Procrustes Analysis and Geodetic Sciences

Fabio Crosilla

Abstract

Procrustes analysis is a well known technique to provide least squares matching of two or more factor loading matrices or for the multidimensional rotation and scaling of different matrix configurations. Applied at first as a useful tool in factor analysis, today it has become a popular method of shape analysis (Goodall 1991, Dryden and Mardia 1998).
This paper reviews the development of the most significant algorithms used in this particular field. Starting from the solution of the classical "orthogonal procrustes problem"(Schönemann 1966) a first extension including a scaling factor and a central dilation will be presented (Schönemann and Carroll 1970).
The solution of the "generalized orthogonal procrustes problem" to sets of more than two matrices will be then reported (Gower 1975, Ten Berge 1977).
Furthermore, "weighted procrustes analysis" will be considered for the cases in which the residuals of a matching procedure are differently weighted across columns (Lissitz et al. 1976) or across rows (Koschat and Swayne 1991) of a matrix configuration .
Finally, some possible applications of procrustes methods for point coordinates transformations in geodesy and photogrammetry will be mentioned. All this makes it possible to emphasize the capabilities of the method proposed.

1. Unweighted procrustes analysis

The so-called "orthogonal procrustes problem" (Schönemann, 1966) is the least squares problem that makes it possible to transform a given matrix A into a given matrix B by an orthogonal transformation matrix T in such a way to minimize the sum of squares of the residual matrix $E = AT - B$, that is

$$tr\,(E'\,E) = min,$$

under the orthogonal condition for matrix T, that is $T'T = TT' = I$. In order to satisfy the minimum condition, a Lagrangean function, defined as

$$F = tr\,(E'\,E) + tr\,[\,L\,(T'T - I)\,],$$

where L is a matrix of Lagrangean multipliers, must be minimized, by setting the partial derivative of F with respect to T equal to zero, that is

$$\partial F/\,\partial T = (\,A'A + A'A)\,T - 2\,A'B + T\,(L + L') = 0 \tag{1.1}$$

Putting $A'A = P$, $A'\,B = S$ and $(L+ L')/2 = Q$, one can note that matrices P and Q are symmetric, so that multiplying (1.1) on the left by T', it results that

$$Q = T'S - T'\,P\,T = Q'.$$

Now, since $T'\,PT$ is symmetric, $T'S$ must be symmetric too; In this way

$$T'S= S'T. \tag{1.2}$$

Multiplying (1.2) on the left by T

$$S = T\,S'T$$

and on the right by T'

$$T'\,S\,T' = S'$$

we finally have

$$SS' = TS'\,ST'. \tag{1.3}$$

Matrices S'S and SS' are symmetric matrices, both of which can be transformed in a diagonal form by orthonormal matrices and both of which have the same eigenvalues, according to Schönemann et al. (1965). Equation (1.3) can be finally written as

$$W\,D_S\,W' = T\,V\,D_S\,V'T'$$

so that W= TV or

$$T = WV'$$

Matrix T is the orthogonal matrix which satisfies the least squares principle defined by tr (E'E) = min. A first generalization to the Schönemann (1966) orthogonal procrustes problem was given by Schönemann and Carroll (1970) when a least squares method for fitting a given matrix A to another given matrix B under choice of an unknown rotation, an unknown translation and an unknown central dilation was presented. The chosen model is the following

$$B = c\,AT + J\,g' + E \tag{1.4}$$

where J' = (1111...1), A, B and T have the same meaning as before, g is a vector of translation components, c is a scalar of central dilation and E is a matrix of residuals. To obtain the least squares solution for model (1.4), the Lagrangean function, written as

$$F = \mathrm{tr}\,(E'\,E) + \mathrm{tr}\,[\,L\,(T'T - I)\,]$$

where

$$\mathrm{tr}\,(E'\,E) = \mathrm{tr}\,B'B + c^2\,\mathrm{tr}\,T'A'\,AT + p\,g'g - 2\,c\,\mathrm{tr}\,B'\,AT - 2\mathrm{tr}\,B'J\,g' + 2\,c\,\mathrm{tr}\,T'A'Jg'$$

and p is the number of rows of matrices A and B, must be differentiated with respect to the unknowns T, g and c. Once the derivatives are set at zero, it results that

$$\partial F/\,\partial T = 2\,c^2\,A'A\,T - 2\,c\,A'B + 2\,c\,A'\,J\,g' + TQ = 0 \tag{1.5}$$

where Q = (L + L')/2 = Q' is an unknown symmetric matrix,

$$\partial F/\,\partial g = 2p\,g - 2\,B'J + 2c\,T'A'J = 0 \tag{1.6}$$

$$\partial F/\,\partial c = 2\,c\,\mathrm{tr}\,T'A'AT - 2\,\mathrm{tr}\,B'AT + 2\,\mathrm{tr}\,T'A'J\,g' = 0 \tag{1.7}$$

From equation (1.6) it follows that

$$g = (B - c\,AT)'\,J/p \tag{1.8}$$

and from (1.5), according to some considerations already done about symmetry of matrices T'A'AT and Q, it results that

$$T'A'B - T'A'Jg' = \text{symmetric matrix}$$

which can also be written, according to (1.8) ,as

$$T'A'B - T'A'(JJ'/p)(B - cAT) = \text{symm.}$$

so that

$$T'A'(I - JJ'/p)\, B = \text{symm.} \tag{1.9}$$

because cT'A'(JJ'/p)AT is symmetric. Equation (1.9) does not consider g and c and a solution for T can be found following the already mentioned procedure applied for the "orthogonal procrustes" problem.

Letting $$S^{*}= A'(I - JJ'/p)\, B,$$

matrix products $S^{*\prime}S^{*}$ and $S^{*}S^{*\prime}$ will be computed, and from their singular value decomposition

$$S^{*\prime}S^{*} = V\, D_s V'$$
$$S^{*}S^{*\prime} = WD_s W'$$

the solution for T = WV' can be found. Substituting (1.8) in (1.7), equation (1.7) can be solved for the contraction factor c

$$c= \text{tr}\; T'A'(I- JJ'/p)\, B/ \;\text{tr}\; A'(I - JJ'/p)\, A \tag{1.10}$$

Inserting the result for c in equation (1.8) the value of the estimated g can be finally obtained.

One interesting thing to note is that the matrix of best fit $\hat{B}$ and that of the residuals E= B - $\hat{B}$ given respectively by

$$\hat{B} = cAT + Jg' = c\, AT + (JJ'/p)(B - c\, AT) = (JJ'/p)B + c(I - JJ'/p)\, AT$$

and by

$$E = B - \hat{B} = (I- JJ'/p)(B- c\, AT)$$

do not involve g. The fit is the same independent of the origin of both data set configurations. For this reason, from the practical point of view in terms of computation, the first recommended step to take consists of the calculation of the two column centered matrices

$$A^{*} = A - J\, k' \quad \text{and}$$
$$B^{*} = B - J\, h'$$

where k = A'J/p and h = B'J/p.

Afterward one has to enter a standard orthogonal procrustes algorithm to obtain the transformation matrix T and the matrix A^*T. The final computations are related to the scalar c, given by

$$c = \text{tr} [(T' A^{*\prime}) B^*] / \text{tr} A^{*\prime} A^*$$

and the matrix of best fit

$$\hat{B} = c (A^*T) + J h'.$$

To measure the least squares fit, the criterion function itself, that is

$$e = \text{tr} E'E = \text{tr} B'(I - JJ')/p B - (\text{tr} T'A' (I - JJ')/p B)^2 / \text{tr} A'(I- JJ')/p A$$

is commonly adopted in the literature. This is however not symmetric in the sense that a fit of A to B generates different values for the elements of e than a fit of B to A. In order to satisfy a symmetry condition, Lingoes and Schönemann (1974) defined at first a new symmetric measure of fit given by

$$e_s = e u^{-1/2}$$

where $$u = \text{tr} B' (I - JJ'/p) B/ \text{tr} A' (I - JJ'/p) A$$

which satisfy e_s (A,B) = e_s (B,A),and does not depend on the order of matching and u is a constant. As reported by Lingoes and Schönemann (1974) e_s depends upon the norm of the target matrix B. Such dependency must be avoided when the comparison of the fits for different target matrices is wanted. To achieve the desired scale invariance, the following measure was finally proposed

$$S = e / \text{tr} B'(I - JJ'/p) B$$

In this case the measure remains invariant for different orders of fitting and is 0 - 1 bounded.

2. Generalized Procrustes analysis

A generalization of the classical procrustes analysis was given by Gower (1975) and Ten Berge (1977) where the problem of best fitting more than two matrices was taken into account. Instead of considering the matching of all possible independent matrix pairs, the procrustes analysis is generalized in such a way that m matrices are simultaneously subjected to similarity transformations until a proper fit criterion is reached. The criterion adopted consists in the minimization of the sum of square distances between each point of the m ones $P_j^{(i)}$ (i = 1...m) belonging to the same cluster and their centroid G_j (j= 1...n), summed for all n clusters. The problem of rotating, translating and scaling m matrices (m ≥ 2) toward a best least squares fit consists in finding orthonormal matrices T_i (i= 1...m), translation vectors t_i, and scale factors c_i, for which the function

$$S = \text{tr} \sum_{i<j}^{m} [(c_i A_i T_i + t_i) - (c_j A_j T_j + t_j)]' \; [(c_i A_i T_i + t_i) - (c_j A_j T_j + t_j)] \qquad (2.1)$$

is minimized (Gower 1975).
The Gower's method starts with an initial centering of the matrices, so that all column sums are zero and a successive scaling of each A_i by a general $w^{1/2}$ so that $\sum$ w tr $A'_i A_i$ = m.

In the following, with the symbol A_i, column centered and scaled matrices will be considered. Rotation matrices and scaling constants are adjusted in sequence. The solution of the rotation problem consists of finding orthonormal matrices T_i for which the function

$$f(T_1...T_m) = \sum_{i<j} \mathrm{tr}\ (A_iT_i - A_jT_j)'\ (A_iT_i - A_jT_j)$$

is minimized or equivalently the function

$$g(T_1...T_m) = \sum_{i<j} \mathrm{tr}\ T_i'\ A_i'\ A_jT_j$$

is maximized. Ten Berge (1977) suggests the following iterative procedure to solve the problem:

Step 1: Rotate A_1 to $\sum_{j=2}^{m} A_j$, thus yielding $A_1T_1^{(1)}$

Step 2: Rotate A_2 to $A_1T_1^{(1)} + \sum_{j=3}^{m} A_j$, thus yielding $A_2T_2^{(1)}$

Step m: Rotate A_m to $\sum_{j=1}^{m-1} A_jT_j^{(1)}$, thus yielding $A_mT_m^{(1)}$

Step m+1 : Rotate $A_1T_1^{(1)}$ to $\sum_{j=2}^{m} A_jT_j^{(1)}$, thus yielding $A_1T_1^{(2)}$

The procedure terminates when the combined effect of some steps does not raise g above a certain threshold value. The procedure will converge and g will be maximized if, and only if, all the matrix products

$$T_i'A_i'A_{i+1}$$

are symmetric and positive semi-definite . Proof of the theorem is reported in Ten Berge 1977, page 269. Let us now consider the problem of computation of the scaling constants c_i. Let it be $\sum$ tr $A_i'A_i$ = m for which scaling constants $c_1....c_m$ are wanted to maximize

$$h(c_1....c_m) = \Sigma_{i<j}\ c_ic_j\ \mathrm{tr}\ A_i'A_j \tag{2.2}$$

with the constraint

$$\Sigma c_i^2\ \mathrm{tr}\ A_i'A_i = \quad \mathrm{tr}\ A_i'A_i = m \tag{2.3}$$

that satisfies the condition of maximizing (2.2) and minimizing the least squares function

$$\Sigma_{i<j}\ \mathrm{tr}\ (c_iA_i - c_jA_j)'\ (c_iA_i - c_jA_j).$$

If we consider the particular rescaled matrix A_i^* as

$$A_i^* = (\mathrm{tr}\ A'_i\ A_i)^{-1/2}\ A_i$$

it follows that tr $A_i^{*\prime} A_i^* = 1$, for i= 1...m. We look for m scalars d_i (i= 1..m), able to maximize

$$h^* (d_1...d_m) = \Sigma_{i<j} d_i d_j \text{ tr } A_i^{*\prime} A_j^*, \tag{2.4}$$

with the constraint

$$\Sigma d_i^2 \text{ tr } A_i^{*\prime} A_i^* = \Sigma d_i^2 = \Sigma c_i^2 \text{ tr } A_i' A_i = \Sigma \text{ tr } A_i' A_i = m \tag{2.5}$$

Now let the mxm matrix Y be written as

$$Y = \begin{bmatrix} trA_1' A_1 & trA_1' A_2 & .... & trA_1' A_m \\ .... & .... & .... & .... \\ trA_m' A_1 & trA_m' A_2 & .... & trA_m' A_m \end{bmatrix}$$

and Y_D = diag (Y), and $F = Y_D^{-1/2} Y Y_D^{-1/2}$.

Putting d_i in a vector d, than condition (2.4) can be written as

$$h^*(d) = 1/2 \, d' (F - I) d \tag{2.6}$$

which must be maximized subject to

$$d'd = m.$$

Considering the singular value decomposition of matrix F, (F= PLP'), and letting z = P' d it follows that

$$h^*(d) = 1/2 \, d' (F - I) d = 1/2 \, d' (PLP' - I) d =$$

$$= 1/2 (z'Lz - m) \leq 1/2 (l_1 z'z - m) = 1/2 \, m (l_1 - 1)$$

where l_1 is the greatest eigenvalue of matrix L. Condition (2.6) is maximized when $d = m^{1/2} p_1$. In this case

$$h^*(m^{1/2}p_1) = 1/2 \, m \, p_1' (F-I)p_1 = 1/2 \, m \, p_1' (PLP' - I) p_1 =$$

$$= 1/2 \, m(e_1'L e_1 - p_1'p_1) = 1/2 \, m (l_1-1).$$

From this result and the constraint (2.5) it follows that

$$c_i = (m / \text{tr } A_i' A_i)^{1/2} p_i \qquad \text{(Ten Berge (1977)).}$$

Up to now the so-called unweighted least squares solutions have been taken into account. This is appropriate when the residuals have equal variance and hence should be weighted equally. If one wishes to weight the residuals differently, a weighted least squares criterion is more appropriate.

3. Weighted procrustes analysis

Two different ways of weighting the residuals are usually applied in the procrustean literature: across columns or across rows. The corresponding least squares criteria are then:

$$\mathrm{tr}(AT - B)' W_n^2 (AT - B) \tag{3.1}$$

$$\mathrm{tr}(AT - B) W_p^2 (AT - B)' \tag{3.2}$$

where W_n and W_p are diagonal weight matrices containing information about the dispersion of the residuals.
To find an orthogonal matrix T minimizing (3.1) is easy, since it is equivalent to minimize (3.1) replacing B by W_n B and A by W_n A, respectively (Lissitz et al. 1976). The second problem is more difficult to solve. A very interesting algorithm was introduced by Koschat and Swayne (1991). The algorithm is based on the possibility of considering the general problem like a specific one for which it is simple to find a valid solution. If matrix A is characterized by orthogonal column vectors characterized by the same euclidean lenght l, the problem is to find an orthogonal matrix T that minimizes (3.2), that is

$$\begin{aligned}\mathrm{tr}(B - AT)W_p^2 (B-AT)' &= \mathrm{tr}(B W_p^2 B') - 2\mathrm{tr}(ATW_p^2B') + \mathrm{tr}(W_p^2T' A'AT)\\ &= \mathrm{tr}(B W_p^2 B') - 2\mathrm{tr}(W_p^2B'AT) + l^2 \mathrm{tr}(W_p^2)\end{aligned}$$

or equivalently, that maximizes

$$\mathrm{tr}(W_p^2 B' AT),$$

which is very similar to the solution of the classical unweighted least-squares problem, previously reported. Writing the singular value decomposition of $W_p^2B'A$ as U L V' the solution can be found as

$$T = VU' \tag{3.3}$$

In case the column vectors of A are not orthogonal with respect to each other, a connection with the case reported above can be made in the following way. Once the nxp matrices A and B are given, and the (n+p)xp matrices A* and B* are defined as

$$A^* = \begin{bmatrix} A \\ A^0 \end{bmatrix} \qquad\qquad B^* = \begin{bmatrix} B \\ B^0 \end{bmatrix}$$

one has to fix matrix A° so that

$$A^{*\prime} A^* = l^2 I_p, \text{ for some } l\,. \tag{3.4}$$

Of course the matrix A* satisfies the condition (3.4) if and only if A° satisfies

$$A^{\circ\prime}A^\circ = l^2 I_p - A'A.$$

In order to satisfy a positive-definite right hand-side of this equation, a sufficiently large l must be fixed. In this case an infinite number of solutions for A° are possible. Koschat and Swayne (1991) suggested the use of the Cholesky decomposition

$$A°' = \mathrm{chol}\,(l^2 I_p - A'A)$$

where l^2 is set equal 1,1 times the largest eigenvalue of A'A. The pxp matrix B° can be chosen arbitrarily. The algorithm reported by Koschat and Swayne (1991), allows definition of a sequence (T_i, B_i*), and is of the form: for i = 2..., define the (n+p)xp matrix B_i* as

$$B_i{}^* = \begin{bmatrix} B \\ A^0 T_{i-1} \end{bmatrix}$$

and the corresponding weighted procrustes residuals

$$\mathrm{tr}\,(B_i{}^* - A^*T)\, W_p{}^2\, (B_i{}^* - A^*T)' \qquad (3.5)$$

As A* is characterized by orthogonal columns of equal lenght, matrix T_i (i= 1,2,3…) reported in formula (3.5), can be computed at each iteration by formula (3.3). The solution method of this problem can be successfully applied in image analysis and in photogrammetry where, if matrices A and B contain the centred coordinates of corresponding control points on two different images or in two different reference systems of coordinate, it is desired to test whether the object described by A can be transformed into the object described by B through rotation and dilation along specified directions.
The problem may be formulated as a regression problem

$$B= AX + E$$

under the constraint

$$X = TK$$

where T'T = TT' = I and K is diagonal with positive values. The solution can be obtained by minimizing

$$\mathrm{tr}\,(B- ATK)(B- ATK)' = \mathrm{tr}\,(BK^{-1} - AT)K^2\,(BK^{-1}- AT)' \qquad (3.6)$$

If K is known this problem corresponds to the problem of minimizing (3.2). The algorithm just described can be successfully applied to find T. For a given T, the diagonal values in K are given as (Koschat and Swayne, 1991)

$$K_{ii}= \frac{(B')_i\,(AT)_i}{(AT)'_i\,(AT)_i}$$

where $(B')_i$ and $(AT)_i$ denote the i-th column vectors of the matrices B and AT. If K and T are unknown, an iterative algorithm can be used. This permits determination of a sequence (K_i, T_i) whose limit is the solution (K,T). Koschat and Swayne (1991) recommend to start by choosing K_1 to be the identity matrix.

4. Procrustes analysis and geodetic applications

Geodetic data analysis often requires the application of rescaling, rotation and translation procedures of different data matrix configurations.

It seems therefore very strange that up to now procrustes analysis has not been widely applied in the geodetic literature. With this technique linearization problems of non linear equations systems and iterative procedures of computation could be avoided, in general, with significant computational time saving and less analytical difficulties.

To the author's knowledge only a single geodetic application of the procrustes technique was done in the 1980s for the construction of an ideal variance-covariance matrix (criterion matrix) of a control net point coordinates.

In that case the solution of the "classical procrustes problem" by Schönemann (1966) was applied to compute an unknown rotation matrix T able to guarantee a least squares matching of the matrices AT and B, where A is a variance covariance eigenvector matrix of a control net point coordinate vector and B is an ideal pseudo eigenvector matrix for the same vector. This last matrix was created a priori by 2D rotations of the "essential" eigenvector component pairs of the net point coordinates in such a way to orient them to the greatest possible extent along a direction orthogonal to that of the movement predicted by the deformation model. See for instance Crosilla (1983a, 1985) for the basic methodology and Crosilla (1983b) for further numerical developments of this technique.

The procedure known in the literature as a generalization of the orthogonal procrustes problem, given by Schönemann and Carroll (1970), could be applied with success for the transformation problems solutions of point coordinates between different reference systems.
In geodesy it is a common practice to transform by similarity 3D coordinates related to WGS 84 reference ellipsoid into 3D coordinates of a local reference system. For this purpose it is necessary to know in advance the approximate values of the unknown transformation parameters. Sometimes it is not easy to fix some of these values, like, for instance, in close range photogrammetry where often rotation angles between the model and the absolute reference systems are difficult to identify in advance.

In these cases procrustes methods are powerful because they do not require the knowledge of a priori unknown parameters values; from the computational point of view they just require some products of matrices containing point coordinates in different reference frames and the eigenvalue-eigenvector decomposition of a 3x3 matrix.

Some first numerical results of coordinate transformations with procrustes seem really satisfactory when compared with the results obtained with the classical methods and are worth of more deep investigations. A paper in progress will report these results and some further considerations.

Promising results are also expected by using generalized procrustes analysis, by Gower (1975) and Ten Berge (1977), for the computation of the International Terrestrial Reference System. As is well known the ITRS is based on the idea that each individual set of coordinates obtained by space geodesy measurements is related to a particular reference system. According to the procrustes approach, to combine all these coordinates into a unique frame, it is necessary to transform each solution by a 7 parameter similarity to an unknown common system satisfying the (2.1) Gower (1975) least squares function, previously reported.

In the author's opinion the same model could be expanded to compute the International Terrestrial Reference Frame (ITRF96) recently introduced by Sillard, Altamini and Boucher (1998) where a 14-parameter similarity is proposed to transform station positions and velocities into a combined system of reference.

Finally, a recent book by Dryden and Mardia (1998), presents very interesting applications of procrustes techniques and related statistics for the definition of objects' size-and-shape and the study of their variations. These examples are very interesting and worth applying to the deformation analysis of geodetic networks carried out by the comparison of two or more network adjustment results, relating to measurements made at different times.

Conclusions

Procrustes analysis seems to be a very promising technique in geodetic applications where transformation problems between systems of reference often have to be solved.
With respect to the classical transformation methods of solution, procrustes procedures take advantage of the symmetrical property of two matrices obtained by simple products of the original ones containing the point coordinate values to be matched. Spectral decomposition of these matrix products makes it possible to then compute the transformation parameters without any approximate value of the unknown parameters and with less computational time.
Very stimulating applications might be possible for the International Terrestrial Reference System and Frame computations and for the analysis of a deformation network by repeated measurements made at different times.

References

Crosilla F. 1983a, “A criterion matrix for the second order design of control networks”, Bull. geod. 57: 226-239

Crosilla F. 1983b “Procrustean transformation as a tool for the construction of a criterion matrix for control networks” Manusc. geod. 8: 343-370

Cosilla F. 1985 “ A criterion matrix for deforming networks by multifactorial analysis techniques”, in E. Grafarend F. Sansò eds. Optimization and design of geodetic networks, Springer Verlag, Berlin Heidelberg New York Tokyo

Dryden I.L. Mardia K.V., 1998 “Statistical Shape Analysis”,J. Wiley, Chichester

Goodall C. 1991 “ Procrustes methods in the statistical analysis of shape”, Journal Royal Statist. Soc. B 53, N° 2, pp 285-339

Gower J.C. 1975 “ Generalized procrustes analysis”, Psychometrika Vol. 40, N°1, pp 33-51

Koschat M. A. Swayne D. F. 1991” A weighted procrustes criterion” , Psychometrika vol 56 N° 2 pp 229- 239

Lingoes J.C. Schönemann P.H. 1974 “ Alternative measures of fit for the Schönemann- Carroll matrix fitting algorithm”, Psychometrika Vol 39, N°4 pp 423-427

Lissitz R.W. Schönemann P.H. Lingoes J.C. 1976 “A solution to the weighted procrustes problem in which the transformation is in agreement with the loss function”, Psychometrika vol 41 N° 4, pp 547-550

Schönemann P.H. Bock R.D. Tucker L.R. 1965 “Some notes on a theorem by Eckart and Young” Univ. of North Carolina Psychometric Laboratory Research Memorandum N°25

Schönemann P.H. 1966 “ A generalized solution of the orthogonal procrustes problem”, Psychometrika, vol 31, N°1, pp 1-10

Schönemann P.H. Carroll R.M. 1970 “Fitting one matrix to another under choice of a central dilation and a rigid motion”, Psychometrika, vol 35, N°2 pp245-255

Sillard P., Altamini Z., Boucher C. 1998 “ The ITRF 96 realization and its associated Velocity field”, Geophysical research letters, vol 25 N° 17, pp 3223-3226

Ten Berge J.M.F. 1977” Ortogonal procrustes rotation for two or more matrices”,Psychometrika vol 42, N° 2, pp 267-276