QUADRATIC SURFACE EMBEDDING

JAN DE LEEUW

ABSTRACT. We discuss various primal and dual algorithms for multidimensional scaling with spherical, ellipsoidal, and more general quadratic constraints. Distances between points on the surface can be either measured along Euclidean lines or along shortest geodesics on the surface.

1. INTRODUCTION

Early in the history of Multidimensional Scaling (MDS) its inventors, notably Torgerson [1958] and Shepard [1962a,b], discovered that points in MDS solutions often fell on or close to quadratic manifolds such as circles, ellipses, or parabolas. Some of the first examples analyzed with the new techniques were the color similarity data of Ekman [1954] and the color naming data of Fillenbaum and Rapaport [1971]. Another early application was the triadic comparisons of musical intervals [Levelt et al., 1966], where the points appeared to fall on a parabola (see Shepard [1974, pages 386–388] for some critical discussion). And, around the same time, the triadic comparisons of Dutch political parties [De Gruijter, 1967], which showed a curved left-right dimension, with parties ordered along an ellipse.

Before MDS there was already Guttman's Radex Theory [Guttman, 1954], developed in the context of factor analysis of correlation matrices. It looked for (possibly concentric) circular patterns, called

Date: May 24, 2007.

2000 *Mathematics Subject Classification.* 62H25.

Key words and phrases. Multivariate Analysis, Multidimensional Scaling.

circumplexes, and it found such patters in many observed correlation matrices. See Borg and Groenen [2005, Chapter 5] for a recent review, in which the radex and related patterns are related to Guttman's facet theory.

The fact that quadratic surfaces frequently show up empirically leads to some interesting technical and methodological problems. First, in some cases it may be appropriate to require that the points computed by MDS are indeed located exactly on some parametric surface. If we measuring distances between cities on Earth, for example, an exact spherical or elliptical representation of the cities makes perfect sense [Cox and Cox, 1991]. Second, it may be appropriate for these nonlinear configurations to measure distance as the shortest geodesic on the non-linear manifold. Again, using Eearth as an example, the Euclidean distance, which goes through Earth, may not be as relevant as the geodesic distance measured over the surface of Earth. And third, since a sphere in three dimensions is really two-dimensional, it may be sensible to look for ways to portray non-linear structures in higher dimensions locally faithfully in fewer dimensions. This is basically the classical problem of cartography, in which we compute a suitable projection of the surface of the earth on the plane.

Much more recently, the problem of finding the best MDS representation with points on a sphere came up in computer vision. Ron Kimmel and his group at the Technion in Haifa have published a number of papers in which approximate geodetic distances along arbitrary surfaces are used in MDS with great-circle distances along a sphere. The two-dimensional spherical coordinates and then used to flatten the sphere. A representative paper is Elad et al. [2005].

2. Multidimensional Scaling

In Euclidean MDS[1] we minimize the loss function usually called *stress*, defined as

$$\sigma(X) = \sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}(\delta_{ij} - d_{ij}(X))^2 \tag{1}$$

over $X \in \mathbb{R}^{n\times p}$ [Kruskal, 1964a,b]. Here $\Delta = \{\delta_{ij}\}$ is a given symmetric and hollow[2] matrix of non-negative *dissimilarities*, $W = \{w_{ij}\}$ is a given symmetric and hollow matrix of non-negative *weights*, X is an unknown $n \times p$ *configuration*, and

$$d_{ij}(X) = \sqrt{\sum_{s=1}^{p}(x_{is} - x_{js})^2}$$

is the *Euclidean distance* between *points* x_i and x_j in the configuration X. More generally, we can consider the *constrained case*, in which configurations are restricted to a set $S \subseteq \mathbb{R}^{n\times p}$.

2.1. **Quadratic Constraints.**

In this paper we are interested in the case in which the points in the configuration are constrained to lie on a quadratic surface in $\mathbb{R}^p$. We will look separately at $p = 2$ and $p = 3$, at the special cases of a circle, ellipse, hyperbola, and parabola in $\mathbb{R}^2$, and also at the special cases of a sphere and an ellipsoid in $\mathbb{R}^3$.

We call the technique to place the points on the quadratic MDS with quadratic constraints, abbreviated to MDS-Q. In the most general form of MDS-Q the points x_i must satisfy

$$x_i'\Lambda x_i + 2x_i'\beta + \gamma = 0,$$

[1] In this paper we use the term *multidimensional scaling* (MDS) to refer to *least squares metric multidimensional scaling*.We do not specifically discuss non-metric MDS in this paper, because the algorithms for the metric case can be easily extended to the non-metric case using *alternating least squares*.

[2] A matrix is *hollow* if it has a zero diagonal.

for some $p \times p$ matrix Λ, some p-element vector β, and some constant γ. Because of the invariance of the distance function under translations we can put the center of the surface in the origin. And because distance is invariant under rotation, we can also require without loss of generality that Λ is diagonal. This covers conics (ellipse, hyperbola, parabola) in $\mathbb{R}^2$, and the various kinds of ellipsoids, hyperboloids, paraboloids, and cylinders in $\mathbb{R}^2$. In the case of ellipsoids and hyperboloids we can choose $\beta = 0$ and $\gamma = -1$, so that the constraints become $x_i'\Lambda x_i = 1$. For ellipsoids the matrix Λ is positive definite, which means that we can also write

$$x_i = \Lambda^{1/2} z_i, \text{ where } \|z_i\| = 1 \text{ for all } i. \tag{2}$$

And, of course, spheres are ellipses in which the matrix Λ is scalar, i.e. $\Lambda = \lambda I$.

2.2. **Geodesic Distance.** Besides Euclidean MDS-Q, we also discuss *Quadratic MDS*, or Q-MDS. Again the points are required to lie on a quadratic surface, but now we define distance to be the length of the shortest geodesic along the surface.

This is easy to do in the case of a sphere with radius λ. The metric embedding problem for spherical space is discussed extensively by Blumenthal [1953, Chapter VII]. Define

$$\breve{d}_{ij}(X) = \lambda \operatorname{Arccos}\left(\frac{x_i'x_j}{\lambda^2}\right). \tag{3a}$$

Thus $\breve{d}_{ij}(X)$ is the great-circle distance between the points measured along the sphere. Spherical distance is monotonic with Euclidean distance, because

$$\breve{d}_{ij}(X) = \lambda \operatorname{Arccos}\left(\frac{2\lambda^2 - d_{ij}^2(X)}{2\lambda^2}\right). \tag{3b}$$

The two distance scales are quite different at the higher end, because Euclidean distance between two points on the sphere is bounded

above by 2λ, while spherical distance is bounded by $2\lambda\pi$. This is illustrated in Figure 1. If points are close together the two distances are, of course, approximately equal.

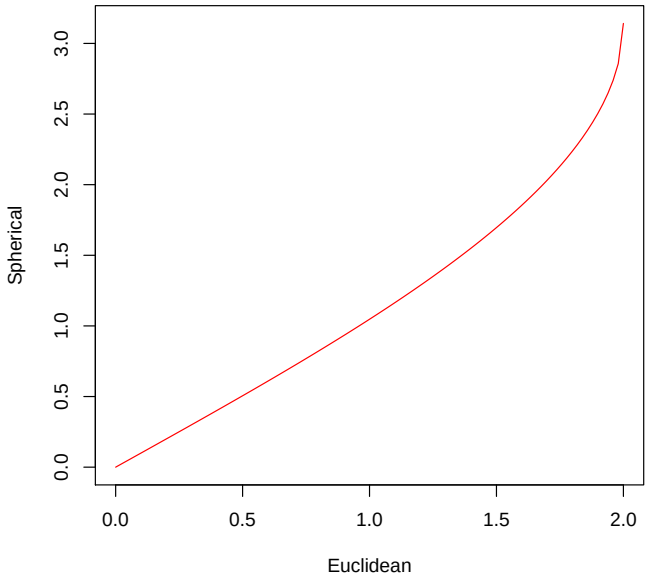


FIGURE 1. Euclidean and Sperical Distance

Unfortunately, matters become more complicated if we go from the sphere to the ellipsoid. In $\mathbb{R}^2$ computing the length of an elliptical arc means evaluating an incomplete elliptic integral of the second kind. Suppose the ellipse is $\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1$. The length of arc from point $(0,1)$ to point $(a\sin\theta, b\cos\theta)$, with $0 \leq \theta \leq \frac{1}{2}\pi$, is

$$\check{d}(\theta) = \int_0^\theta \sqrt{a^2\sin^2 t + b^2\cos^2 t}\,dt.$$

Arc length distances between arbitrary points on the ellipse can be computed by adding and subtracting integral terms of this form. Distance along the ellipse is still monotonic with Euclidean distance, but the relationship is no longer simple.

For a parabola $y = \frac{x^2}{a}$ the situation is somewhat simpler because we can measure along the horizontal axis. We find

$$\begin{aligned}\check{d}(x_0, x_1) &= \frac{1}{2a}\int_{2x_0}^{2x_1}\sqrt{a^2+u^2}du = \\ &= \frac{u}{4a}\sqrt{a^2+u^2} + \frac{1}{4}a\log(u+\sqrt{a^2+u^2})\Big|_{2x_0}^{2x_1}.\end{aligned}$$

Not that this is a nice formula to work with, but at least the integral can be evaluated in closed form.

Computing geodesics on the ellipsoid in $\mathbb{R}^3$ is even more complicated than the ellipse. Since the work of Jacobi and Weierstrass we know what the geodesics on ellipsoids look like, and we know how to compute them, but the analytic expressions are complicated and difficult to work with in an MDS context [Knörrer, 1980; Tabanov, 1996; Perelomov, 2000]. In geodesy there are many programs easily available to do the computations, but for now we have found no easy way to fit geodesics on ellipsoids with MDS. The same thing is true for higher dimensions, and for hyperboloids and paraboloids.

So, summarizing, it looks like MDS-Q may be feasible for any dimension and for any quadratic surface. Q-MDS, however, seems limited to spheres in any dimension, with the possible exception of ellipses in $\mathbb{R}^2$.

3. Previous Work

3.1. **Two-step Methods for MDS-Q.** In order to analyze the quadratic curves in the plane that occurred in Levelt et al. [1966] and in De Gruijter [1967] a quadratic fitting procedure was described by Van De Geer [1968]. The procedure is already explained and used in Levelt et al. [1966, Footnote 1, Page 172]. The method proposed by Van De Geer was also discussed subsequently by Gnanadesikan [1977, Section 2.4], as a form of nonlinear component analysis. In pattern recognition the method is know as *algebraic fitting*.

The method is easy to explain. If we have n pairs of points (x_{i1}, x_{i2}) in the plane, then we form the $n \times 6$ matrix Z with rows

$$z_i = (x_{i1}^2 \mid x_{i2}^2 \mid x_{i1}x_{i2} \mid x_{i1} \mid x_{i2} \mid 1).$$

The n points are on some conic section if there is a v such that $Zv = 0$. To find the best fitting conic we minimize $v'Z'Zv$ over all normalized v. The nornalization used by Van De Geer is $v'v = 1$, which means the optimal v is the eigenvector corresponding with the smallest eigenvalue of $Z'Z$. This, of course, creates the problem of what to do with the other eigenvectors, especially if they have eigenvalues close to the smallest one.

Bookstein [1979] observed that the normalization $v'v = 1$ has the disadvantage that rotations and translations of the set of points will lead to a different fitted quadratic curve. He proposed using the normalization $v_1^2 + v_2^2 + \frac{1}{2}v_3 = 1$. This gives the desired invariance. The algebraic method has a somewhat peculiar way of weighting and combining the residuals. Sampson [1982] improves Bookstein's method by iteratively reweighting the deviations by their approximate standard errors. A third disadvantage of algebraic fitting is that even if we set out to fit an ellipse, the procedure may give us a hyperbola. Fitzgibbon et al. [1999] propose the quadratic normalization $v_1v_2 - \frac{1}{4}v_3^2 = 1$. This guarantees that the matrix of the quadratic form is positive definite, and that consequently the quadratic is an ellipse. It turns out that actually the constraint makes the solution unique, in the sense that the other eigenvectors do not define ellipses. Observe that we could also require $v_1v_2 - \frac{1}{4}v_3^2 = 0$ or $v_1v_2 - \frac{1}{4}v_3^2 = -1$, which would define a parabola or hyperbola. Also note that these quadratic constraints do not imply invariance under translation and rotation.

Gander et al. [1994]

Alternatively, we can normalize v by requiring that it's last element is -1. This amounts to minimizing the loss function

$$\mathcal{F}(\mu, A) = \sum_{i=1}^{n} ((x_i - \mu)' A (x_i - \mu) - 1)^2.$$

Thus the proposed procedure has two steps: first an unrestricted MDS solution is found, and then the best fitting quadratic curve or surface is fitted.

There is a huge literature on fitting ellipses and other quadratic curves and surfaces to sets of points in $\mathbb{R}^p$. The problem has been studied and applied in astronomy, geology, meteorology, computer vision, pattern recognition and statistics. It is difficult to summarize the literature, but we'll give it a try.

3.2. **Primal Methods for MDS-Q.** In earlier work, there has been some discussion on Euclidean MDS with points restricted to lie on a sphere. Initially, constrained MDS concentrated on simple linear constraints on the configuration, familiar from structural equation modeling. Bentler and Weeks [1978] apply Gauss-Newton methods with linear constraints to restrict the primary colors from the example of Fillenbaum and Rapaport [1971] to lie equally spaced at specific places on a circle.

A general theory for least squares Euclidean MDS with restrictions on the configuration was developed by De Leeuw and Heiser [1980]. Also see Borg and Groenen [2005, Chapter 10] for a review with examples.

3.3. **Penalty Methods for MDS-Q.**

3.4. This approach was extended by Lee and Bentler [1980] to nonlinear equality constraints using Gauss-Newton methods with penalty terms to enforce the constraints. In Lee [1984] this is further extended to nonlinear equality and inequality constraints, so

that the configuration can, for instance, be restricted to lie in a circular ring or in a segment of the sphere. The penalty method is replaced by a combination of Gauss-Newton and an augmented Lagrangian method.

In the work of Bentler and his colleagues powerful general purpose optimization methods are used, which are not specifically tailored to MDS. There are both *primal methods*, in which the constraints are incorporated in parametric form directly into the loss function, and *dual methods*, where constraints are imposed at convergence by using penalty or Lagrangian terms.

3.4.1. *CMDA.* A different dual method was proposed by Borg and Lingoes [1979, 1980]. It is also discussed in Borg and Groenen [2005, Section 10.4]. The difference is not to impose restrictions on the configuration, but to impose them directly on the distances. This makes the method more specific to MDS.

For Euclidean MDS with points constrained on a circle or a sphere Borg and Lingoes introduce an extra point x_0 into the MDS problem, and define the family of penalized loss functions

$$\sigma_\kappa(X) = \min_{\Delta\in\mathcal{D}_L} \sigma_L(X,\Delta) + \kappa \min_{\Delta\in\mathcal{D}_C} \sigma_C(X,\Delta) \tag{4}$$

where $\sigma(X)$ is given by (1) and $\tilde\sigma(X,\Delta)$ is given by (1) as well. The set $\mathcal{D}$ consists of all non-negative and hollow symmetric matrices that satisfy the constraints. The non-negative quantity κ is a penalty parameter. If $\kappa\to\infty$ the second term is forced to zero, and we minimize the first term under the conditions that the second term is zero, i.e. that the x_i are on a sphere with center at $x_0$and with radius λ.

The CMDA approach has the advantage it can be implemented quite simply by using the standard Euclidean MDS majorization method (cf. below). It has the usual disadvantage that we have

to select a penalty parameter, or a sequence of penalty parameters, in some way or another. Moreover the Hessian of *stress* will become increasingly ill-conditioned for large penalties, and convergence can consequently be quite slow.

3.5. **Methods for Q-MDS.** The first systematic discussion of Circular and Spherical MDS is in Cox and Cox [1991]. Also see Cox and Cox [1994, Section 4.4.1]. They set out to solve the non-metric Circular and Spherical MDS problems. Since great-circle distance and Euclidean distance are monotonically related, and they are using non-metric MDS, they argue that they can solve the problem by using Euclidean distance. This is true, as long as one realizes that the choice of the distance function does influence how deviations from perfect fit are measured. They then parametrize points on the circle or sphere by using polar coordinates and use a gradient method to minimize *stress*. Polar coordinates become somewhat unwieldy in higher dimensions.

Elad et al. [2005] solve the metric Spherical MDS problem in the context of texture mapping of sphere-like surfaces. They use great-circle distance, and minimize *stress* by a gradient method with line search.

4. Basic Majorization Theory

To define the main aspect of the majorization theory of De Leeuw and Heiser [1980] we need a number of definitions. Let $A_{ij} = (e_i - e_j)(e_i - e_j)'$, where the e_i are unit vectors. Thus e_i has zeroes everywhere, except for element i, which is equal to one. The matrix A_{ij} has zeroes everywhere, except for elements (i,i) and (j,j), which are equal to $+1$, and elements (i,j) and (j,i), which are

equal to -1. Also define

$$B(X) = \sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}s_{ij}(X)A_{ij},$$

$$V = \sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}A_{ij},$$

where

$$s_{ij}(X) = \begin{cases} \frac{\delta_{ij}}{d_{ij}(X)} & \text{if } d_{ij}(X) > 0, \\ 0 & \text{if } d_{ij}(X) = 0. \end{cases}$$

Also define

$$\rho(X) = \mathbf{tr}\ X'B(X)X,$$

$$\eta^2(X) = \mathbf{tr}\ X'VX.$$

Then

$$\sigma(X) = 1 - 2\rho(X) + \eta^2(X).$$

The Cauchy-Schwartz Inequality implies that for all pairs of configurations X and Y we have

$$\rho(X) \geq \mathbf{tr}\ X'B(Y)Y,$$

and thus

$$\sigma(X) \leq 1 - 2\mathbf{tr}\ X'B(Y)Y + \eta^2(X).$$

Define the Guttman transform [Guttman, 1968] of a configuration X as

$$\overline{X} = V^+B(X)X,$$

with superscript $+$ denoting the Moore-Penrose inverse. Then, for all pairs X and Y, we can majorize *stress* at each configuration Y by a quadratic in X. Completing the square gives

$$\sigma(X) \leq \sigma(Y) + \eta^2(X - \overline{Y}) - \eta^2(Y - \overline{Y}).$$

In each iteration of the majorization algorithm we first compute the Guttman transform $\overline{X}^{(k)}$ of our current best configuration, and

then solve a *configuration projection problem* of the form

$$\min_{X\in S}\mathbf{tr}\ (X-\overline{X}^{(k)})'V(X-\overline{X}^{(k)}).$$

In other words, we project $\overline{X}^{(k)}$ on the manifold of constrained configurations. This projection gives us the update $X^{(k+1)}$, with $\sigma(X^{(k+1)})<\sigma(X^{(k)})$.

4.1. **Majorization for CMDACMDS.**

4.1.1. *Sphere.* We can compare this primal gradient projection algorithm with the dual algorithm suggested by Borg and Groenen [2005, Section 10.4]. The majorization algorithm for the penalty function (4) uses the iterations

$$X^{(k+1)}=(V+\kappa I)^{-1}(V\overline{X}^{(k)}+\kappa\tilde{X}^{(k)}),$$

where $\tilde{x}_i^{(k)}=\lambda x_i^{(k)}/\|x_i^{(k)}\|$. For large κ this means

$$X^{(k+1)}\approx\tilde{X}^{(k)}+\frac{1}{\kappa}V(\overline{X}^{(k)}-\tilde{X}^{(k)}),$$

which indicates that convergence will tend to be slow.

4.1.2. *Ellipse.* We know that an ellipse can be defined as the locus of all points that have equal sum of distances from two focal points. Focal points can be chosen on the horizontal axis, at equal distances from the origin. This means that we can fit and ellipse with CMDA by introducing two additional points x_0 and $-x_0$ on the horizontal axis into our MDS problem, and minimize

$$\sigma_\kappa(X,\lambda)=\sigma(X)+\kappa\sum_{i=1}^n(d(x_i,x_0)+d(x_i,-x_0)-\lambda)^2. \tag{5}$$

4.1.3. *Hyperbola.* The hyperbola is the locus of all points that have equal difference of distances from two focal points. Thus, analogous to the case of the ellipse, we minimize

$$\sigma_\kappa(X,\lambda)=\sigma(X)+\kappa\sum_{i=1}^n(|d(x_i,x_0)-d(x_i,-x_0)|-\lambda)^2. \tag{6}$$

4.1.4. *Parabola.* The parabola is the locus of all points that have equal distance to a line (the *directrix*) and a point (the *focal point*). The directrix can be chosen to be the horizontal axis. More generally, for any conic section, the distance to the focal point must be a constant multiple of the distance to the directrix. This multiple, the *eccentricity*, is equal to one for the parabola, large than one for the hyperbola, and less than one for the ellipse.

Suppose x_0 is the focal point, and $\overline{x}_i$ is the projection of x_i on the horizontal axes. Then we must minimize

$$\sigma_\kappa(X,\lambda) = \sigma(X) + \kappa\sum_{i=1}^n (d(x_i,x_0) - \lambda d(x_i,\overline{x}_i))^2. \tag{7}$$

5. Configuration projection

The constraints (2) lead to the problem of minimizing $\mathbf{tr}\ (\lambda Z - Y)'V(\lambda Z - Y)$, over all scalars λ and over all Z with $\mathbf{diag}\ ZZ' = I$. The optimum λ for given Z is

$$\hat{\lambda} = \frac{\mathbf{tr}\ Y'VZ}{\mathbf{tr}\ Z'VZ},$$

and the problem we need to solve is equivalent to the maximization of

$$\rho(Z) = \frac{[\mathbf{tr}\ Y'VZ]^2}{\mathbf{tr}\ Z'VZ}.$$

In order to maximize the function $\rho(Z)$ we use the fractional programming technique of Dinkelbach [1967]. Suppose $\tilde{Z}$ is our current best configuration. Define

$$\eta(Z,\tilde{Z}) = [\mathbf{tr}\ Y'VZ]^2 - \rho(\tilde{Z})\mathbf{tr}\ Z'VZ.$$

If we find Z^+ such that $\eta(Z^+,\tilde{Z}) > \eta(\tilde{Z},\tilde{Z}) = 0$, then $\rho(Z^+) > \rho(\tilde{Z})$. Thus for global convergence it is sufficient to increase $\eta(Z,\tilde{Z})$.

We increase $\eta(Z,\tilde{Z})$ by *block relaxation* [De Leeuw, 1994], i.e. we cycle through all z_i, optimizing over each of the z_i in turn, while keeping the others fixed at their current best values. Various different strategies are possible. We can perform a number of cycles

updating Z, while keeping $\rho(\tilde{Z})$ fixed at its current value. Or we could compute a new $\rho(\tilde{Z})$ after computing the update of each single row z_i of Z. It is unclear which strategy is best, and some numerical experimentation will be useful.

Expanding we find

$$\eta(Z,\tilde{Z}) = z_i'u_iu_i'z_i + 2z_i'h_i + \text{ terms not depending on } z_i,$$

where $U = VY$ and

$$h_i = u_i(\sum_{j\neq i}^n u_j'z_j) - \rho(\tilde{Z})\sum_{j\neq i}^n v_{ij}z_j.$$

Diifferentiate and introduce a Lagrange multiplier θ for the side condition $z_i'z_i = 1$. The stationary equations are

$$(u_i'z_i)u_i + h_i = \theta z_i.$$

Premultiplying both sides by z_i' shows that if there are multiple solutions, we want the one with the largest value of θ.

The stationary equations have the solution $z_i = -(u_iu_i' - \theta I)^{-1}h_i$. Thus if we define

$$\mathcal{F}(\theta) = h_i'(u_iu_i' - \theta I)^{-2}h_i$$

we can find θ by solving the equation $\mathcal{F}(\theta) = 1$. Such equations, often called *secular equations*, have been studied systematically in great detail in numerical mathematics, possibly starting with Forsythe and Golub [1965]; Spjøtvoll [1972]. There are excellent reviews of secular equation theory in Tao and An [1995] and of solvers in Conn et al. [2000, Chapter 7].

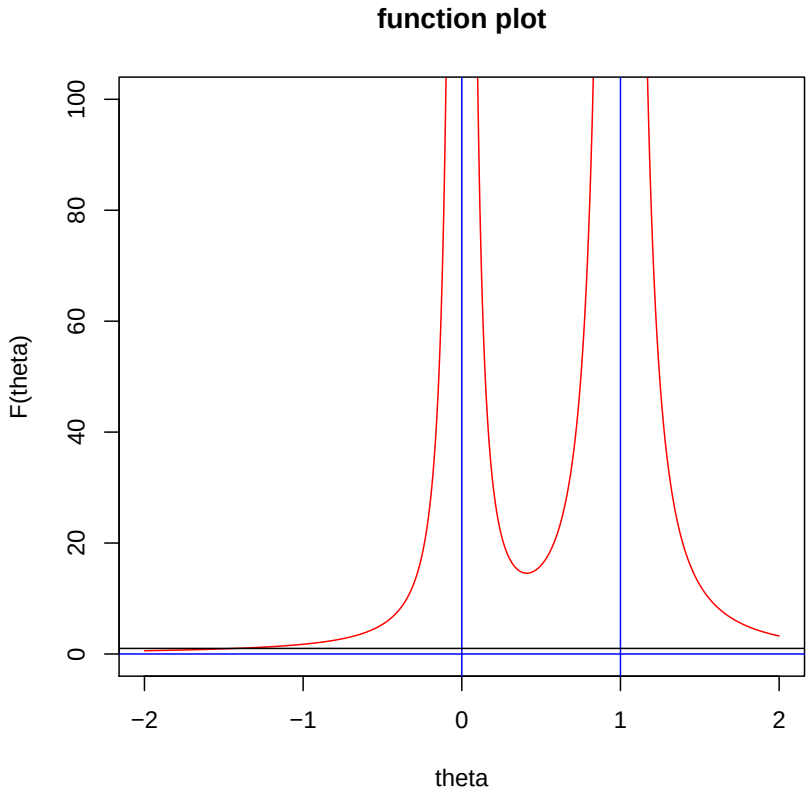
Since we are dealing with a simple special case, we can actually solve the secular equation quite simply. Define $\tau_i = u_i'u_i$, the projector $P_i = u_iu_i'/\tau_i$, and its orthogonal complement $Q_i = I - P_i$. The equation becomes

$$\mathcal{F}(\theta) = \frac{p_i}{(\tau_i - \theta)^2} + \frac{q_i}{\theta^2} = 1,$$

where $p_i = h_i'P_ih_i$ and $q_i = h_i'Q_ih_i$. We can find the solutions by solving the quartic equation

$$\begin{aligned}(8)\quad &\theta^2 p_i + (\tau_i - \theta)^2 q_i - \theta^2(\tau_i - \theta)^2 = \\ &= -\theta^4 + 2\tau_i\theta^3 + (p_i + q_i - \tau_i^2)\theta^2 - 2\tau_i q_i\theta + \tau_i^2 q_i = 0\end{aligned}$$

for its largest real root.



This makes it possible to describe the behavior of the function $\mathcal{F}$ and to show where the relevant roots are located. The plot in Figure 5 is typical. It has $\tau_i = 1$. The function is always positive, it has a horizontal asymptote at zero and two vertical asymptotes, one at zero and the other at τ_i. Between 0 and τ_i it has its unique stationary value, a local minimum, at

$$\hat{\theta} = \frac{\sqrt[3]{q_i}}{\sqrt[3]{q_i} + \sqrt[3]{p_i}}\tau_i,$$

equal to

$$\mathcal{F}(\hat{\theta}) = \frac{(\sqrt[3]{p_i} + \sqrt[3]{q_i})^3}{\tau_i^2}.$$

Thus the equation $\mathcal{F}(\theta) = 1$ has one negative root and one root larger than τ_i. There may or may not be two additional roots, depending on whether the value of the function at the local minimum is smaller than or larger than one. Correspondingly, the quartic (8) has either two real roots (and two complex conjugates) or four real roots (of which two could be equal).

6. Using the Great-Circle Distance

In Spherical MDS we have to minimize

$$\sigma(X) = \sum_{i=1}^n \sum_{j=1}^n w_{ij}(\delta_{ij} - \breve{d}_{ij}(X))^2,$$

where $\breve{d}_{ij}(X)$ is defined by (3a) or equivalently (3b). From (3b), if I am not mistaken,

$$\left.\frac{\partial \breve{d}_{ij}}{\partial d_{ij}}\right|_Y = \frac{d_{ij}(Y)}{\lambda \sin \frac{\breve{d}_{ij}(Y)}{\lambda}},$$

provided, of course that $\breve{d}_{ij}(Y)$ is not equal to zero or to $\lambda\pi$. Write this partial as $g_{ij}(Y)$. Then we can make the Gauss-Newton approximation

$$\delta_{ij} - \breve{d}_{ij}(X) \approx g_{ij}(Y)(\delta_{ij}(Y) - d_{ij}(X))$$

where

$$\delta_{ij}(Y) = \frac{\delta_{ij} - \breve{d}_{ij}(Y) + g_{ij}(Y)d_{ij}(Y)}{g_{ij}(Y)},$$

and thus

$$\sigma(X) \approx \sum_{i=1}^n \sum_{j=1}^n w_{ij} g_{ij}^2(Y)(\delta_{ij}(Y) - d_{ij}(X))^2.$$

This means that if suitable precautions against non-convergence are used, such as choosing a step-size when necessary, we can apply our algorithm for Euclidean MDS with spherical restrictions to

Spherical MDS. In most cases finding the Euclidean spherical solution will provide a good initial estimate for Spherical MDS.

7. Elliptical and More General Quadratic Constraints

Elliptical Constraints lead to the problem of minimizing

$$\eta(Z,\Lambda) = \mathbf{tr}\ (Z\Lambda - Y)'V(Z\Lambda - Y), \tag{9}$$

over all diagonal matrices Λ and over all Z with $\mathbf{diag}\ ZZ' = I$. In the spherical case we minimized out Λ first, and then applied the Dinkelbach trick to get from a ratio to a simple quadratic. In the elliptical case following the same route does not lead to the desired simplifications. But we can use alternating least squares [De Leeuw, 1994]. This produces a different, but closely related algorithm. Clearly this new algorithm could also be applied to the spherical case.

Thus we alternate minimization over Λ for fixed Z and minimization over Z for fixed Λ. The optimum Λ for given Z is simply

$$\hat{\lambda}_s = \frac{y_s'Vz_s}{z_s'Vz_s}.$$

To tackle the problem of finding the optimal Z for given Λ we again minimize over each row of Z separately, using block relaxation. Now

$$\eta(Z,\Lambda) = v_{ii}z_i'\Lambda^2 z_i + 2z_i'h_i + \text{terms independent of } z_i,$$

where

$$h_i = \Lambda^2 \sum_{j\neq i}^{n} v_{ij}z_j - u_i,$$

and $U = VY\Lambda$. As before this gives

$$z_i = -(v_{ii}\Lambda^2 - \theta I)^{-1}h_i,$$

where θ is now chosen such that

$$\mathcal{F}(\theta) = \sum_{s=1}^{p} \frac{h_{is}^2}{(v_{ii}\lambda_s^2 - \theta)^2} = 1.$$

Again $\mathcal{F}$ is positive, with a horizontal asymptote at zero, and with vertical asymptotes at the p poles $v_{ii}\lambda_s^2$. In this case we want the smallest solution $\theta < v_{ii}\min_s \lambda_s^2$. If $p = 2$ we can solve a quartic again, but in general we may have to use one of the very efficient methods to solve secular equations discussed by Melman [1995, 1997, 1998].

8. CODE

```
1  sphereProj<-function(y,v,init=FALSE,immediate=FALSE,ktmax=100,itmax=5,oeps=1e
       -10,ieps=1e-6,iverbose=FALSE,overbose=FALSE) {
2  require("polynom"); n<-nrow(y); nn<-1:n; itel<-1
3  u<-v%*%y; syy<-sum(y*u); tau<-rowSums(u^2)
4  if (is.matrix(init)) z<-init else z<-y/sqrt(rowSums(y^2))
5  repeat {
6  szy<-sum(z*u); vz<-v%*%z; szz<-sum(z*vz); rho<-(szy^2)/szz
7  if (overbose) cat("Iteration:␣",formatC(itel,digits=3,width=3),
8  "Rho:␣",formatC(rho,digits=6,width=10,format="f"),
9  "\n")
10 kmax<-1
11 repeat {
12 for (i in nn) {
13 ui<-u[i,]; zi<-z[i,]; vz<-v%*%z; vi<-vz[i,]; tti<-tau[i]; szy<-sum(z*u)
14 if (immediate) {
15 szz<-sum(z*vz)
16 rho<-(szy^2)/szz
17 }
18 hi<-ui*(szy-sum(ui*zi))-rho*(vi-v[i,i]*zi)
19 sold<-(sum(ui*zi)^2)+2*sum(hi*zi)
20 ppi<-(sum(ui*hi)^2)/tti; qqi<-sum(hi^2)-ppi
21 p0<-qqi*tti^2; p1<--2*tti*qqi; p2<-ppi+qqi-tti^2; p3<-2*tti; p4<--1
22 pol<-polynomial(c(p0,p1,p2,p3,p4)); the<-max(Re(solve(pol)))
23 zi<-(hi/the)-((1/the)+(1/(tti-the)))*(sum(ui*hi)/tti)*ui; z[i,]<-zi
24 snew<-(sum(ui*zi)^2)+2*sum(hi*zi)
25 if (iverbose) cat("Iteration:␣",formatC(itel,digits=3,width=3),
26 "sol:␣",formatC(sold,digits=6,width=10,format="f"),
27 "sne:␣",formatC(snew,digits=6,width=10,format="f"),
28 "\n")
29 }
30 if ((kmax == ktmax) || (ieps<0)) break()
31 kmax<-kmax+1
32 }
33 if ((itel == itmax) || (oeps<0)) return((szy/szz)*z)
34 itel<-itel+1
35 }
36 }
```

References

P.M. Bentler and D.G. Weeks. Restricted Multidimensional Scaling Models. *Journal of Mathematical Psychology*, 17:138–151, 1978.

L.M. Blumenthal. *Theory and Applications of Distance Geometry*. Clarendon Press, Oxford, England, 1953.

F. Bookstein. Fitting Conic Sections to Scattered Data. *Computer Graphics and Image Processing*, 9:56–71, 1979.

I. Borg and P.J.F. Groenen. *Modern Multidimensional Scaling: Theory and Applications*. Springer, Second edition, 2005.

I. Borg and J.C. Lingoes. Multidimensional Scaling with Side Constraints on the Distances. In J.C. Lingoes and E.E. Roskam and I. Borg, editor, *Geometric Representations of Relational Data*. Mathesis Press, Ann Arbor, Michigan, 1979.

I. Borg and J.C. Lingoes. A Model and Algorithm for Multidimensional Scaling with External Constraints on the Distances. *Psychometrika*, 45:25–38, 1980.

A.R. Conn, N.I.M. Gould, and P.L. Toint. *Trust-Region Methods*. MPS-SIAM Series on Optimization. SIAM, Philadelphia, PA, 2000.

T.F. Cox and M.A.A. Cox. *Multidimensional Scaling*. Chapman & Hall, London, England, 1994.

T.F. Cox and M.A.A. Cox. Multidimensional scaling on a sphere. *Communications in Statistics*, 20:2943–2953, 1991.

D.N.M De Gruijter. The Cognitive Structure of Dutch Political Parties in 1966. Report E019-67, Psychological Institute, University of Leiden, 1967.

J. De Leeuw. Block Relaxation Methods in Statistics. In H.H. Bock, W. Lenski, and M.M. Richter, editors, *Information Systems and Data Analysis*, Berlin, 1994. Springer Verlag.

J. De Leeuw and W. J. Heiser. Multidimensional Scaling with Restrictions on the Configuration. In P.R. Krishnaiah, editor, *Multivariate Analysis, Volume V*, pages 501–522, Amsterdam, The Netherlands, 1980. North Holland Publishing Company.

W. Dinkelbach. On Nonlinear Fractional Programming. *Management Science*, 13:492–498, 1967.

G. Ekman. Dimensions of Color Vision. *Journal of Psychology*, 38: 467–474, 1954.

A. Elad, Y. Keller, and R. Kimmel. Texture Mapping via Spherical Multi-dimensional Scaling. In R. Kimmel, N. Sochen, and J. Weikert, editors, *Scale-Space 2005, LNCS 3459*, pages 443–455, Berlin, Heidelberg, New York, 2005. Springer-Verlag.

S. Fillenbaum and A. Rapaport. *Structure in the Subjective Lexicon*. Academic Press, New York, New York, 1971.

A. Fitzgibbon, M. Pilu, and R.Fisher. Direct Least-squares Fitting of Ellipses. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 21:476–480, 1999.

G.E. Forsythe and G.H. Golub. On the Stationary Values of a Second Degree Polynomial on the Unit Sphere. *Journal of the Society for Industrial and Applied Mathematics*, 13:1050–1068, 1965.

W. Gander, G. Golub, and R. Strebel. Least Squares Fitting of Circles and Ellipses. *BIT*, 43:558–578, 1994.

R. Gnanadesikan. *Methods for Statistical Data Analysis of Multivariate Observations*. Wiley, New York, 1977.

L. Guttman. A General Nonmetric Technique for Fitting the Smallest Coordinate Space for a Configuration of Points. *Psychometrika*, 33:469–506, 1968.

L. Guttman. A New Approach to Factor Analysis: The Radex. In P. Lazarsfeld, editor, *Mathematical Thinking in the Behavioral Sciences*. Free Press, New York, N.Y., 1954.

H. Knörrer. Geodesics on the Ellipsoid. *Inventiones Mathematicae*, 59:119–143, 1980.

J.B. Kruskal. Multidimensional Scaling by Optimizing Goodness of Fit to a Nonmetric Hypothesis. *Psychometrika*, 29:1–27, 1964a.

J.B. Kruskal. Nonmetric Multidimensional Scaling: a Numerical Method. *Psychometrika*, 29:115–129, 1964b.

S.Y. Lee. Multidimensional Scaling Models with Equality and Inequality Constraints. *Communications in Statistics*, 13:127–140, 1984.

S.Y. Lee and P.M. Bentler. Functional Relations in Multidimensional Scaling. *British Journal of Mathematical and Statistical Psychology*, 33:142–150, 1980.

W.J.M. Levelt, J.P. Van De Geer, and R. Plomp. Triadic Comparions of Musical Intervals. *British Journal of Mathematical and Statistical Psychology*, 19:163–179, 1966.

A. Melman. A Unifying Convergence Analysis of Second-Order Methods for Secular Equations. *Mathematics of Computation*, 66: 333–344, 1997.

A. Melman. Numerical Solution of a Secular Equation. *Numerische Mathematik*, 69:483–493, 1995.

A. Melman. Analysis of Third-order Methods for Secular Equations. *Mathematics of Computation*, 67:271–286, 1998.

A.M. Perelomov. A Note on Geodesics on an Ellipsoid. *Regular and Chaotic Dynamics*, 5:89–94, 2000.

P. D. Sampson. Fitting Conic Sections to "Very Scattered" Data: An Iterative Refinement of the Bookstein Algorithm. *Computer Graphics and Image Processing*, 18:97–108, 1982.

R.N. Shepard. The Analysis of Proximities: Multidimensional Scaling with an Unknown Distance Function (Part I). *Psychometrika*, 27:125–140, 1962a.

R.N. Shepard. The Analysis of Proximities: Multidimensional Scaling with an Unknown Distance Function (Part II). *Psychometrika*, 27:219–246, 1962b.

R.N. Shepard. Representation of Structure in Similarity Data: Problems and Prospects. *Psychometrika*, 39:373–421, 1974.

E. Spjøtvoll. A Note on a Theorem of Forsythe and Golub. *SIAM Journal on Applied Mathematics*, 23:307–311, 1972.

M.B. Tabanov. New Ellipsoidal Confocal Coordinates and Geodesics on an Eliipsoid. *Journal of Mathematical Sciences*, 82:3851–3858,

1996.

Pham Dinh Tao and Le Thi Hoai An. Lagrangian Stability and Global Optimality in Nonconvex Quadratic Minimization Over Euclidean Balls and Spheres. *Journal of Convex Analysis*, 2:263–276, 1995.

W.S Torgerson. *Theory and Methods of Scaling*. Wiley, New York, 1958.

J.P. Van De Geer. Fitting a Quadratic Function to a Two-dimensional Set of Points. Research Note RN 004-68, Department of Data Theory, University of Leiden, 1968.

Department of Statistics, University of California, Los Angeles, CA 90095-1554

E-mail address, Jan de Leeuw: deleeuw@stat.ucla.edu

URL, Jan de Leeuw: http://gifi.stat.ucla.edu