which gives an isocontour:

$$2tx^2 + (2t-1)y^2 + \tfrac{3}{2}t = 0\,.$$

Thus, we must have $0 = t < ½$, so the contours are hyperbolic, as is shown in Figure 9.

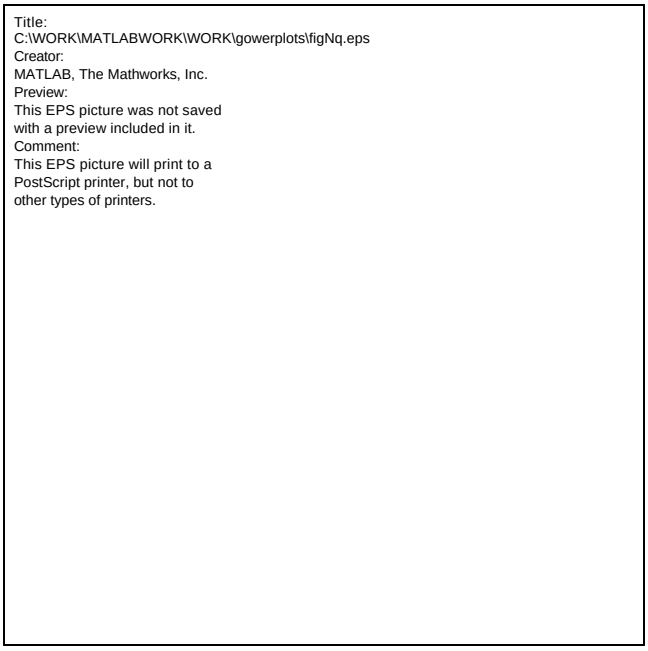

Figure 9. Hyperbolic contours for $\sum q_i^2$ around *N*-center.

3.2.2 The L_1-norm and the *K*-center

From Appendix A3.3, to minimize $q_1 + q_2 + q_3$ is the same as minimizing:

$$2\Delta\left(\frac{s_1}{a_1}+\frac{s_2}{a_2}+\frac{s_3}{a_3}\right)$$

subject to the constraint $s_1 + s_2 + s_3 = 1$. This is a simple linear programming problem with solution $s_i = 1$ when a_i is the longest side, else $s_i = 0$. Thus,

$$t = \frac{2\Delta}{max(a_1, a_2, a_3)}.$$

When $max(a_1,a_2,a_3) = a_2$ it follows from Appendix A3.4 that $t = y/a_2$. This is constant when $sin(A_1)$ is constant, that is on the line A_1A_3. Similarly, when $max(a_1,a_2,a_3) = a_1$, t is constant on the line A_2A_3. However, when $max(a_1,a_2,a_3) = a_3 = 1$, the maximum length side is A_1A_2 and t is constant when y is constant. This state of affairs holds within the intersection of the two circles, both with unit radius, one centered at A_1 and the other at A_2.

Figure 10. Multiple straight line contours for $\sum q_i$ around *K*-center. Each contour has two parts except for those that pass through the intersection of the two unit circles, which have three parts.

Thus the contours are made up of two or three linear segments as shown in Figure 10. The two-line contours occur for $t \geq \sqrt{3}/2$.

3.2.3 Minimax and the Incenter

The incenter is the point *I* equidistant from the three sides of the triangle and hence gives the center that minimizes $max(q_1,q_2,q_3)$. We have that $q_1 = q_2 = q_3 = r$, the inradius, given by:

$$r = \frac{2?}{a_1 + a_2 + a_3}.$$

Figure 11. Hyperbolic shaped contours for Minimax(q_i) around Incenter.

Using Appendix A3.4 shows that the isocontours do not have a simple equation in Cartesian coordinates. However the isocontours are readily computed and are shown in Figure 11.

3.3. Functions of d_i

We continue with functions of d_i, the lengths of the sides of an interior triangle.

3.3.1. The L_2-norm and the *N*-center

We adopt a geometrical approach, referring to Figure 12(i). Consider the set of points F that have the same value of $d_1^2 + d_2^2 + d_3^2$ with D_1 and D_2 as does D_3. This is a circle centered at the midpoint of D_1D_2 which meets A_1A_2 at D_3 and one other point $F\not{c}$, say. Between D_3 and $F\not{c}$ will be points that give smaller values of $d_1^2 + d_2^2 + d_3^2$. It follows that for fixed D_1, D_2 the smallest value of $d_1^2 + d_2^2 + d_3^2$ is obtained when D_3 is chosen to be the point where a concentric circle touches A_1A_2. The normal at this point, being also a radius, passes through the midpoint of D_1D_2; the same applies to the

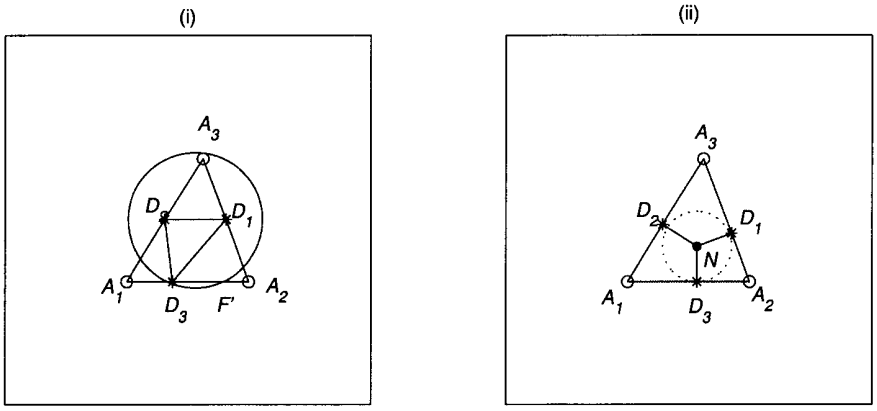

Figure 12. Derivation of the *N*-center. (i) shows any interior triangle D_1, D_2, D_3. $F\not{c}$ is the point that has the same sum-of-square with D_1 and D_2 as has D_3. In the limit (ii) shows the center *N* of a pedal triangle.

normals at D_1 and D_2. The geometry is shown in Figure 12(ii). The normals meet at *N*, the centroid of the triangle $D_1D_2D_3$. It follows that $d_1^2 + d_2^2 + d_3^2 = 3(q_1^2 + q_2^2 + q_3^2)$ so this leads to the same solution as in Section 3.2.1. We have defined D_1, D_2 and D_3 as the vertices of any interior triangle but the above shows that at the minimum, they define a pedal triangle.

3.3.2. The L_1-norm and the Orthocenter

The orthocenter, *H*, is at the intersection of the altitudes from A_1, A_2 and A_3. When one of the angles is obtuse, the orthocenter will lie outside the triangle. For acute angled triangles the pedal triangle of the orthocenter has minimal perimeter among all interior triangles – that is where D_1, D_2 and D_3 are not necessarily given by orthogonal projections of any center. To see this refer to Figure 13(i) in which $D_1D_2D_3$ represents any internal triangle and consider the set of points F that have the same value of $d_1 + d_2 + d_3$ with D_1 and D_2 as does D_3. This is an ellipse with foci at D_1 and D_2 that passes through D_3 and will intersect A_1A_2 at one other point $F\not{c}$, say (compare with Section 2.1). Between D_3 and $F\not{c}$ will be points that give smaller values of $d_1 + d_2 + d_3$. It follows that for fixed D_1, D_2 the smallest value of $d_1 + d_2 + d_3$ is obtained when D_3 is chosen to be the point where a confocal ellipse touches A_1A_2. From the standard property that a normal to an ellipse bisects the directions to its foci, D_1D_3 and D_2D_3 make equal angles with the normal at D_3. The same result must apply for normals at D_1 and D_2. This is a standard result for the pedal triangle associated with the orthocenter, showing that the minimal perimeter occurs when $D_1D_2D_3$ is the pedal triangle of the ortho-

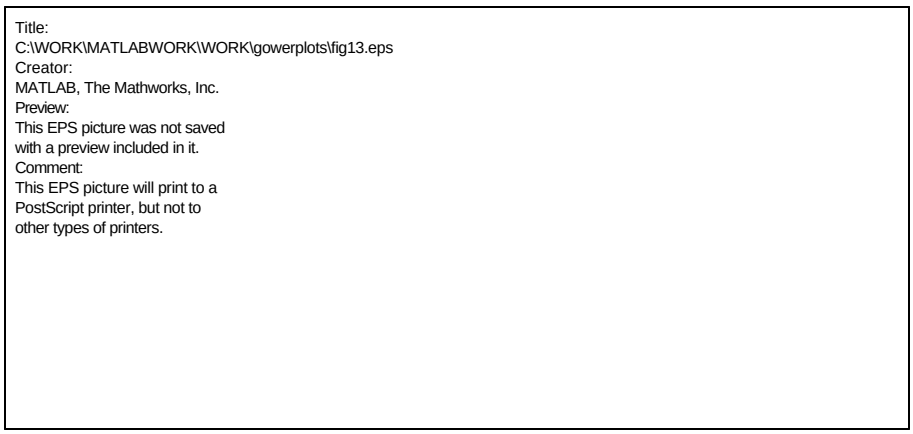

Figure 13. Derivation for the orthocenter having minimal perimeter of interior triangle. (i) shows any interior triangle D_1, D_2, D_3. F¢ is the point that has the same sum of distances with D_1 and D_2 as has D_3. In the limit, not shown, the triangle is pedal with center at the orthocenter, H. In (ii) the angle at A_2 is obtuse and the geometry of (i) has to be modified as described in the text.

center. This derivation breaks down when the triangle has an obtuse angle. Figure 13(ii) shows that the minimum is then given by the pathological pedal triangle with two coincident vertices at the obtuse-angled corner and the other vertex at its projection on the opposite side, giving a minimal value of twice the smallest altitude.

The minimal value of $d_1 + d_2 + d_3$ for an acute-angled triangle may be derived from $d_i = a_i \cos A_i$ to give

$$t = \frac{8\Delta^2}{a_1 a_2 a_3} = \frac{2\Delta}{R},$$

giving isocontours shown in Figure 14 with equation (Appendix A3.4):

$$4y^4 = t^2\left\{\left(y^2 + x^2 + \tfrac{1}{4}\right)^2 - x^2\right\}$$

3.3.3 Minimax and the *J*-center

For the minimax solution, we seek an interior triangle that is equilateral. This case is discussed in detail in Appendix A3.5, where it is shown that the minimax solution defines an equilateral pedal triangle where $s_i = a_i^2\left(\frac{1}{2}s + \frac{2}{\sqrt{3}}\Delta - a_i^2\right)$. We find that the minimum is given by

$$t = D_2D_3 = D_1D_3 = D_1D_2 = 2\Delta/k$$

where $k^2 = \tfrac{1}{2}s + 2\sqrt{3}\,\Delta$, as for the distance around the mediancenter (Sec-

Figure 14. Contours for **?** d_i around Orthocenter. Outside the vertical lines and inside the circle the contours are straight. Otherwise, between the vertical lines the contours are curved. The overall effect is that of hyperbolic-shaped contours.

tion 3.1.2) but now valid for obtuse angles greater than 120º. The equation of a contour is:

$$y^2 = t^2(\sqrt{3}y + (x^2 + y^2 + \tfrac{3}{4}))$$

as shown in Figure 15, but this is valid only for $t^2 < 1$. Then the contours are hyperbolic but only the upper branch of the hyperbola refers to valid triangles. For triangles reflected in A_1A_2 the appropriate contour is the reflection of the corresponding upper branch, not the lower branch.

The results of section 3 are gathered together in Table 1, which lists the canonical triadic distances together with their optimal properties (column 4). Appendix A3 gives the mathematical tools we have found useful in obtaining these results and gives a few examples of their use; we do not give detailed derivations in every case. Recall that if the coordinates of $A_1A_2A_3$ are presented in a 3×2 matrix **X**, any center S may be written in the form $\mathbf{s}'\mathbf{X}$

Figure 15. Contours for Minimax(d_i) around J-center. These are true hyperbolas but the lower half is a reflection of the upper half in the line A_1, A_2. The true lower half of the hyperbola contains impermissible points A_3 and is not shown. This is a double contour.

where the elements of the centering column-vector **s** are simple functions of a_1, a_2 and a_3 and $\mathbf{s}'\mathbf{1} = 1$ (see Appendix A3.1). The different settings of **s** are shown, in unstandardized form, in column 3 of Table 1.

4. Non-canonical Triadic Distances

Section 3 discussed canonical triadic distances that have the optimal properties given in Table 1. Thus, $p_1^2 + p_2^2 + p_3^2$ is minimized when $S = G$. We could easily calculate sums-of-squares about H or $p_1 + p_2 + p_3$ about G but there are very many such non-canonical distances and they do not merit separate discussion. We content ourselves with showing in Figure 16 the isocontours for (i) sums of squares around the orthocenter and (ii) $max(p_i)$ from the mediancenter.

The contours in Figure 16 are bizarre and it is hard to see how they may have any potential practical value.

Table 1. Some optimal properties for centers†

Optimize	Center	s_i‡	Optimum
$p_1^2+p_2^2+p_3^2$	G	1	$\frac{1}{3}\Sigma(a_i-\overline{a})^2$
$p_1+p_2+p_3$	M	$(s-2a_i^2+\frac{4}{\sqrt{3}}\Delta)^{-1}$	$\mathrm{Min}(k, a_1+a_2+a_3-\max(a_1,a_2,a_3))$
$\mathrm{Min}(\mathrm{Max}(p_1,p_2,p_3))$	C	$a_i^2(s-2a_i^2)$	$R=\frac{a_1a_2a_3}{4\Delta}$
$q_1^2+q_2^2+q_3^2$	N	a_i^2	$\frac{4\Delta^2}{s}$
$q_1+q_2+q_3$	K	1 for $\min(a_1,a_2,a_3)$	$\frac{2\Delta}{\mathrm{Max}(a_1,a_2,a_3)}$
$\mathrm{Min}(\mathrm{Max}(q_1,q_2,q_3))$	I	a_i	$r=\frac{2\Delta}{a_1+a_2+a_3}$
$d_1^2+d_2^2+d_3^2$	N	a_i^2	$\frac{12\Delta^2}{s}$
$d_1+d_2+d_3$	H	$(s-2a_i^2)^{-1}$	$\frac{8\Delta^2}{a_1a_2a_3}$
$\mathrm{Min}(\mathrm{Max}(d_1,d_2,d_3))$	J	$a_i^2(\frac{1}{2}s+\frac{2}{\sqrt{3}}\Delta-a_i^2)$	$\frac{2\Delta}{k}$

†G is the centroid; I, the incenter is the center of a circle, radius r, that internally touches the sides of the triangle; C, the circumcenter is the center of the circle, radius R, that passes through the three vertices; and the orthocenter H is the point where the three altitudes meet; M is the mediancenter; K, J, and N are new unnamed centers described in the text. Δ is the area of the triangle, $s=a_1^2+a_2^2+a_3^2$, and $k=\sqrt{\frac{1}{2}s+2\sqrt{3}\Delta}$.

‡We give the unstandardized s_i so that their sum is not equal to 1.

(i)

(ii)

Figure 16. Sum of squares around orthocenter and Max (p_i) around mediancenter. These are examples of non-canonical distances that have some unexpected features.

5. Discussion

We may group contours in various ways. Some contours are single (e.g. Figures 2, 3 and 4) and others are double (e.g. Figures 8 and 16). Double contours occur for a given value of t when one loop reflects into another: e.g. isocontours for a fixed circumradius R. When reflection maps a loop into itself we have single contours: e.g. the perimeter distance. The two loops of double contours necessarily intersect at A_1 and A_2, usually defining a region between A_1 and A_2 where t changes very rapidly. Some isocontours include disjoint sections: e.g. the variance distance (Figures 5) and the product model (Figure 6). Most contours are closed (approximately elliptical or ovoid) but some are open (hyperbolic, or approximately so: e.g. Figures 9, 10, 11, 14 and 15).

Closed single contours are the simplest to interpret because they have similar properties to the familiar circular contours of dyadic distances. Hyperbolic contours admit extreme points that have the same triadic distances with A_1 and A_2 as do less extreme points. However this hyperbolic behavior accurately reflects the properties of the triadic distance adopted and should not necessarily be regarded as a deficiency. We note the analogous properties of representations of asymmetry in terms of area (Gower, 1977; Constantine and Gower, 1978) where zero area with an origin O may be given by two adjacent points or two distant points collinear with O. Other criteria model non-size properties of triadic data. For example, the variance criterion models shape; a zero value implies an equilateral triangle and increasing values departures from equilateralism. In such cases it is better to replace the term *triadic distance* by *triadic shape* and other appropriate terminology. Just as with interpreting diagrams of asymmetry, we have to learn to understand the geometry of such spaces before we can use them for interpretation. Some of our results are a first step in that direction.

The distinction between triadic distance and triadic shape is related to the triadic metric concept. Gower and De Rooij (2003) define a triadic metric to be one that may be expressed as a function of dyadic distances that satisfy the usual dyadic metric property that $d_{ij} + d_{ik} \geq d_{jk}$ for all triads i,j,k. With this definition all the triadic distances discussed in this paper are metric. However, Heiser and Bennani (1997) propose the more stringent definition that $t_{ijl} + t_{ikl} + t_{jkl} \geq 2t_{ijk}$ for all tetrads i,j,k,l. Yet another definition, Joly and Le Calvé (1995), requires that $t_{ijk} + t_{ijl} \geq max(t_{ikl}, t_{jkl})$. The Perimeter, generalized Euclidean and $max(a_i, a_j, a_k)$ definitions satisfy both these definitions provided that the dyadic distances satisfy the triangle inequality. For all other triadic distances we have found counter-examples showing they do not follow the axioms, except for the L_1-norm around the mediancenter which remains undecided. However, although $p_1^2 + p_2^2 + p_3^2$

is clearly not a triadic metric by the definition of Heiser and Bennani, (1997), its square root obviously is; simple transforms of other triadic distances may also satisfy the strict triadic metric conditions. Contours are not affected by making simple transformations, such as square root; only the values of t on the contours change. The same is true of dyadic distances when non-metric scaling methods are used, and the associated monotonic transformations are not sensitive to metric properties. These observations suggests to us that, however defined, the triadic metric properties themselves may not be as relevant as the shapes of the isocontours. We believe that the distinction between closed, hyperbolic and disjoint contours is more important. Thus, we emphasize two points: (i) non-triadic metrics may often be simply transformed into triadic metrics and (ii) satisfaction of the triadic metric property is not necessary for a triadic coefficient to be useful.

We have discussed many triadic distances and their properties but not how to fit them to data. Although some distances are easy to fit (see Gower and De Rooij, 2003), it might be a formidable problem to fit others. In principle, optimization software (e.g. MATLAB's optimization toolbox) may be used to minimize a loss function for any definition of distance. The result of any fitting process is a set of estimates of the underlying dyadic distances. A crucial assumption that lies behind all our isocontours is that triadic distances are functions of dyadic distances. The plots are entirely based on properties of triangles, so the only assumption actually used for drawing the isocontours is that the metric (i.e. triangle) inequality holds. However, if more than three points are plotted, then a full Euclidean representation is envisaged, at least as a good approximation. This may be represented visually by a multidimensional scaling of the estimated dyadic distances (Gower and De Rooij, 2003).

If the triadic model fits well, then the triadic terms may be represented as discussed above and interpreted in terms of our isocontours. For example, Gower and De Rooij (2003) found that the perimeter model fitted very well to Hayashi's (1972) "teams of three" data and in their Figure 4 presented a multidimensional scaling of the underlying dyadic distances. Therefore, a team of three people is represented by a triangle of three points whose perimeter approximates the observed triadic value. Further, if we are interested in how other people may work with the first two members relative to the way that does the third member, then we must construct an ellipse passing through the third vertex with foci at the vertices representing the first two members. Then we can see which other persons lie on, or near this ellipse, and which lie inside (implying a better performance) and which outside (implying a poorer performance). This demonstrates one kind of use of our isocontour results.

Gower and De Rooij (2003) presented evidence suggesting that the estimates of dyadic distances are robust relative to a range of triadic distances. Then, the same multidimensional scaling diagram may be used to inspect the effect of adopting different triadic models. As noted above, the isocontours may be classified into a few general types of shape. This suggests that it may be difficult to discriminate between triadic models that generate the same general shape of isocontour (e.g. the elliptical contours of the perimeter model and the elliptical-shaped contours of the sum of distances from the mediancenter). A corollary would be that one may as well use the most simple model within each class. It is less clear to us how to interpret the concentration of contours within the intersection region of double contours and the disjoint regions of some triadic models. These anomalies occur only for small dyadic distances relative to the base distance; for points remote from this region the approximation to the observed triadic value should be acceptable. Gower and De Rooij (2003) give an example where the variance triadic distance gives good estimates of the dyadic distances in spite of the disjoint contours. It seems that it is only part of the triadic information that may be estimated very imprecisely. Thus a second use of the isocontours is as a mode of thinking about triadic models that highlights properties that need examination.

The symmetry condition implies a special type of model, suitable for special forms of three-way table typified by the teams of three data. We might also fit triadic distances to an $n \times n \times n$ table, in the same way that distances are sometimes fitted to square tables. Furthermore, just as unfolding models can be used to fit distance models to rectangular tables, one might seek to develop methodology where any $p \times q \times r$ table is regarded as a corner of a symmetric $n \times n \times n$ table, most of which is missing (see De Rooij and Heiser, 2000).

Hitherto we have regarded triadic distance as a function of dyadic distance, but there is no formal need for this restriction when triadic terms are parameters of statistical models. Then, distance properties are less important and other forms might be used, although the symmetry condition would be retained when the three classifying factors have no natural ordering and hence are interchangeable. The triadic distance terms discussed above do not include main effects, and refer only indirectly to two-way interactions. Such additional terms may be included in any model; indeed Gower and De Rooij (2003) found it desirable to add a constant to each triadic observation of the teams of three data before fitting a triadic distance. Triadic terms express the three-way interaction as a function of dyadic parameters that must be estimated. This is a very special form of three-way interaction, which may be contrasted with the general three-way interactions

of additive and generalized linear models, or to the three-way multiplicative interaction CANDECOMP model

$$t_{ijk} = \sum_{r=1}^{p} a_{ir} b_{jr} c_{kr}$$

of rank p (Harshman, 1970; Carroll and Chang, 1970). We may consider the generalization: $t_{ijk} = \sum_{r=1}^{p} A_{jkr} B_{ikr} C_{ijr}$, where A_{jkr} is the (j,k)th term of the matrix $\mathbf{A}_r$. Our model can be seen as the special case where (i) $\mathbf{A}_r = \mathbf{B}_r = \mathbf{C}_r$, and are symmetric with elements that may be interpreted as distances, (ii) $p = 1$ and (iii) the product operation is replaced by the triadic functions $f(.,.,.)$. In summary, triadic distances are just one possible term to include in statistical models together with possible main effect and other interaction terms. Triadic distances that adequately model triadic interaction need further development.

We have focused on isocontours for points A_3 with A_1 and A_2 fixed. A further development would be to consider only A_1 fixed and find regions for A_2 and A_3 with constant triadic distance. We observe that A_2 may be moved in a circle, center A_1, and for each point on the circumference A_3 has the desired triadic distance with A_1 and A_2. However, in the above we have $a_3 = 1$, a constant, whereas we can maintain the same triadic distance while varying a_3. To give a complete picture we need the extreme setting of A_2 obtained in this way; this extends the permissible region for A_3. Things can get very complicated and in our opinion little is added to the understanding of triadic distances or triadic shapes. Moreover, the interest is usually in the question of whether a given point has a smaller or larger triadic distance with A_1 and A_2 than with A_3 and then the isocontour lines presented above are relevant.

Finally, we remark on the process we have used for making our isocontours. One would have thought that it would be a simple matter to present the formula for constant t for the different criteria and let a computer contour-plotting program do the hard work. We found this not to be the case, especially in regions where the contours change rapidly or there are singular points. In practice we found it necessary first to do a detailed mathematical analysis for each set of contours to discover the general behavior and whether there were any regions where special care was needed. Then the computer program can be controlled to ensure that interesting details are not smoothed out of existence.

Appendix

A1 The Variance Distance

The variance distance (see Gower and De Rooij, 2003) is defined in section 2 as:

$$t^2 = var(a_1,a_2,a_3)=(a_1^2 + a_2^2 + a_3^2) - \tfrac{1}{3}(a_1 + a_2 + a_3)^2 ,$$

It is more convenient to work in terms of the equivalent form:

$$L = 3t^2 = (a_2 - a_3)^2 + (a_1 - a_3)^2 + (a_1 - a_2)^2 .$$

If we let $u=a_1-a_3$ and $v=a_2-a_3$ then

$$L = (u - v)^2 + u^2 + v^2 ,$$

i.e.

$$L = \tfrac{3}{2}(u - v)^2 + \tfrac{1}{2}(u + v)^2 ,$$

which represents an ellipse with orthogonal axes $u - v = 0$ and $u + v = 0$. The metric constraints $a_1+a_2>a_3$, $a_1+a_3>a_2$ and $a_2+a_3>a_1$ transform into $u + v > -a_3$, $u > v - a_3$ and $v > u - a_3$ which, together with $a_3 = 1$ define the feasible region shown in Figure A1. For L = ½ the ellipses lie entirely within the feasible region. When ½ < L = 1½ then a horse-shoe portion of the ellipse lies within the feasible region, bounded by the line $u + v + 1 = 0$. For 1½ < L = 2 the ellipse is cut into three sections bounded by all three metric constraints, where the main part remains of horse-shoe shape while a small portion is acceptable near u = -1, v = 0 and u = 0, v = -1. At the extreme, when L = 2, the latter portion collapses into two single points. For L > 2, only the horse-shoe portion remains but increasingly has a more circular shape. When the u,v space is transformed back into a_2,a_3 space we arrive at Figure 5 where the same structure is clearly preserved.

A2 The Structure of the Isocontours for the Median L_1 norm.

The mediancenter is the center M that minimizes $t = p_1 + p_2 + p_3$. When all angles of the triangle are less than 120°, then each pair of vertices of the triangle subtend an angle of 120° at M (here termed a *proper mediancenter*). Otherwise, M is at the vertex with the obtuse angle (*improper mediancenter*). Referring to Figure A2, remembering now that the center S is M, with 120° angles at the center, we have for the three smaller triangles:

Figure A1. Derivation of contours for triadic variance distance. The plot is of $a_1 - a_3$ against $a_2 - a_3$. Only the unshaded region pertains to real triangles. The elliptical contours are of constant $3t^2$.

$$\left.\begin{aligned} a_1^2 &= p_2^2 + p_3^2 + p_2 p_3 \\ a_2^2 &= p_1^2 + p_3^2 + p_1 p_3 \\ a_3^2 &= p_1^2 + p_2^2 + p_1 p_2 \end{aligned}\right\}$$

Summing, $$a_1^2 + a_2^2 + a_3^2 = 2P_2 + P_{12}, \qquad \text{(A2.1)}$$

where $P_2 = p_1^2 + p_2^2 + p_3^2$ and $P_{12} = p_2p_3 + p_1p_3 + p_1p_2$. Summing the areas of the triangles gives:

$$4\Delta = \sqrt{3}\ P_{12}. \qquad \text{(A2.2)}$$

Expanding t^2 gives: $$t^2 = P_2 + 2P_{12}. \qquad \text{(A2.3)}$$

Eliminating P_2 and P_{12} from (A2.1), (A2.2) and (A2.3) gives:

$$2t^2 - (a_1^2 + a_2^2 + a_3^2) = 4\sqrt{3}\,\Delta. \qquad \text{(A2.4)}$$

(i)

(ii)

Figure A2. Detailed view on the complexity of L_1-norm with mediancenter. The shaded areas denote region with improper mediancenters and the unshaded areas denote regions with proper mediancenters. In (i) $t > 2/\sqrt{3}$ and the contour has four parts, all circular. In (ii) $t < 2/\sqrt{3}$ and the contour is in eight parts made up of different circles and of two ellipses. In both cases the overall effect is elliptical.

With the normalization $a_3 = 1$ and $A_3(x, y)$ referred to our standard coordinate system (Section A3.4), (A2.4) becomes:

$$x^2 + y^2 + ¾ + \sqrt{3}\, y = t^2,$$

i.e.
$$x^2 + (y + \tfrac{\sqrt{3}}{2})^2 = t^2. \qquad \text{(A2.5)}$$

This is a circle center $E(0, -\frac{\sqrt{3}}{2})$ with radius t. The geometry is shown in Figure A2(i). Denoting $min(p_1 + p_2 + p_3)$ by k, we have from (A2.4) that:

$$k^2 = \tfrac{1}{2} s + 2\sqrt{3}\Delta. \qquad \text{(A2.6)}$$

The lightly shaded area shows all points that generate improper mediancenters with A_1 and A_2; A_1A_2E is an equilateral triangle. The circle, center E, radius t is shown; its circumference includes all points that form triangles with A_1, A_2 with proper mediancenters. However, when M coincides with A_1 (or A_2), the angle $A_3A_1A_2$ (or $A_3A_2A_1$) reaches 120° and any A_3 outside the boundaries defined by EA_1 and EA_2 define triangles with improper mediancenters at A_1 (or A_2). Then $t = a_2 + a_3 = a_2 + 1$. It follows that A_3, for the improper mediancenters, must lie on a circle, radius $t - 1$, center A_1 (or A_2). Returning to the proper mediancenters, from the fundamental property of cyclic quadrilaterals that $A_1M´A_2E + A_1E´MA_2 = ME´A_1A_2$, we see that $A_1M + MA_2 = ME$, so that $t = A_1M + MA_2 + A_3M = A_3E$ and the points A_3, M, E are collinear (see e.g. Wells (1991), under the entry "Fermat Point", a synonym for the mediancenter, for further references). The perpendicular to A_1E at the point where it meets the circles for the proper and improper mediancenters is a tangent to both and hence there is a smooth transition from one circle to the other.

Things are more complicated than just described when the diameter of the circumcircle of A_1A_2E is $\frac{2}{\sqrt{3}} > 1$. Then, if $1 < t < \frac{2}{\sqrt{3}}$ the circle center E, radius t intersects the circumcircle and part of it lies within the shaded area of improper mediancenters. Between the intersections putative positions of A_3 have angles $A_3 > 120°$. Thus, now $t = a_1 + a_2$, which defines an ellipse as in Section 2.1. The equation of this ellipse is as given in Section 2.1 but with t replaced by t +1. It can be shown that at the join of this ellipse to the adjoining circular arcs, there is a common tangent so, again, the join is smooth. The geometry for this case is shown in Figure A2 (ii).

To the naked eye, the contours show no sign of their complex structure. Indeed, from Figure A2(i), we see that we may approximate the isocontours by an ellipse with major axis t - ½ and minor axis $t - \frac{\sqrt{3}}{2}$. This approximation also applies quite well to Figure A2(ii) with maximum deviation when $t = 1$, when the isocontour collapses to the line A_1A_2 giving a

maximal difference of $1-\frac{\sqrt{3}}{2} = 0.134$ between the "ellipse" (now a line) and the circle it replaces.

A3 Tools for Handling Triadic Distances Derived from Centers

In this appendix we present some results that we have found useful in deriving the optimal properties associated with various centers. In some cases we have sought optimal properties that have themselves defined centers.

A3.1 The Centering Vector s

The concept of a distance matrix **D** is basic to the following. **D** is a symmetric $n \times n$ matrix with zero diagonal and element $-\frac{1}{2}d_{ij}^2$ when $i \neq j$. The rows and columns of **D** represent n points and d_{ij} is the Euclidean distance between the ith and jth points. A standard result (Gower, 1982) is that if **s** is any column-vector of length n such that $\mathbf{s}'\mathbf{1} = 1$ and if **X** satisfies the decomposition:

$$\mathbf{XX}' = (\mathbf{I} - \mathbf{1s}')\mathbf{D}(\mathbf{I} - \mathbf{s1}') \tag{A3.1}$$

then, **X** gives a set of coordinates that generate the distances d_{ij} and **X** is centered so that $\mathbf{s}'\mathbf{X} = 0$. In particular, the squared distances of each point from this center are given by $diag(\mathbf{XX}')$ which, when expressed as a column-vector, $\mathbf{p} = (p_1, p_2, \ldots, p_n)'$ may be written:

$$\mathbf{p} = (\mathbf{s}'\mathbf{Ds})\mathbf{1} - \mathbf{2Ds}. \tag{A3.2}$$

In our case $n = 3$ and we write $d_{ij} = a_k$ where $i \neq j \neq k$. Thus:

$$\mathbf{D} = -\frac{1}{2}\begin{pmatrix} 0 & a_3^2 & a_2^2 \\ a_3^2 & 0 & a_1^2 \\ a_2^2 & a_1^2 & 0 \end{pmatrix}$$

And

$$\mathbf{p} = \begin{pmatrix} a_3^2 s_2 + a_2^2 s_3 \\ a_3^2 s_1 + a_1^2 s_3 \\ a_2^2 s_1 + a_1^2 s_2 \end{pmatrix} - (a_1^2 s_2 s_3 + a_2^2 s_1 s_3 + a_3^2 s_1 s_2)\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \tag{A3.3}$$

which gives the squares of the quantities p_1, p_2 and p_3.

A3.2 Geometrical Interpretation of s

Algebraically, **s** gives a set of weights which when attached to the vertices A_1, A_2 and A_3 give the coordinates of S; the geometrical interpretation of **s** is useful and interesting. In Figure 1, we shall write q_k for the angle between SA_i and SA_j, i^1j^1k. Then we have that:

$$\mathbf{XX}' = \begin{pmatrix} p_1^2 & p_1p_2\cos(q_3) & p_1p_3\cos(q_2) \\ p_1p_2\cos(q_3) & p_2^2 & p_2p_3\cos(q_1) \\ p_1p_3\cos(q_2) & p_2p_3\cos(q_1) & p_3^2 \end{pmatrix}$$

$$= \begin{pmatrix} p_1 & & \\ & p_2 & \\ & & p_3 \end{pmatrix} \begin{pmatrix} 1 & \cos(q_3) & \cos(q_2) \\ \cos(q_3) & 1 & \cos(q_1) \\ \cos(q_2) & \cos(q_1) & 1 \end{pmatrix} \begin{pmatrix} p_1 & & \\ & p_2 & \\ & & p_3 \end{pmatrix}. \quad \text{(A3.4)}$$

Now $q_1 + q_2 + q_3 = 2p$, so that:

$$\begin{pmatrix} 1 & \cos(q_3) & \cos(q_2) \\ \cos(q_3) & 1 & \cos(q_1) \\ \cos(q_2) & \cos(q_1) & 1 \end{pmatrix} \begin{pmatrix} \sin(q_1) \\ \sin(q_2) \\ \sin(q_3) \end{pmatrix} = 0.$$

It follows from (A3.4) that:

$$\mathbf{XX}' \begin{pmatrix} p_1^{-1} & & \\ & p_2^{-1} & \\ & & p_3^{-1} \end{pmatrix} \begin{pmatrix} \sin(q_1) \\ \sin(q_2) \\ \sin(q_3) \end{pmatrix} = 0$$

so that $\mathbf{s}'$ is proportional to $\left(\sin(?_1)/p_1,\ \sin(?_2)/p_2,\ \sin(?_3)/p_3\right)$. The area of triangle A_iA_jS is

$$\tfrac{1}{2}(p_1p_2p_3)\frac{\sin(q_k)}{p_k}.$$

It follows that the elements of **s** are proportional to the areas of the triangles A_2A_3S, A_1A_3S, A_1A_2S, which sum to Δ. Hence:

$$\mathbf{s} = \begin{pmatrix} \text{Area } A_2A_3S \\ \text{Area } A_1A_3S \\ \text{Area } A_1A_2S \end{pmatrix} / \Delta \quad \text{(A3.5)}$$

is the centering vector for the coordinates. That is $\mathbf{s}'\mathbf{X}$ gives the coordinates of S. This result is especially useful if a center is geometrically defined, in which case the areas are usually easily derived, and we wish to find the corresponding algebraic form for the elements of $\mathbf{s}$.

A3.3 Pedal Centers and Concurrency

How to calculate the quantities q_1, q_2 and q_3 for pedal triangles follows easily from the observation that the area of triangle A_jA_kS is $\frac{1}{2}a_iq_i$. Thus from the preceding paragraph, we have:

$$q_i = 2? \, s_i/a_i. \tag{A3.6}$$

Because for pedal triangles, A_1D_2 D_3S is a cyclic quadrilateral, it is easy to see that:

$$d_i = p_i \, sin(A_i). \tag{A3.7}$$

Writing $A_2D_1 = l_1a_1$, $A_3D_2 = l_2a_2$ *and* $A_1D_3 = l_3a_3$ then we quote without proof that:

D_1, D_2 and D_3 are the vertices of a pedal triangle if and only if

$$l_1^2a_1^2 + l_2^2a_2^2 + l_3^2a_3^2 = (1-l_1)^2a_1^2 + (1-l_2)^2a_2^2 + (1-l_3)^2a_3^2,$$

which is equivalent to:

$$l_1a_1^2 + l_2a_2^2 + l_3a_3^2 = \tfrac{1}{2}(a_1^2 + a_2^2 + a_3^2) = \tfrac{1}{2}s. \tag{A3.8}$$

Turning now to concurrency, let $\mathbf{s}$ define a center S given by $s_1\mathbf{x}_1 + s_2\mathbf{x}_2 + s_3\mathbf{x}_3$. With the above notation, E_1 has coordinates $l_1\mathbf{x}_3 + (1-l_1)\mathbf{x}_2$, then E_1, S and A_1 are collinear if and only if:

$$det\begin{pmatrix} 0 & 1-l_1 & l_1 \\ s_1 & s_2 & s_3 \\ 1 & 0 & 0 \end{pmatrix} = 0, \text{ i.e. } l_1s_2 - (1-l_1)s_3 = 0.$$

This, with similar expressions for the collinearity of E_2, S, A_2 and E_3, S, A_3 are consistent if:

$$\begin{pmatrix} 0 & -l_1 & (1-l_1) \\ (1-l_2) & 0 & -l_2 \\ -l_3 & (1-l_3) & 0 \end{pmatrix}\begin{pmatrix} s_1 \\ s_2 \\ s_3 \end{pmatrix} = 0 \tag{A3.9}$$

has a nontrivial solution. The condition for this is that the l-matrix has a zero determinant, i.e.

$$l_1 l_2 l_3 = (1-l_1)\,(1-l_2)\,(1-l_3). \qquad \text{(A3.10)}$$

Thus, (A3.10) is the condition for concurrency, which may be compared with (A3.8). When (A3.10) is satisfied, **s** is the null vector of the l-matrix. This null vector is easily constructed from the cofactors of any row of (A3.10). Using ~ to denote "is proportional to", the three rows generate the following possibilities:

$$\mathbf{s} \sim \begin{pmatrix} l_2(1-l_3) \\ l_2 l_3 \\ (1-l_2)(1-l_3) \end{pmatrix} \text{ or } \begin{pmatrix} (1-l_1)(1-l_3) \\ l_3(1-l_1) \\ l_1 l_3 \end{pmatrix} \text{ or } \begin{pmatrix} l_1 l_2 \\ (1-l_1)(1-l_2) \\ l_1(1-l_2) \end{pmatrix}. \qquad \text{(A3.11)}$$

These three versions of **s** must be proportional and any one may be used. However, it would be more elegant to have a simple symmetrical form. One way of proceeding is to multiply (term by term) the three forms and rescale the result. Writing $m_i = l_i/(1-l_i)$, this gives:

$$\mathbf{s}' \sim \left(\frac{m_2}{m_3}\right)^{\frac{1}{3}}, \left(\frac{m_3}{m_1}\right)^{\frac{1}{3}}, \left(\frac{m_1}{m_2}\right)^{\frac{1}{3}} \qquad \text{(A3.12)}$$

with a normalizer N which after some detailed algebraic manipulation is found to be given by:

$$N^3 = \left(\frac{m_2}{m_3} + \frac{m_3}{m_1} + \frac{m_1}{m_2}\right) + 3\left(\frac{1}{l_1} + \frac{1}{l_2} + \frac{1}{l_3}\right) + 3\left(\frac{1}{1-l_1} + \frac{1}{1-l_2} + \frac{1}{1-l_3}\right) - 12$$

The lack of complete symmetry derives from the necessary asymmetric relationships of the l_i and seems to be unavoidable, e.g. A_1 is adjacent to l_3 and to $(1-l_2)$. Thus, when (A3.10) is valid, we have concurrency and then (A3.12) gives the centering vector for the center of concurrency.

A3.4 Coordinates for Isocontours and Trigonometrical Relationships

We have adopted the convention that $a_3 = 1$ and that $A_1 = (-½,0)$ and $A_2 = (½,0)$ and are interested in the locus of $A_3 = [x,y]$. There is a need to

express x and y in terms of a_1 and a_2. From considerations of area we immediately have that:

$$y = 2?.$$

Also:
$$\left.\begin{aligned} a_1^2 &= y^2 + (x - \tfrac{1}{2})^2 \\ a_2^2 &= y^2 + (x + \tfrac{1}{2})^2 \end{aligned}\right\}$$

and hence:
$$x = \tfrac{1}{2}(a_2^2 - a_1^2).$$

Many centers lend themselves to trigonometrical expression. We have chosen to write our results in terms of a_1, a_2 and a_3 (see Table 1). The two forms may be linked by the elementary trigonometrical relationships: (i) the cosine formulae $2a_ia_j\cos(A_k) = a_i^2 + a_j^2 - a_k^2$ and (ii) the sine formulae:

$$\frac{\sin A_1}{a_1} = \frac{\sin A_2}{a_2} = \frac{\sin A_3}{a_3} = \frac{2\Delta}{a_1 a_2 a_3}.$$

A3.5 Example - Minimizing the Maximum Side of an Interior Triangle.

Minimizing the maximum side of an interior triangle implies finding the smallest equilateral interior triangle. An equilateral triangle with smallest perimeter, smallest sums-of squares and smallest area are equivalent criteria. We may construct an interior equilateral triangle as follows. First imagine a point D_1 on A_2A_3 and construct two lines D_1D_2 and D_1D_3 so that the angle $D_2D_1D_3$ is 60° and denote the angle $D_2D_1A_3$ by q; this construction is always possible for some value of q. If $D_1D_2 < D_1D_3$ then, keeping the same angle q, we may slide D_1 towards A_2 until $D_1D_2 = D_1D_3$; $D_1D_2D_3$ is then equilateral with side d, say. Using the notation z_i to denote what was written l_ia_i in A3.3, we then have:

$$\left.\begin{aligned} \frac{z_1}{\sin(60 + q - A_2)} &= \frac{d}{\sin(A_2)} \\ \frac{a_1 - z_1}{\sin(A_3 + q)} &= \frac{d}{\sin(A_3)} \end{aligned}\right\}. \qquad \text{(A3.13)}$$

Eliminating z_1 yields:

$$a_1 = d\left(\frac{\sin(A_3 + q)}{\sin(A_3)} + \frac{\sin(60 + q - A_2)}{\sin(A_2)}\right). \qquad \text{(A3.14)}$$

This is valid for any interior equilateral triangle. Now we may seek the value of q that minimizes d. Differentiation gives:

$$\frac{\cos(A_3+q)}{a_3}+\frac{\cos(60+q-A_2)}{a_2}=0 \qquad \text{(A3.15)}$$

Eliminating d in (A3. 13) gives:

$$\frac{a_1-z_1}{z_1}=\frac{a_2}{a_3}\frac{\sin(A_3+q)}{\sin(60+q-A_2)} \qquad \text{(A3.16)}$$

which with (A3.15), after some manipulation, may be used to evaluate q, giving:

$$\frac{z_1}{a_1}=\frac{a_3(a_3-a_2\cos(60+A_1)}{a_2^2+a_3^2-2a_2a_3\cos(60+A_1)}=\frac{k^2+a_3^2-a_2^2}{2k^2}$$

where $k^2=\frac{1}{2}s+2\sqrt{3}\Delta$ (see Section A2.6). Gathering this result together with the similar results for z_2 and z_3, we have:

$$\left.\begin{aligned}\frac{z_1}{a_1}&=\frac{k^2+a_3^2-a_2^2}{2k^2}\\ \frac{z_2}{a_2}&=\frac{k^2+a_1^2-a_3^2}{2k^2}\\ \frac{z_3}{a_3}&=\frac{k^2+a_2^2-a_1^2}{2k^2}\end{aligned}\right\} \qquad \text{(A3.17)}$$

from which it is easy to show that $a_1z_1 + a_2z_2 + a_3z_3 = \frac{1}{2}s$, the condition required by Appendix (A3.8) for $D_1D_2D_3$ to be a pedal triangle. Thus the minimax equilateral triangle must be a pedal triangle.

Next, we derive the minimal value of d. We do this by finding the areas Δ_1, Δ_2, Δ_3 of the triangles $D_2D_3A_1$, $D_1D_3A_2$, $D_2D_1A_3$, which, when subtracted from Δ, gives the area required. We have: $\Delta_1 = \frac{1}{2}z_3(a_2-z_2)\sin A_1 = \Delta z_3(a_2-z_2)/(a_2a_3)$ which from (A3.17) gives:

$$\Delta_1=\frac{\Delta}{4}\left[\frac{k^4+k^2(a_2^2+a_3^2-2a_1^2)+(a_2^2-a_1^2)(a_3^2-a_1^2)}{k^4}\right].$$

Summing this with the similar formulae for Δ_2 and Δ_3 and subtracting from the total area gives, after a little manipulation, the area of the required equilateral triangle as:

$$\Delta - \Delta_1 - \Delta_2 - \Delta_3 = \frac{\Delta}{4k^4}\left(k^4 - a_1^4 - a_2^4 - a_3^4 + a_2^2a_3^2 + a_1^2a_3^2 + a_1^2a_2^2\right). \quad \text{(A3.18)}$$

From $(k^2 - \frac{1}{2}s)^2 = 12\Delta^2$, we have:

$$k^4 = sk^2 - a_1^4 - a_2^4 - a_3^4 + a_2^2a_3^2 + a_1^2a_3^2 + a_1^2a_2^2.$$

Combining this with the area $\sqrt{3}d^2/4$ of the equilateral triangle, (A3.18) becomes:

$$d^2 = \frac{\Delta}{\sqrt{3}k^4}\left(2k^4 - sk^2\right) = \frac{\Delta}{\sqrt{3}k^2}\left(2k^2 - s\right) = \frac{4\Delta^2}{k^2}.$$

Thus, finally:
$$d = \frac{2\Delta}{k} \quad \text{(A3.19)}$$

a remarkably simple formula for the side of the equilateral triangle. Although k in (A3.19) has the same algebraic form as for the median distance, here it is valid even when $A_1A_2A_3$ contains an angle greater than 120º.

Having shown that the optimal interior equilateral triangle is pedal, it must have a center J, say. We now derive the centering vector $\mathbf{s}$ for which $\mathbf{s}'\mathbf{X}$ gives the coordinates of J. We saw in (A3.2) that this is most easily obtained by finding the areas of A_2A_3J, A_1A_3J, A_2A_1J. Denoting the equal angles JA_1D_2 and JD_3D_2 by α, we have:

(i) $q_2 = p_1 \sin a$, and (ii) $\frac{q_2}{\sin a} = \frac{d}{\sin B}$ and hence:

$$p_1 = \frac{a_2a_3}{k} \quad \text{(A3.20)}$$

and similarly for p_2 and p_3. It follows that:

$$area(A_2A_3J) = area(p_2,p_3,a_1) = \frac{a_1^2}{k^2}\, area(a_3,a_2,k). \quad \text{(A3.21)}$$

This area may be evaluated by expansion but is more easily found by noting that the triangle with sides (a_3,a_2,k) is EA_1A_3 of Figure A2(i), immediately giving the area as $\frac{1}{2}a_2a_3 \sin(A_1+60)$. Thus, $s_1 \sim a_1^2a_2a_3 \sin(A_1+60)$ where the symbol ~ means "up to a factor of proportionality". Thus:

$$s_1 \sim a_1^2(4\Delta + \sqrt{3}(s - 2a_1^2)) \text{ or finally:}$$

$$s_1 \sim a_1^2((\tfrac{2}{\sqrt{3}}\Delta + \tfrac{1}{2}s) - a_1^2)\,. \qquad \text{(A3.22)}$$

The normalizing factor, found by summation, is $\frac{2}{\sqrt{3}}\Delta s + 8\Delta^2$.

References

APOSTOL, T.M. (1967), *Calculus*, Vol. 1, Singapore, John Wiley & Sons, Inc.

CARROLL, J.D. and CHANG, J. (1970). "Analysis of individual differences in multidimensional scaling via an N-way generalization of 'Eckart-Young' decomposition", *Psychometrika, 35*, 283--319.

CARROLL, J.D. and PRUZANSKY, S. (1980), "Discrete and Hybrid scaling models". In: E. D. Lantermann and H. Feger (Eds), *Similarity and Choice*. 108-139, Bern: Hans Huber.

CARROLL, J.D. and PRUZANSKY, S. (1983), "Representing proximities data by discrete, continuous, or 'hybrid' models". In: J. Felsenstein (Eds), *Numerical taxonomy.* 229-248, NY: Springer-Verlag.

CARROLL, J.D. and PRUZANSKY, S. (198), "Discrete and Hybrid models for proximity data". In: W. Gaul and M. Schader (Eds), *Classification as a tool of research*. 47-59, Amsterdam: North-Holland.

CONSTANTINE, A.G. and GOWER, J.C. (1978), "Graphical representations of asymmetry", *Applied Statistics, 27*, 297 -304.

COX, T.F., COX, M.A.A., and BRANCO, J.A. (1991), "Multidimensional scaling or *n*-tuples", *British Journal of Mathematical and Statistical Psychology, 44*, 195-206.

DAWS, J.T. (1996), "The analysis of free-sorting data: Beyond pairwise cooccurences", *Journal of Classification, 13*, 57-80.

DENIS, J-P., GOWER, J.C.(1996), "Asymptotic confidence regions for biaddittive models: Interpreting genotype-environment interactions", *Applied Statistics, 45*, 479 – 493.

DE ROOIJ, M. (2001), *Distance models for transition frequency data*, Unpublished doctoral thesis, Leiden University.

DE ROOIJ, M. (2002), "Distance models for three-way tables and three-way association", *Journal of Classification, 19*, 161-178.

DE ROOIJ, M. and HEISER, W.J. (2000), "Triadic distance models for the analysis of asymmetric three-way proximity data", *British Journal of Mathematical and Statistical Psychology, 53*, 99-119.

DE ROOIJ, M. and HEISER, W.J. (2003a). "A distance representation of the quasi-symmetry model and related distance models". In H. Yanai, A. Okada, K. Shigemasu, Y. Kano, and J.J. Meulman (Eds). *New developments in Psychometrics*. 487- 494. Tokyo: Springer-Verlag.

DE ROOIJ, M. and HEISER, W.J. (2003b, submitted paper), "Graphical representations and odds ratios in a distance-association model for the analysis of cross-classified data".

GOWER, J.C. (1974), "Algorithm AS 78. The mediancentre", *Applied Statistics, 23*, 466 – 470.

GOWER, J.C. (1975), "Generalized Procrustes Analysis", *Psychometrika, 40*, 33 –51.

GOWER, J.C. (1977), "The analysis of asymmetry and orthogonality", In: J. R. Barra, F. Brodeau, G. Romer, and B. van Cutsem (Eds.) *Recent developments in statistics*, 109-123, Amsterdam, North-Holland.

GOWER, J.C. (1982), "Euclidean distance geometry", *The Mathematical Scientist, 7*, 1 –14.

GOWER, J.C., DE ROOIJ, M. (2003), "A comparison of the multidimensional scaling of triadic and dyadic distances", *Journal of Classification*, 20, 115-136.

HALDANE, J.B.S. (1948), "Note on the median of a multivariate distribution", *Biometrika, 35*, 414 –415.

HAYASHI, C. (1972), "Two dimensional quantifications based on the measure of dissimilarity among three elements", *Annals of the Institute of Statistical Mathematics, 24*, 251-257.

HARSHMAN, R.A. (1970). "Foundations of the PARAFAC procedure: model and conditions for an 'explanatory' multi-mode factor analysis". *UCLA working papers in phonetics, 16*, 1.

HEISER, W.J. and BENNANI, M. (1997), "Triadic distance models: axiomatization and least squares representation", *Journal of Mathematical Psychology, 41*, 189-206.

JOLY, S. and LE CALVE, G. (1995), "Three-way distances", *Journal of Classification, 12*, 191-205.

KEMPTON, R.A. (1984), "The use of biplots in interpreting variety by environment interactions", *Journal of Agricultural Science, Cambridge, 103*, 123 - 135.

PAN, G. and HARRIS, D.P. (1991), "A new multidimensional scaling technique based upon association of triple objects P_{ijk} and its application to the analysis of geochemical data", *Mathematical Geology, 23*, 861-886.

TAKANE, Y. BOZDOGAN, H. and SHIBAYAMA, T. (1987). "Ideal point discriminant analysis", *Psychometrika, 52*, 371-392.

TAKANE, Y. (1987). "Analysis of contingency tables by ideal point discriminant analysis", *Psychometrika, 52*, 493-513.

STILLWELL, J. (1998), *Numbers and Geometry*, New York, Springer-Verlag.

WELLS, D. (1991), *The Penguin dictionary of curious and interesting geometry*, Harmondsworth, England, Penguin Books Ltd.