# FITTING ELLIPSOIDS BY LEAST SQUARES

JAN DE LEEUW

ABSTRACT. The data are $n$ points $y_i$ in $\mathbb{R}^m$. The problem we solve in this paper is to find the least-squares best fitting ellipsoid to these points. An eigenvalue algorithm and an alternating least squares algorithm are presented.

## 1. INTRODUCTION

The problem of fitting an ellipse to a two-dimensional set of of points occurs in quite a few different contexts.

The most obvious one is astronomy, in which orbits of binary stars must be fitted to observational data. In astronomy, the problem has a very long and intricate history, with major contributions by Kepler, Herschel, Thiele and others.

A second area of application is in multidimensional scaling. In various applications of scaling the points are mapped to quadratic manifolds. Colors, for instance, are often represented on a sphere [**?**]. Dutch political parties show a folded elliptical left-right dimension, with the left and the right being bend close together **?**. In correspondence analysis, we very often find a horseshoe structure, which is basically a parabola Gifi [1990]; **?**]; Hill [1974]. Guttman Guttman [1941] has described radex, simplex and circumplex similarity matrices, which also produces quadratic structures.

Pattern recognition

Geology

*Date*: January 25, 2012.

Ellipsoid. Other conics. Spheres.

## 2. Parametrization

2.1. **Algebraic Representation.** An ellipsoid $\mathcal{E}(\mu, V)$, with parameters $\mu$ and $V$, is the set of all $y$ in $\mathbb{R}^m$ such that

$$(y-\mu)'V^{-1}(y-\mu)=1. \tag{1}$$

Here $\mu$ is the *center* of the ellipsoid, and $V$ is a positive semidefinite matrix that determines its *shape*. Clearly we need $m+\frac{1}{2}m(m+1)=\frac{1}{2}m(m+3)$ parameters to uniquely define an ellipsoid in $\mathbb{R}^m$.

The explain the role of $V$, consider the problem of finding the extreme values of $(y-\mu)'(y-\mu)$ over all $y$ that satisfy (1). The maximum will define the largest sphere centered at $\mu$ that intersects the ellipsoid, which is also the smallest sphere centered at $\mu$ that contains the ellipsoid. The minimum defines the smallest sphere centered at $\mu$ that intersects the ellipsoid, which is the largest sphere centered at $\mu$ contained in the ellipsoid.

The stationary equations for this optimization problems are $\tilde{y}=\lambda^2V^{-1}\tilde{y}$, with $\lambda^2$ a Lagrange multiplier, and $\tilde{y}=y-\mu$. It follows that the solutions are given by choosing $\lambda^2$ to be the eigenvalues of $V$ and $\tilde{y}$ the corresponding eigenvectors. Number the eigenvalues as $\lambda_1^2\geq\cdots\geq\lambda_m^2$ and choose corresponding orthonormal eigenvectors $\tilde{y}_1,\cdots,\tilde{y}_m$. Then $y_j=\mu+\lambda_j y_j$ are the solutions to the stationary equations, with corresponding function values $\lambda_j^2$. The $\lambda_j$ are the lengths of the principal axes of the ellipsoid, measured from the center to the surface. The eigenvectors are the direction cosines of the principal axes, with the origin of the space translated to the center of the ellipsoid.

2.2. **Parametric Representation.** Alternatively we can describe an ellipsoid as the image of the unit sphere under a linear transformation. Thus

$$\mathcal{E}(\mu, V) = \{y \mid y = \mu + Tx\} \tag{2}$$

if $x$ varies over all vectors satisfying $x'x = 1$ and $T$ is any matrix with $TT' = V$. One convenient choice for $T$ is $K\Lambda$, where $V = K\Lambda^2 K'$ is the eigen-decomposition (or spectral decomposition) of $V$. Here $K$ is a rotation matrix with $K'K = KK' = I$ and $\Lambda$ is diagonal. Another possibility would be to choose $T$ lower-triangular, using the Cholesky decomposition of $V$.

## 3. Least Squares Using Algebraic Distances

One convenient way to fit an ellipse to observed data $y_i$ is to try to solve the system

$$(y_i - \mu)' V^{-1} (y_I - \mu) = \tau \tag{3}$$

for $\mu$, $V$ and $\tau$. We use $\tau$ instead of 1 for convenience in later computations, after we have found a solution we can always divide both sides by $\tau$ to renormalize the ellipse.

To write system (3) in a more convenient form, define the $(m+1) \times (m+1)$ matrix

$$W = \begin{bmatrix} V^{-1} & -V^{-1}\mu \\ -\mu' V^{-1} & \mu' V^{-1}\mu - \tau \end{bmatrix},$$

and the $(m+1)$ element vector

$$\overline{y} = \begin{bmatrix} y \\ 1 \end{bmatrix}.$$

Observe that we can recover $\tau$, $\mu$, and $V^{-1}$ uniquely from $W$, for instance by using the convenient result

$$W^{-1} = \begin{bmatrix} V - \frac{1}{\tau}\mu\mu' & -\frac{1}{\tau}\mu \\ -\frac{1}{\tau}\mu' & -\frac{1}{\tau} \end{bmatrix}.$$

The system of equations (3) is equivalent to the homogeneous system

$$\overline{\mathcal{Y}}_i' W \overline{\mathcal{Y}}_i = 0,$$

which is linear in $W$. Obviously we can always find the trivial solution $W = 0$, and in order to exclude this we need some kind of normalization to rule this solution out. Of course in practice the points are not exactly on an ellipsoid and the system will not be exactly solvable. Thus we will have to settle for an approximate solution.

In computing least squares solutions we find an approximate solution by minimizing

$$\sigma_{\text{QCA}}(W) = \sum_{i=1}^{n} (\overline{\mathcal{Y}}_i' W \overline{\mathcal{Y}}_i)^2$$

over all $W$ that satisfy some normalization condition.

### 3.1. **LCA.**

### 3.2. **QCA.** **tr** $W^2 = 1$

In this section we discuss an approximation method that was proposed in statistics independently by Gnanadesikan and Wilk [1969] and by Van De Geer [1968]. It was used explicitly by Van De Geer [1968] to fit ellipses to two-dimensional sets of points. See Gnanadesikan [1977, section 2.4] for an accessible review. We shall call it Quadratic Component Analysis (QCA), because it is a straightforward generalization of Principal Component Analysis (PCA) in the way it was introduced originally by Pearson [1901]. QCA leads to the eigenvalue problem

$$\sum_{i=1}^{n} (\overline{\mathcal{Y}}_i' W \overline{\mathcal{Y}}_i) \overline{\mathcal{Y}}_i \overline{\mathcal{Y}}_i' = \lambda W, \tag{4}$$

in which we want the solution corresponding to the smallest nontrivial eigenvalue. In matrix notation this implies that **vec**$(W)$ is an

eigenvector of the $(m+1)^2 \times (m+1)^2$ Kronecker product matrix

$$\sum_{i=1}^{n} (\overline{\mathcal{Y}}_i \overline{\mathcal{Y}}_i') \otimes (\overline{\mathcal{Y}}_i \overline{\mathcal{Y}}_i').$$

We emphasize that we are interested in the smallest *nontrivial* eigenvalue, because any anti-symmetric matrix $W$ defines a solution to (4) with eigenvalue equal to zero. Since $W$ is of order $m+1$, there are obviously $\frac{1}{2}m(m+1)$ linearly independent anti-symmetric solutions. And since all other eigenvectors are orthogonal to these trivial solutions, they lead to a symmetric $W$. So we want the smallest eigenvalue corresponding with a symmetric $W$. This eigenvalue is non-negative, and zero if and only if we have a perfectly fitting ellipse.

QCA is obviously useful if we can attain perfect fit, and it will presumably work well if the errors are small. But in the general case solutions computed by QCA, which work by "ignoring errors", do not necessarily have good statistical properties and by definition cannot introduce measurement errors into the estimating process. Also the choice of the normalization $\mathbf{tr}\ W^2 = 1$ is just for computational convenience, and has no other justification.

The attractiveness of QCA is its simplicity, and the fact that there are no local minima. If, in astronomical applications, we impose constraints on the ellipsoid dictated by Kepler's laws the simplicity disappears. Also observe that there is no guarantee that the best solution actually defines an ellipsoid. The $V$ we find is not necessarily positive semi-definite, and thus we can actually find that we have computed the best fitting hyperboloid. Nevertheless we can expect that QCA will produce good initial estimates for more sophisticated iterative procedures, which we will discuss next.

## 4. Least Squares Using Orthogonal Distances

In this section we use the representation (2) of the ellipsoid as the linear transformation of the unit sphere. We measure badness-of-fit with a least squares loss function that takes measurement errors into account. In astronomy, this is traditiionally known as *Chi-square*. It is defined by

$$\chi^2(X,K,\Lambda,\mu)=\sum_{i=1}^n(y_i-\mu-K\Lambda x_i)'D_i(y_i-\mu-K\Lambda x_i),$$

where the $D_i$ are known $m\times m$ positive definite matrices with the inverse of the measurement errors (not necessarily diagonal). The problem is to minimize $X,K,\Lambda,\mu$ over the free parameters. There are $n\times(m-1)$ free parameters in $X$, $m$ free parameters each in $\mu$ and in $\Lambda$, and $m^2-\frac{1}{2}m(m+1)=\frac{1}{2}m(m-1)$ in $K$.

4.1. **Algorithm.** The basic algorithm is *block relaxation*, i.e. we minimize by cycling over the four sets of parameters, keeping three of them fixed at their current best values while optimizing over the fourth. This produces a decreasing, and consequently convergent, sequence of chi-square values. See De Leeuw [1994] for more details. If the loss function that we minimize is a least squares loss function, then block relaxation is often called *alternating least squares*.

In presenting the algorithm we proceed along a slightly devious path. We first derive it for the simple case in which the $D_i$ are scalar. Thus there exist $\delta_i^2$ such that $D_i=\delta_i^2 I$. Then, in the next section, we adapt to the general case in which $D_i$ are arbitrary positive semi-definite matrices.

QCA can be used to compute initial estimates for $K$, $\mu$, and $\Lambda$. Thus we first study how to compute the corresponding optimal $x_i$, satisfying $x_i'x_i=1$, for given $K$, $\mu$, and $\Lambda$. Clearly each $x_i$ can be computed separately.

The part of the loss function that depends on $x_i$ is simply $x_i'\Lambda^2 x_i - 2x_i' g_i$, where $g_i = \Lambda K'(y_i - \mu)$. Then the optimal $x_i$ must satisfy $(\Lambda^2 - \theta I)x_i = g_i$ as well as $x_i' x_i = 1$. This means we must solve the one-dimensional *secular equation*

$$\psi(\theta) = g_i'(\Lambda^2 - \theta I)^{-2} g_i = \sum_{j=1}^{m} \frac{g_{ij}^2}{(\lambda_j - \theta)^2} = 1$$

for $\theta$ and then set $x_i = (\Lambda^2 - \theta I)^{-1} g_i$.There are many algorithms to efficiently solve such secular equations [Gander, 1981; Gander et al., 1989; Melman, 1997, 1998].

The function $\psi(\theta)$ can be described rather precisely in quantitative terms. Suppose $\lambda_1 < \cdots < \lambda_r$ are the $r$ different elements of $\Lambda$, so possibly $r < m$ if some elements of $\Lambda$ are the same. The function $\psi$ starts at zero at $-\infty$, then increases to $+\infty$ if $\theta$ approaches $\lambda_1$. It comes back from $+\infty$ to the right of $\lambda_1$, decreases to a non-negative minimum between $\lambda_1$ and $\lambda_2$, and increases again to $+\infty$ as $\theta$ approaches $\lambda_2$ from the left. The same applies to all intervals between any two consecutive eigenvalues, the function will always have a non-negative valley in that interval, with the slopes of the valley increasing to $+\infty$ on both sides. Finally the function decreases from $+\infty$ to zero if $\theta$ moves from $\lambda_r$ to $+\infty$.

It is easy to see we want the smallest root of $\psi(\theta) = 1$, which is the one on the branch $\theta < \lambda_1$. On that branch the function $\psi$ is both increasing and convex, which makes it easy to find the root.

Define

$$\overline{x} = \frac{\sum_{i=1}^n \delta_i^2 x_i}{\sum_{i=1}^n \delta_i^2},$$
$$\overline{y} = \frac{\sum_{i=1}^n \delta_i^2 y_i}{\sum_{i=1}^n \delta_i^2}.$$

It is easy to see that the optimal $\mu$ for given $X$, $K$, and $\Lambda$ is

$$\mu = \overline{y} - K\Lambda\overline{x}.$$

To compute the optimal $\Lambda$ for given $X$, $K$, and $\mu$ we define

$$\overline{A} = \frac{\sum_{i=1}^n \delta_i^2 x_i x_i'}{\sum_{i=1}^n \delta_i^2},$$
$$\overline{B} = \frac{\sum_{i=1}^n \delta_i^2 y_i x_i'}{\sum_{i=1}^n \delta_i^2}.$$

The part of the loss function depending on $\Lambda$ can be written as $\mathbf{tr}\ \Lambda^2\overline{A} - 2\mathbf{tr}\ \Lambda K'(\overline{B} - \mu\overline{x}')$ and thus the optimal $\Lambda$ is the diagonal matrix

$$\Lambda = \mathbf{diag}^{-1}(\overline{A})\mathbf{diag}(K'(\overline{B} - \mu\overline{x}')).$$

Observe that only the diagonal elements of the matrices $A$ and $K'(\overline{B} - \mu\overline{x}')$ need to be computed. Also, it is easy at this stage to incorporate constraints on $\Lambda$. We can require, for instance, that its elements must be non-negative, or that some elements are zero (which means we require the ellipsoid to lie in a subspace of $\mathbb{R}^m$).

The optimal $K$ for given $X$, $\mu$ and $\Lambda$ is found by maximizing the linear function $\mathbf{tr}\ K'(\overline{B} - \mu\overline{x}')\Lambda$ over $K'K = I$. This is a classical Procrustus problem [Borg and Groenen, 2005, Chapter 20]. If $Z = (\overline{B} - \mu\overline{x}')\Lambda$, then the optimal $K$ is given by $Z(Z'Z)^{-\frac{1}{2}}$, using the inverse of the symmetric square root of $Z'Z$. Alternatively, if $Z = P\Xi Q'$ is the singular value decomposition of $Z$, then the optimal $K$ is $PQ'$.

It is clear that we can cycle through the four subproblems in various ways, as long as we design our iterative strategy in such a way that eventually each subproblem will be solved a large number of times. Some guidance on the various strategies is in De Leeuw and Michailides [in preparation].

## 5. Majorization

In general, the $D_i$ will not be scalar. In fact, they may not even be diagonal. But we can use *majorization* [De Leeuw, 1994; Heiser,

1995; Lange et al., 2000] to reduce the problem to one with a scalar weight matrix.

We discuss this trick in general terms. Suppose we want to minimize the least squares loss function $\sigma(f) = (y-f)'W(y-f)$ over $f \in F$. Suppose $\tilde{f} \in F$ is our current best guess. Expand the loss as

$$\sigma(f) = [(y-\tilde{f}) - (f-\tilde{f})]'W[(y-\tilde{f}) - (f-\tilde{f})].$$

If $\gamma > 0$ is any constant larger than or equal to the largest eigenvalue of $W$ we have

$$\begin{aligned} \sigma(f) &\leq \sigma(\tilde{f}) - 2(y-\tilde{f})'W(f-\tilde{f}) + \gamma(f-\tilde{f})'(f-\tilde{f}) = \\ &= \sigma(\tilde{f}) + \gamma(f-\tilde{y})'(f-\tilde{y}) - \gamma(\tilde{y}-\tilde{f})'(\tilde{y}-\tilde{f}), \end{aligned} \tag{5}$$

where

$$\tilde{y} = \tilde{f} + \frac{1}{\gamma}W(y-\tilde{f}). \tag{6}$$

If $f^+ \in F$ is such that

$$(f^+-\tilde{y})'(f^+-\tilde{y}) < (\tilde{f}-\tilde{y})'(\tilde{f}-\tilde{y})$$

then we have from (5) that $\sigma(f^+) < \sigma(\tilde{f})$. Thus, basically, we can solve weighted least squares problems of any form by constructing pseudo-data using (6) and then making one or more steps of an unweighted least squares algorithm to fit these pseudo-data.

In the ellipsoid fitting example, in which we choose $\gamma_i$ to be the largest eigenvalue of $D_i$, the pseudo-data are

$$\tilde{y}_i = (\tilde{\mu} + \tilde{K}\tilde{\Lambda}\tilde{x}_i) + \gamma_i^{-1}D_i(y_i - (\tilde{\mu} + \tilde{K}\tilde{\Lambda}\tilde{x}_i)).$$

Of course if $D_i$ is actually scalar, the pseudo-data are equal to the data. The majorization algorithm now alternates computation of the pseudo-data with the alternating least squares to fit the blocks of parameters. Many variations are possible depending for instance on how many alternating least squares cycles we perform before computing new pseudo-data.

## 6. Applications in Astronomy

Kepler's laws for the two-body problem tell us the smaller body is rotating around the larger body in an ellipse. There are three important complications with the applications of our methods to fitting ellipses in astronomy.

6.1. **Ellipses in Three-space.** The first complication is not at all essential. In astronomy the true orbits of the smaller body around the larger one are two-dimensional ellipses imbedded in three-dimensional space. This is not a problem for the QCA method, which will continue to give a perfect solution in the case in which there exists a perfect solution. And it is not a problem for the DLS method, because it is trivial to require the diagonal matrix $\Lambda$ to have only two non-zero elements.

6.2. **True and Apparent Orbit.** The second complication is that we do not observe the true orbit of the body, but only the apparent orbit, which is the projection of the true orbit on the observation plane. Again, this will be an ellipse. The relation between the two ellipses is illustrated in Figure 6.2, which is taken from Tatum [2005].

We could use our methods to fit an ellipse to the apparent orbit, but obviously we are interested in the parameters describing the true orbit. We can suppose, without loss of generality, that the observation plane is horizontal and that the centroid of the true orbit is in the observation plane. Also, of course, the errors of measurement apply to the points in the apparent orbit. This means that the (majorized) DLS loss function becomes

$$\chi^2(X,K,\Lambda,\mu)=\sum_{i=1}^n\delta_i^2(y_i-S(\mu+K\Lambda x_i))'(y_i-S(\mu+K\Lambda x_i)),$$

where $S$ is a $2\times 3$ binary matrix that selects the first two coordinates from a three-dimensional vector.

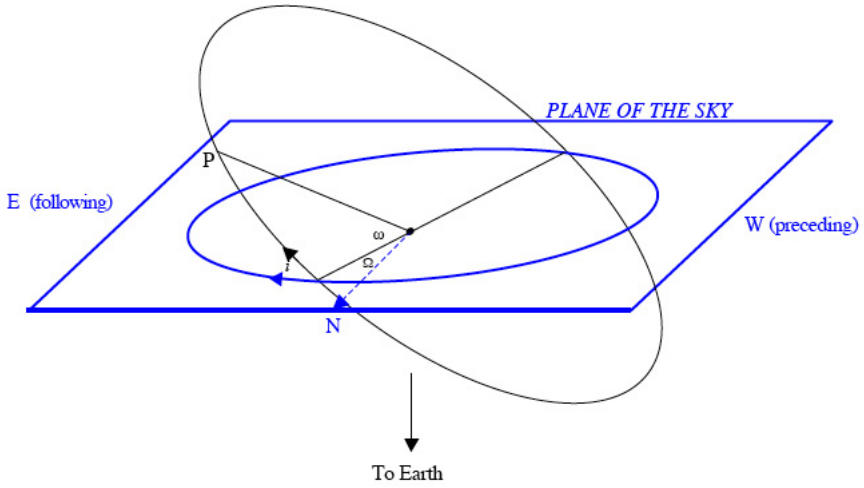


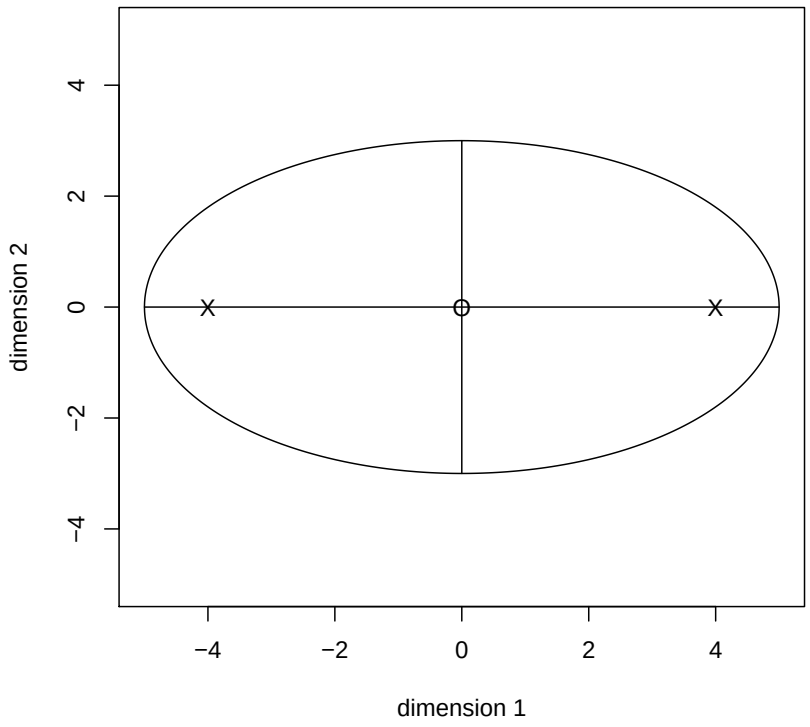

FIGURE 1. True and Apparent Orbit

The DLS method can be adapted by using *augmentation* [De Leeuw and Michailides, in preparation]. We introduce an additional unknown $\eta_i$ for each $i$, which is the third coodinate of $y_i$. The augmented $y_i$ are now written as $\tilde{y}_i$ and we minimize

$$\chi^2(X,K,\Lambda,\mu,\eta) = \sum_{i=1}^{n} \delta_i^2(\tilde{y}_i - \mu + K\Lambda x_i)'(y_i - S\mu + K\Lambda x_i),$$

in the usual way, treating $\eta$ as another block of variables in the alternating least squares algorithm. Obviously the optimal $\eta_i$ for given $X$, $K$, $\Lambda$, and $\mu$ is the third element of $\mu + K\Lambda x_i$. Remember we keep $\mu_3 = 0$ throughout for identification purposes, while $\Lambda$ also has one element restricted to zero

6.3. **Kepler's Laws.** And thirdly, and most importantly, Kepler's laws impose constraints on the parameters of the ellipses. And astronomical observations are not just static points $y_i$ in space, they are also ordered in time. Clearly both for purposes of interpretation and to improve the quality of the estimates it will be essential to take these constraints into account.




## Appendix A. Ellipses in Two Dimensions

Let us start with the simple ellipse $\Lambda x$, with $\lambda_1 > \lambda_2$. The center is in the origin, and the length of the *semi-major axis* is $\lambda_1$, the length of the *semi-minor axis* is $\lambda_2$. Figure A illustrates the case with $\lambda_1 = 5$ and $\lambda_2 = 3$.

We have also indicated the two *foci* of the ellipse, which are at $(\pm\sqrt{\lambda_1^2 - \lambda_2^2}, 0)$, in the example at $(\pm 4, 0)$. The *eccentricity* of the ellipse is

$$e = \frac{\sqrt{\lambda_1^2 - \lambda_2^2}}{\lambda_1}.$$

One of the foci is the *primary focus*, which is the point where the object being orbited resides. The point on the orbit closest to the primary focus is the *periastron*, the distance from the periastron

to the primary focus is clearly $\lambda_1 - \sqrt{\lambda_1^2 - \lambda_2^2}$. The *semi-latus rectum* is the distance from a focus to the ellipse, measured along a line perpendicular to the major axis. This is equal to

$$\ell = \frac{\lambda_2^2}{\lambda_1}.$$

The temporal description of an orbit uses the *period* $P$, which is the time for one revolution, and a point in time $T$ where the satellite passes through the periastron. If the mass $M$ of the object being orbited is known, then *Kepler's third law* says that $\lambda_1^3 = P^2 M$.

## APPENDIX A. CODE

```
makeData<-function(a=2,m=c(1,2),g=diag(c(2,1)),u=0,v=1,s
    =.1) {
t<-seq(u,v,by=s)
v<-matrix(c(sin(t),cos(t)),length(t),2)
k<-makeK(a)
l<-makeL()
t((m+k%*%g%*%t(v)))
#t(l%*%(m+k%*%g%*%t(v)))
}

makeL<-function() {
l<-matrix(rnorm(6),3,2)
e<-eigen(crossprod(l))
l<-l%*%e$vectors
d<-colSums(l^2)
l/outer(c(1,1,1),sqrt(d))
}

makeK<-function(a) {
matrix(c(sin(a),-cos(a),cos(a),sin(a)),2,2,byrow=TRUE)
}

perturb2<-function(mat,sigma) {
n<-dim(mat)[1]
mat+(matrix(rnorm(2*n),n,2)*outer(rep(1,n),sigma))
}

project<-function(mat) {
return(mat[,c(1,3)])
}

gpca<-function(y){
```

```
n<-dim(y)[1]; m<-dim(y)[2]; mm<-m+1; mmm<-mm*mm
cc<-matrix(0,mmm,mmm)
for (i in 1:n) {
  yi<-c(y[i,],1); c<-outer(yi,yi)
  cc<-cc+kronecker(c,c)
  }
ee<-eigen(cc)
for (i in mmm:1) {
  w<-matrix(ee$vectors[,i],mm,mm)
  w<-(w+t(w))/2; if (sum(w^2)>1e-6) break()
}
return(w)
}

drawEllipse<-function(mu=c(0,0),a=0,lambda=c(1,1),n=1000)
    {
k<-matrix(c(cos(a),sin(a),-sin(a),cos(a)),2,2)
x<-seq(0,2*pi,length=1000); ff<-k%*%c(sqrt((lambda[1]^2)-(
    lambda[2]^2)),0)
f1<-mu+ff; f2<-mu-ff
y<-mu+k%*%rbind(lambda[1]*sin(x),lambda[2]*cos(x))
yup<-max(y); ylo<-min(y)
plot(t(y),type="n",xlim=c(ylo,yup),ylim=c(ylo,yup),xlab="
    dimension 1",ylab="dimension 2")
lines(t(y)); text(matrix(mu,1,2),"O")
text(matrix(f1,1,2),"X"); text(matrix(f2,1,2),"X")
lines(matrix(c(mu-lambda[1]*k[,1],mu+lambda[1]*k[,1]),2,2,
    byrow=TRUE))
lines(matrix(c(mu-lambda[2]*k[,2],mu+lambda[2]*k[,2]),2,2,
    byrow=TRUE))
}
```

## References

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

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 G. Michailides. *Block Relaxation and Majorization Algorithms in Statistics*. Springer, in preparation.

W. Gander. Least Squares with a Quadratic Constraint. *Numerische Mathematik*, 36:291–307, 1981.

W. Gander, G.H. Golub, and U. Von Matt. A Constrained Eigenvalue Problem. *Linear Algebra and Its Applications*, 114/115:815–839, 1989.

A. Gifi. *Nonlinear Multivariate Analysis*. Wiley, Chichester, England, 1990.

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

R. Gnanadesikan and M.B. Wilk. Data Analytic Methods in Multivariate Statistical Analysis. In P.R. Krishnaiah, editor, *Multivariate Analysis*, volume II, New York, 1969. Academic Press.

L. Guttman. The Quantification of a Class of Attributes: A Theory and Method of Scale Construction. In P. Horst, editor, *The Prediction of Personal Adjustment*, pages 321–348. Social Science Research Council, New York, 1941.

W.J. Heiser. Convergent Computing by Iterative Majorization: Theory and Applications in Multidimensional Data Analysis. In W.J. Krzanowski, editor, *Recent Advanmtages in Descriptive Multivariate Analysis*, pages 157–189. Oxford: Clarendon Press, 1995.

M.O. Hill. Correspondence Analysis: a Neglected Multivariate Method. *Applied Statistics*, 23:340–354, 1974.

K. Lange, D.R. Hunter, and I. Yang. Optimization Transfer Using Surrogate Objective Functions. *Journal of Computational and Graphical Statistics*, 9:1–20, 2000.

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

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

K. Pearson. On Lines and Planes of Closest Fit to Systems of Point is Space. *Philosophical Magazine (6)*, 23:559–572, 1901.

J.B. Tatum. Celestial Mechanics, 2005. URL `http://astrowww.phys.uvic.ca/~tatum/celmechs.html`.

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`