10

Rotation

Robert I. Jennrich

Introduction

Rotation arises in exploratory factor analysis and represents the primary distinction between exploratory and confirmatory factor analysis. This chapter attempts to provide an overview of the main components of rotation, from the work of Thurstone to the present. These components are rotation methods, rotation algorithms, and standard errors for rotated factor loadings.

Because it was simpler, early rotation methods used orthogonal rotation, and in particular varimax rotation. Today, because it is more general, oblique rotation is becoming the standard. At first quartimin was the most popular form of oblique rotation, but now qeomin seems to be replacing quartimin because geomin seems to work better for more complex structures.

At first, every rotation method came with a rotation algorithm designed specifically for the method, or more precisely for the criterion that defined the method. Today we have algorithms designed to optimize arbitrary criteria. A new method is proposed by defining its criterion, and a general algorithm is used to optimize it. This makes it much easier to investigate new methods.

A number of rotation algorithms including general closed-form pairwise algorithms for quartic criteria and general pairwise and gradient projection algorithms for arbitrary criteria, are discussed below.

Standard errors for rotated loadings using linear approximation methods are also reviewed. These include methods using the asymptotic distribution of the initial loadings, constrained maximum likelihood methods, nonparametric constrained minimum deviance methods, and nonparametric pseudo-value methods.

The chapter concludes with a comparison of some popular rotation methods, the difficulties that arise when doing this, some advice on choosing a method, and some real data examples.

The Wiley Handbook of Psychometric Testing: A Multidisciplinary Reference on Survey, Scale and Test Development, First Edition. Edited by Paul Irwing, Tom Booth, and David J. Hughes.

Exploratory Factor Analysis

The exploratory factor analysis (EFA) model has the form

$$x = \Lambda f + u, \tag{10.1}$$

where x is a vector of observed responses, f is a vector of common factors, u is a vector of unique factors, and Λ is a $p \times k$ matrix of factor loadings. It is assumed that the vectors f and u have mean zero and are uncorrelated, and that the components of u are uncorrelated.

Let Φ and Ψ be the covariance matrices of f and u. Then the covariance matrix of x is

$$\Sigma = \Lambda\Phi\Lambda' + \Psi.$$

Extraction

An EFA begins with an extraction step. Assume $\Phi = I$. Under this assumption an EFA model is called an orthogonal model because the factors are uncorrelated. With this assumption, $\Sigma = \Lambda\Lambda' + \Psi$. Let S be a sample covariance matrix and $D(S, \Sigma)$ be a deviance function that measures how close Σ is to S. Two popular deviance functions are the least squares deviance function,

$$D(S,\Sigma) = \|S-\Sigma\|^2,$$

and the maximum likelihood deviance function,

$$D(S,\Sigma) = \log(|\Sigma|) + \operatorname{tr}(S\Sigma^{-1}) - \log(|S|) - p.$$

The extraction step minimizes

$$D(S,\Lambda\Lambda' + \Psi)$$

with respect to Λ and Ψ. Let A and $\hat{\Psi}$ be the minimizing values of Λ and Ψ produced by the algorithm used for minimization. The matrix A is called an initial loading matrix. Note that A is a function of the data, but by convention it does not carry a hat. Finally, let

$$\hat{\Sigma} = AA' + \hat{\Psi}$$

be the estimate of Σ defined by A and $\hat{\Psi}$.

Oblique rotation

Our interest is primarily in oblique rotation. Let

$$\hat{\Lambda} = AT^{-1},$$

where A is an initial loading matrix and T is an arbitrary $k \times k$ nonsingular matrix with rows of length one. For reasons that will become clear shortly this is called an oblique rotation of the initial loading matrix A.

Let $\hat{\Phi} = TT'$, and note that

$$\hat{\Lambda}\hat{\Phi}\hat{\Lambda}' + \hat{\Psi} = AT^{-1}TT'(T')^{-1}A' + \hat{\Psi} = AA' + \hat{\Psi} = \hat{\Sigma}.$$

Thus, $\hat{\Sigma}$ is determined only up to an oblique rotation of A. This is called the rotation problem.

Among all oblique rotations of A, we seek one that looks nice—which more often than not means looks simple.

Orthogonal rotation

Let

$$\hat{\Lambda} = AT',$$

where A is an initial loading matrix and T is an arbitrary $k \times k$ orthogonal matrix. Since T' is an orthogonal matrix, the rows of $\hat{\Lambda}$ are rotations of the rows of A. This motivates calling $\hat{\Lambda}$ an orthogonal rotation of A.

Note that an orthogonal rotation is a restricted form of an oblique rotation, because an oblique rotation

$$\hat{\Lambda} = AT^{-1}$$

becomes an orthogonal rotation when T is an orthogonal matrix. An oblique rotation does not rotate the rows of A. Nevertheless, it is called a rotation because it is a generalization of an orthogonal rotation.

Because oblique rotation is more general than orthogonal rotation current practice favors oblique rotation, and our discussion will be oriented toward the oblique case.

Dealing with the Rotation Problem

There are many orthogonal and oblique rotations of the initial loading matrix A produced by the extraction step. The real rotation problem is choosing a rotation that may be of interest. This is usually done by using a rotation criterion $Q(\Lambda)$ that measures the complexity of Λ and minimizing this over all Λ that are rotations of A.

The main problem is, what does the vague statement "of interest" mean? One case is clear. If each row of Λ has at most one nonzero element, Λ is said to have perfect simple structure, an example of which is displayed in Table 10.1. However, there may not be a rotation of A with perfect simple structure, and this is the usual case. Thurstone (1935) proposed a less demanding definition of simple structure. The second loading matrix in Table 10.1 has Thurstone simple structure, which requires a fair number of zeros, but far fewer than perfect simple structure. The complexity of a row of Λ is the number of

Table 10.1 Examples of perfect and Thurstone simple structure.

Perfect			*Thurstone*		
1	0	0	1	0	0
1	0	0	0	1	0
1	0	0	0	0	1
0	1	0	.89	.45	0
0	1	0	.89	0	.45
0	0	1	0	.71	.71

nonzero elements in the row. Thurstone simple structure can have row complexities greater than one. As with perfect simple structure there may be no rotation of A that has Thurstone simple structure, and this is the usual case. It may, however, be possible to find a rotation of A that approximates Thurstone simple structure, or even perfect simple structure. The phrase "simple structure" is sometimes used to denote "perfect simple structure," and sometimes used to denote "approximate perfect simple structure". Here, "simple structure" will mean "approximate perfect simple structure."

Graphical Methods

The original rotation methods were graphical. The first term $c = \Lambda f$ in the factor analysis model of Equation 10.1 is called the common part of x. The ith component of c has the form

$$c_i = \lambda_{i1} f_1 + \cdots + \lambda_{ik} f_k.$$

This is plotted in Figure 10.1 for the case of two factors. Plotting all c_i gives the representation shown in Figure 10.2. The two-factor solution is obtained by choosing new factors $\tilde{f}_1$ and $\tilde{f}_2$ through the clusters of c_i and updating the λ_{ir}.

For more than two factors one cycles through pairs of factors making similar plots. Because it is better to actually do it than simply talk about doing it, the author attempted to graphically rotate the well-known Thurstone 26-variable box data (Thurstone, 1947, p. 371). Table 10.2 shows Thurstone's results and those of the author.

To aid in comparing these solutions their sorted absolute loading (SAL) plots (Jennrich, 2004) are given in Figure 10.3. The author did not do as well as Thurstone. A proper solution is known to have 27 small values. Thurstone got 27 small values. The author also got 27 small values, but clearly not as small. It is also known that a proper solution has three pure indicators, and these should produce three distinctly larger values. Thurstone found three distinctly larger values. The author failed to find these. The only conclusion one can draw from Figure 10.3 is that Thurstone is much better at graphical rotation than Jennrich.

While this may be the case, with the reader's indulgence the author would like a rematch. Thurstone's 26-variable box problem is known as a hard problem. Many rotation methods fail on this problem. Thurstone's 20-variable box problem, on the other hand, is much simpler. It was used to train students. Figure 10.4 shows the author's and

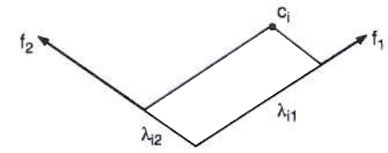

Figure 10.1 Components of the common part of x for two factors.

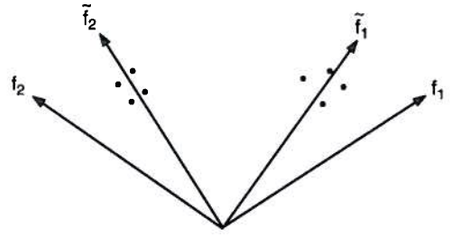

Figure 10.2 All the common part components.

Table 10.2 Graphical solutions to Thurstone's 26-variable box problem. The formulas on the left were used to generate values for the 26 variables from the dimensions x_1, x_2, and x_3 of Thurstone's boxes.

Formula	*Thurstone*			*Jennrich*		
x_1	.95	.01	.01	.98	.02	.02
x_2	.02	.92	.01	.04	.97	−.02
x_3	.02	.05	.91	−.02	−.07	1.02
x_1x_2	.59	.64	−.03	.62	.69	−.05
x_1x_3	.60	.00	.62	.58	−.07	.70
x_2x_3	−.04	.60	.58	−.05	.55	.63
$x_1^2x_2$	.81	.38	.01	.81	.43	−.00
$x_1x_2^2$	.35	.79	.01	.36	.85	−.02
$x_1^2x_3$	.79	−.01	.41	.77	−.05	.46
$x_1x_3^2$	.40	−.02	.79	.42	−.07	.92
$x_2^2x_3$	−.04	.74	.40	−.04	.73	.42
$x_2x_3^2$	−.02	.41	.74	−.05	.35	.80
x_1/x_2	.74	−.77	−.06	.75	−.83	.09
x_2/x_1	−.74	.66	.06	−.75	.83	−.09
x_1/x_3	.74	.02	−.73	.82	.15	−.85
x_3/x_1	−.74	−.02	.73	−.82	−.15	.85
x_2/x_3	−.07	.80	−.76	−.01	.99	−.91
x_3/x_2	.07	−.80	.76	.01	−.99	.91
$2x_1 + 2x_2$	.51	.70	−.03	.53	.76	−.06
$2x_1 + 2x_3$	.56	−.04	.69	.54	−.10	.74
$2x_2 + 2x_3$	−.02	.60	.58	−.03	.55	.62
$(x_1^2 + x_2^2)^{1/2}$	.50	.69	−.03	.52	.74	−.05
$(x_1^2 + x_3^2)^{1/2}$	.52	−.01	.68	.51	−.08	.74
$(x_2^2 + x_3^2)^{1/2}$	−.01	.60	.55	−.01	.56	.59
$x_1x_2x_3$	.43	.46	.45	.43	.45	.47
$(x_1^2 + x_2^2 + x_3^2)^{1/2}$	.31	.51	.46	.32	.49	.48

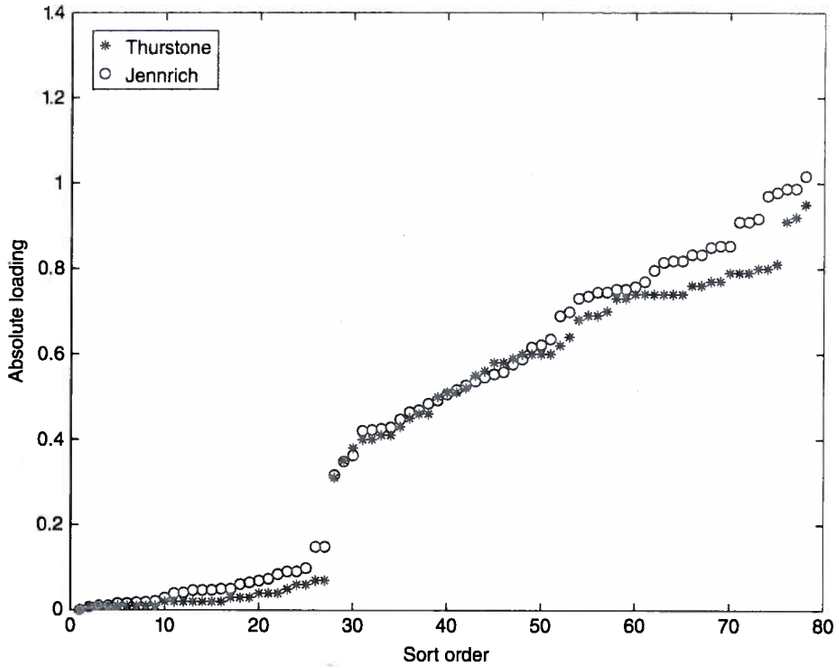

Figure 10.3 Sorted absolute loading plots for the 26-variable box problem.

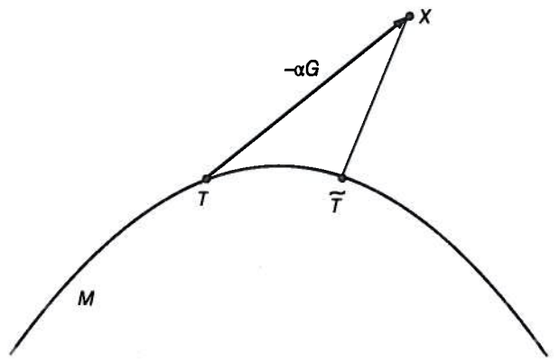

Figure 10.4 Graphical representation of the gradient projection algorithm.

Thurstone's results on this problem. Again, a proper solution should have 27 small values. The author found these almost as well as Thurstone. A proper solution to this problem is known to have nine pure indicators. The author found nine large values, as did Thurstone, but the author's are a little more distinct. The author would like to declare this contest a tie.

Analytic Methods

Analytic methods begin by choosing a rotation criterion Q defined on all $p \times k$ loading matrices Λ. $Q(\Lambda)$ is a measure of the complexity of Λ. An analytic method proceeds by minimizing

$$f(T) = Q(AT^{-1})$$

over all nonsingular T with rows of length one. Using the minimizing value $\hat{T}$,

$$\hat{\Lambda} = A\hat{T}^{-1} \tag{10.2}$$

is the oblique rotation of A corresponding to Q. Some authors replace T by T' and write $f(T) = Q\left(A(T')^{-1}\right)$, but this seems unnecessarily complicated. Many rotation criteria have been proposed. Some of the more popular are identified in the following subsections.

Quartic criteria

By quartic criteria we mean criteria Q such that $Q(\Lambda)$ is a quartic function of the components of Λ. The earliest of these was the quartimin criterion,

$$Q(\Lambda) = \sum_{r \neq s} \sum_{i} \lambda_{ir}^2 \lambda_{is}^2.$$

The quartimin criterion has a nice theoretical property. Quartimin rotation will produce perfect simple structure whenever it exists. If A can be rotated to a loading matrix with close to perfect simple structure, continuity suggests that quartimin rotation will do this as well.

A generalization of the quartimin criterion is the Crawford–Ferguson (Crawford & Ferguson, 1970) family (CF) of criteria. Let $0 \leq \kappa \leq 1$. The CF criterion corresponding to κ is given by

$$Q(\Lambda) = (1-\kappa) \sum_{r \neq s} \sum_{i} \lambda_{ir}^2 \lambda_{is}^2 + \kappa \sum_{i \neq j} \sum_{r} \lambda_{ir}^2 \lambda_{jr}^2.$$

The first term of this criterion is a multiple of the quartimin criterion and can be viewed as a measure of row complexity. The second term is a similar measure of column complexity. The parameter κ weights the two complexities. When $\kappa = 0$ the CF criterion is the quartimin criterion, and when $\kappa = 1/p$ it is the CF-varimax criterion.

The geomin criterion

The best known of the non-quadratic criteria is Yates' (1987) geomin criterion,

$$Q(\Lambda) = \sum_{i} \left(\prod_{r} \lambda_{ir}^2 \right)^{1/k},$$

where k is the number of columns of Λ. Yates called this a geomin criterion because it is the sum of the geometric mean of each row of Λ^2, the element-wise square of Λ.

The geomin criterion Q has a nice theoretical property. If a loading matrix Λ has at least one zero in each row then $Q(\Lambda)$ is zero, which is the smallest value it can have. If there is a rotation Λ of an initial loading matrix A with at least one zero in each row, a geomin rotation of A will have this property.

If there is a rotation of A with perfect simple structure, some geomin rotation of A will have perfect simple structure. Unfortunately this is not true for all geomin rotations of A, but this seems to happen often in practice.

A problem with the geomin criterion is that it is not differentiable at Λ if Λ has one or more components that are zero. This can make it difficult to minimize by analytic methods. Browne (2001) suggested dealing with this problem by replacing λ_{ir} in Yates' formula by $\lambda_{ir} + \epsilon$, where ϵ is a small positive value. This makes the resulting Q differentiable at all Λ. Browne suggested using $\epsilon = .01$. When used here, the geomin criterion will have this modified form.

Bifactor criteria

Bifactor analysis is a form of confirmatory factor analysis using a factor loading matrix of the form

$$\Lambda = \begin{pmatrix} * & * & 0 \\ * & * & 0 \\ * & * & 0 \\ * & 0 & * \\ * & 0 & * \\ * & 0 & * \end{pmatrix}.$$

More precisely, the loadings in the first column are free parameters and after the first column the loading matrix has at most one free parameter in each row. Loading matrices of this form are said to have bifactor structure. In bifactor analysis the first factor is called a general factor and the remaining factors are called group factors. Bifactor analysis is a fairly extensively used form of confirmatory factor analysis.

Recently, Jennrich and Bentler (2011) introduced an exploratory form of bifactor analysis. This was done by using orthogonal EFA with a rotation criterion that does not involve the loadings on the first factor and encourages perfect simple structure for the loadings on the remaining factors. To avoid confusion, they call this confirmatory bifactor analysis to differentiate it from exploratory bifactor analysis.

$B(\Lambda)$ is called a bifactor rotation criterion if Λ minimizes $B(\Lambda)$ if and only if Λ has bifactor structure. The following theorem identifies a specific bifactor rotation criterion.

Theorem 10.1: If Λ is an arbitrary $p \times k$ loading matrix and

$$B_q(\Lambda) = \mathrm{qmin}(\Lambda_2),$$

where Λ_2 is the submatrix of Λ containing its last $k - 1$ columns and qmin() is the quartimin rotation criterion, then B_q is a bifactor rotation criterion.

The criterion B_q in Theorem 10.1 is called the biquartimin criterion. Although $B_q(\Lambda)$ does not depend on the first column of Λ, when $B_q(\Lambda)$ is used for rotation, it is all columns of Λ, including its first, that are rotated.

Some care must be used in defining bifactor rotation criteria. One might be tempted to use

$$B_v(\Lambda) = -\mathrm{vmax}(\Lambda_2),$$

where vmax() is the varimax rotation criterion. The authors show that B_v is not a bifactor rotation criterion and that its use can lead to very poor results.

Other criteria

There are many other rotation criteria, including Bentler's (1977) invariant pattern simplicity criterion, McCammon's (1966) minimum entropy criterion, and McKeon's (1968) infomax criterion. There is a cottage industry of numerous papers proposing new criteria.

Reference structure and early analytic methods

Analytic oblique rotation was not originally formulated as described by Equation 10.2 because the required optimization, which involves T^{-1}, seemed too difficult. Using T^{-1} can be avoided by using reference structures. Following Thurstone (1947) and Harman (1976), let the rows of a nonsingular matrix U be biorthogonal to the rows of T and have length one. Biorthogonal means the rth row of U is orthogonal to the sth row of T whenever $r \neq s$. Let

$$R = AU'.$$

This is called the reference structure corresponding to U. Note that

$$R = AT^{-1}TU' = \hat{\Lambda}D,$$

where D is diagonal because the rows of T and U are biorthogonal. This result is of interest because it means the columns of R are rescaled versions of the columns of $\hat{\Lambda}$, and this suggests that R is simple when $\hat{\Lambda}$ is simple, and conversely. Rather than apply a complexity criterion to Λ one can apply it to R. The analytic rotation problem is then to minimize

$$Q(R) = Q(AU')$$

over all nonsingular U with rows of length one. Here, U' has replaced the T^{-1} in the definition of an oblique rotation. Harman (1976) calls making R simple an indirect method and making Λ simple a direct method because it simplifies the loadings directly.

Carroll (1953) has shown that when Q is the quartimin criterion, $Q(AU')$ viewed as a function of a single row of U is a constant plus a homogeneous quadratic function of the row. Because the row must have length one, an optimal value is the eigenvector of the

matrix defining the quadratic function that corresponds to its smallest eigenvalue. Cycling through the rows of U gives a relatively simple algorithm for minimizing $Q(R)$. Actually, this can be generalized to Carroll's (1960) oblimin family of criteria. It does not, however, seem to generalize to other criteria. For some time, indirect oblimin was the standard method of oblique analytic rotation.

Direct methods

Indirect methods based on criteria applied to reference structures were eventually replaced by methods based on criteria applied to loading matrices directly. The first such method was introduced by Jennrich and Sampson (1966) for direct quartimin rotation. Today, direct methods are standard. Algorithms for these will be discussed in the next section.

Analytic Oblique Rotation Algorithms

In the beginning, all proposed rotation criteria came with a rotation algorithm designed to optimize the specific proposed criterion. These algorithms were sometimes quite complex. Later, algorithms were developed that worked for entire classes of criteria, and later still algorithms that worked for essentially all criteria. Some of these will be discussed here.

Closed-form pairwise algorithms for quartic criteria

Let f_1 and f_2 be a pair of factors and consider rotating f_1 in the plane of f_1 and f_2. More precisely, the new f_1 factor has the form

$$\tilde{f}_1 = \alpha_1 f_1 + \alpha_2 f_2$$

and has variance one. Jennrich and Sampson (1966) showed that the values of the quartimin criterion under such rotations can be expressed as a fourth-degree polynomial $Q'(\delta)$ in $\delta = \alpha_1/\alpha_2$. This may be minimized in closed form without iteration by solving the cubic equation $Q'(\delta) = 0$. Cycling through all ordered pairs of factors gives a pairwise algorithm for minimizing the criterion. This method generalized to the oblimin family of criteria is used in a number of major software systems.

Although to date no formal proof has been given, this approach also works when the quartimin criterion is replaced by any quartic criterion $Q(\Lambda)$ that is invariant under sign changes in the columns of Λ. These include essentially all quartic criteria, and in particular the oblimin and CF families of criteria. Because five values determine a quartic polynomial, $Q(\delta)$ can be found by evaluating Q at five loading matrices corresponding to five values of δ. As a consequence, the only specific information required to implement these methods is a formula for $Q(\Lambda)$.

General pairwise line search algorithms

Browne and Cudeck have developed a pairwise line search algorithm for minimizing arbitrary rotation criteria. Let $Q(\Lambda)$ be any rotation criterion that is invariant under sign changes in the columns of Λ. These include all criteria known to the author. As was done

by Jennrich and Sampson, Browne and Cudeck let f_1 and f_2 be an arbitrary pair of factors and consider rotating f_1 in the plane of f_1 and f_2. For pairwise rotations of this form they have shown that the values of the criterion Q can be expressed as a function $Q(\delta)$ of the parameter δ used by Jennrich and Sampson. In general $Q(\delta)$ will not be quartic, but it can be minimized using a general line search algorithm. This forms a basis for a pairwise algorithm for minimizing Q. This is a remarkable algorithm:

- it works for almost any rotation criterion $Q(\Lambda)$;
- all that is required is a formula for $Q(\Lambda)$;
- it is remarkably simple;
- it has been used successfully for many different criteria.

Unfortunately, Browne and Cudeck have not published an account of their method or observations on its performance. Their method is used, however, by Browne, Cudeck, Tateneni, and Mels (2002) in the CEFA (comprehensive exploratory factor analysis) software. This free software deals with almost every aspect of EFA, including a broad variety of methods for extraction and rotation, factoring correlation matrices, and providing standard errors for the estimates produced. It has a graphical user interface and a nice manual. The software and manual may be downloaded.[1]

Gradient projection algorithms

Jennrich (2002) gave a general gradient method that does not require cycling through pairs of factors. The oblique rotation problem is to minimize

$$f(T) = Q\left(AT^{-1}\right)$$

over all T in the manifold M of nonsingular T with rows of length one. This is a slightly more precise expression of the problem than that given earlier. The gradient projection (GP) algorithm proceeds as follows.

Given a $T \in M$ and an arbitrary scalar $\alpha > 0$, compute the gradient G of f at T and project $X = T - \alpha G$ onto M—see Figure 10.4. The algorithm moves T in a negative gradient direction and then projects the result back onto M. At first it seems like the required projection may be difficult to find because projecting onto a nonlinear manifold is a nonlinear regression problem and these generally require complex iterative procedures. For the manifold M, however, projection is very easy and this is what motivates the method. The projection $\tilde{T}$ of X onto M is simply X scaled to have rows of length one.

Theorem 10.2: If T is not a stationary point of f restricted to M, then

$$f\left(\tilde{T}\right) < f(T)$$

for all $\alpha > 0$ and sufficiently small.

[1] http:/quantrm2.psy.ohio-state.edu/browne/

Using Theorem 10.2, the GP algorithm halves an initial value of α until $f(T) < f(T)$. Replacing T by $\widetilde{T}$ gives a strictly monotone iterative algorithm for minimizing f over M, and hence Q over all rotations of the initial loading matrix A. Strictly monotone algorithms are desirable because they must converge to stationary points. Moreover, since the only points of attraction of a strictly monotone algorithm are local minima, strictly monotone algorithms are almost guaranteed to converge to a local minimum.

This approach requires a formula for the gradient of Q in addition to a formula for Q. Jennrich (2004) has shown that this problem-specific requirement can be removed with almost no loss of precision by using numerical gradients. Using these, the gradient projection algorithm has the same nice properties as the Browne and Cudeck line search algorithm without the need to cycle through pairs of factors.

Free SAS, SPSS, R/S, and Matlab code for GP rotation can be downloaded.[2] Thus, for almost any computing environment one is working in, one can find code written specifically for that environment and hence code that may be used immediately without any need for translation. There is code for both orthogonal and oblique rotation using analytic and numerical gradients. The analytic gradients tend to be quite simple. They are given for the most popular, and some less popular but promising, criteria.

Advantages and disadvantages

The pairwise quartic algorithms have the advantage that no line search is required, and they are probably the fastest of the algorithms discussed. Their main disadvantage is that they are restricted to quartic criteria.

The main advantage of the general pairwise line search algorithm is that it applies to arbitrary criteria and the only method-specific code required is that for evaluating the criterion. Also, it is a very simple algorithm. Disadvantages are that it requires cycling through pairs and requires a line search subalgorithm.

The main advantage of the GP algorithm is that it applies to arbitrary rotation criteria, it does not require stepping through pairs of factors, and when using numerical gradients, it is very simple. When using analytic gradients, it appears to be significantly faster than the general pairwise line search algorithm, at least in the limited experience of the author. Its main disadvantage is that when used with analytic gradients it requires method-specific code to produce these. Fortunately, gradient formulas for popular rotation criteria are given at the URL cited above. While gradient formulas can be avoided with almost no loss of precision by using numerical gradients, their use slows the algorithm.

Choosing a Rotation Method

Up to this point we have not discussed the problem of choosing a rotation method. Unfortunately, there is no right or wrong choice. Exploratory factor analysis is an exploratory method. It is used to suggest a relation between observed variables and latent factors. A rotation method suggests what such a relation may look like, but there are many rotation methods and they can suggest different relations. It is the user who must ultimately decide which if any of these may be of interest. The data can identify an initial loading matrix, but it cannot determine what rotation, if any, is of interest in a specific context.

[2] http://www.stat.ucla.edu/research/gpa

There is some theoretical guidance:

- If an initial loading matrix can be rotated to a loading matrix with perfect simple structure, the quartimin method will do this. Continuity suggests the quartimin method will do this at least approximately when the initial loading matrix can be rotated to a loading matrix that approximates perfect simple structure.
- If an initial loading matrix A can be rotated to a loading matrix with at least one zero loading in each row, a geomin rotation of A will have this property. When A can be rotated to a loading matrix with at least one nearly zero loading in each row one expects a geomin rotation of A will have at least one nearly zero loading in each row.

There is a cottage industry devoted to proposing new rotation methods. Every time a new method is introduced its performance on one or two examples is given and compared with one or two natural alternatives. The purpose is to show that the new method works at least in appropriate circumstances. There are only a limited number of papers that compare a number of popular methods on a number of examples from an unbiased point of view. These include Browne (2001) and Schmitt and Sass (2011).

Browne looked at two classical examples, the 24 psychological tests problem (Harman, 1976) and Thurstone's (1947) 26-variable box problem. In the case of the 24 psychological tests problem Browne found that an initial loading matrix can be rotated to a loading matrix with simple structure using a number of methods, including quartimin, CF-varimax, geomin, and infomax.

To investigate a more complex case Browne considered Thurstone's 26-variable box problem. Here there is a rotation that suggests the way the data were generated, and Browne considered the ability of various methods to obtain this.

Among the methods considered, only geomin and infomax identified the way the data were generated. Browne goes on to show how one can use Cureton and Mulaik (1975) standardization to obtain satisfactory results for other methods as well. We will not attempt to discuss this here.

Using a somewhat different approach, Schmitt and Sass (2011) began with the population loading matrix Λ displayed in the first three columns of Table 10.3 above the horizontal line. This has very a complex structure because it is far from perfect simple structure. They used a population factor correlation matrix Φ given by the 3×3 matrix below the population loading matrix and for Ψ a diagonal matrix of unique variances:

.3188 .3425 .3315 .1845 .1518 .3215 .3315 .0235 .0713 .3695 .4303 .2845

.2123 .0235 .2535 .3915 .2563 .3695

Rather than using Λ, Φ, and Ψ to generate a sample and analyzing its correlation matrix R, Schmitt and Sass constructed the population correlation matrix P and analyzed this rather than R. This avoids random results and the problem of choosing the sample size n. What Schmitt and Sass did was to let $n = \infty$. Their analysis should give results that approximate large-n results. This seems like a good place to start when comparing rotation methods.

Table 10.3 Comparing rotations for a model with complex structure.

Population			*Quartimin*			*Geomin*			*CFvarimax*		
0.75	0.25	0.15	0.79	0.07	0.03	0.74	0.11	0.07	0.72	0.13	0.09
0.75	0.10	0.25	0.78	-0.10	0.17	0.73	-0.06	0.21	0.72	-0.02	0.21
0.75	0.30	0.05	0.82	0.13	-0.09	0.76	0.18	-0.06	0.74	0.19	-0.02
0.75	0.45	0.10	0.77	0.27	-0.05	0.71	0.32	-0.03	0.70	0.32	0.01
0.75	0.15	0.45	0.70	-0.07	0.36	0.65	-0.04	0.40	0.65	0.01	0.38
0.75	0.20	0.20	0.78	0.01	0.09	0.73	0.05	0.13	0.72	0.08	0.14
0.05	0.75	0.30	-0.04	0.72	0.23	-0.04	0.75	0.18	-0.02	0.72	0.21
0.30	0.75	0.45	0.19	0.59	0.33	0.17	0.62	0.31	0.20	0.62	0.13
0.45	0.75	0.25	0.39	0.59	0.11	0.36	0.63	0.09	0.37	0.62	0.13
0.10	0.75	0.20	0.04	0.73	0.11	0.03	0.76	0.06	0.05	0.73	0.11
0.05	0.75	0.05	0.01	0.80	-0.06	0.01	0.84	-0.11	0.02	0.78	-0.04
0.30	0.75	0.15	0.27	0.66	0.02	0.25	0.71	-0.01	0.25	0.68	0.05
0.05	0.45	0.75	-0.13	0.34	0.75	-0.12	0.34	0.73	-0.06	0.38	0.68
0.45	0.30	0.75	0.29	0.10	0.68	0.28	0.12	0.69	0.32	0.17	0.65
0.30	0.20	0.75	0.15	0.04	0.76	0.14	0.04	0.76	0.19	0.10	0.70
0.15	0.10	0.75	-0.01	-0.03	0.84	-0.00	-0.04	0.84	0.05	0.03	0.76
0.25	0.25	0.75	0.09	0.10	0.76	0.09	0.10	0.75	0.14	0.16	0.70
0.20	0.10	0.75	0.05	-0.04	0.82	0.05	-0.04	0.83	0.10	0.02	0.75
1.00	0.40	0.40	1.00	0.66	0.72	1.00	0.64	0.67	1.00	0.59	0.60
0.40	1.00	0.40	0.66	1.00	0.66	0.64	1.00	0.70	0.59	1.00	0.59
0.40	0.40	1.00	0.72	0.66	1.00	0.67	0.70	1.00	0.60	0.59	1.00

Schmitt and Sass considered a number of popular rotation methods, but the primary contenders were geomin, quartmin, and CF-varimax. They used a maximum likelihood factor analysis program applied to P to produce an initial loading matrix A and rotated this.

There is a way to obtain an initial loading matrix A corresponding to P that does not require the use of an EFA program. One can simply unrotate Λ and Φ by using a Cholesky factorization to write Φ in the form $\Phi = TT'$. Then,

$$\Sigma = \Lambda\Phi\Lambda' + \Psi = \Lambda TT'\Lambda' + \Psi.$$

Let $D = \text{diag}(\Sigma)^{-1/2}$. Then

$$P = D\Lambda\Phi\Lambda'D + D\Psi D = (D\Lambda T)(D\Lambda T)' + D\Phi D = AA' + \tilde{\Psi},$$

where

$$A = D\Lambda T.$$

This A is an initial loading matrix that generates P. This is much simpler and more accurate than running an iterative factor analysis extraction program to produce an initial loading matrix.

Using the method just described, the results for quartimin, geomin, and CF-varimax are given in Table 10.3. To the precision displayed, the quartimin and CF-varimax

Table 10.4 Example 1 rotations

Canonical					*Orthogonal*					*Oblique*				
54	42	-13	-0	-7	29	63	4	9	-5	66	0	11	-1	-7
44	27	-8	6	-1	26	45	8	0	-2	47	8	-0	4	-3
59	42	-15	5	8	33	66	4	-4	5	74	1	-6	3	6
48	46	-16	5	6	22	65	2	-3	2	73	-2	-6	-4	1
59	40	-13	-0	-0	33	64	4	6	1	68	1	7	2	-1
50	3	33	22	-5	38	18	48	6	3	3	62	4	-4	1
50	-5	26	18	-5	42	13	39	5	1	-2	52	5	9	0
53	2	28	27	-11	41	20	48	4	-6	4	65	1	4	-9
53	-3	34	14	11	44	14	43	3	20	1	53	1	1	21
54	-5	16	-21	25	48	16	7	12	38	9	0	19	10	44
62	-6	15	-24	-4	56	18	9	31	16	2	4	46	15	16
51	-7	24	-13	20	45	12	18	12	34	2	15	17	5	39
62	9	2	-13	-4	50	34	6	19	7	26	3	29	16	8
63	4	21	-28	-20	51	25	14	47	7	5	9	66	1	3
74	-35	-26	1	-7	84	9	-10	-4	-12	1	-5	8	86	-6
70	-26	-7	8	1	74	11	9	-5	-2	1	17	1	61	4
68	-34	-20	7	8	78	6	-5	-14	-1	1	2	-9	79	8

results in Table 10.3 are the same as those in Table 10.4 of Schmitt and Sass. The geomin results, however, differ significantly.

The value of the geomin criterion for the geomin result in Table 10.3 is 1.3231. Schmitt and Sass do not report the value of the geomin **criterion** for their result, but when it is computed from the rounded numbers in their Table 10.4 the value is 1.3491. Because this value is larger, this suggests that the Schmitt and Sass result may be incorrect.

Schmitt and Sass don't seem to have a favorite method. Instead, they show how the methods they consider compare to one another, and in particular how the CF methods change with increasing values of κ. They consider a variety of values of κ and observe that as κ increases the factor correlations decrease, but the complexity of the rotations increases.

As noted above, the quartimin method is the Crawford–Ferguson method with $\kappa = 0$ and the CF-varimax method corresponds to $\kappa = 1/p$. The factor correlations in Table 10.3 decrease as κ increases, in agreement with their observation. To consider the change in complexity this author constructed SAL plots for the rotations in Table 10.3. These are displayed in Figure 10.5. Schmitt and Sass call a loading λ_{ir} a primary loading if its absolute value is the largest among the loadings in the same row, and call it a cross loading otherwise. There are a total of 18 primary loadings in each loading matrix. These are displayed in the upper left-hand corner of the plot.

The cross loadings for CF-varimax tend to be a bit larger than those for quartimin and geomin. This is in agreement with Schmitt and Sass' observation that complexity increases with κ. Overall, however, the methods displayed are more similar than different. One might seek an example where differences are more dramatic. For the box data considered by Browne, there was a substantial difference between quartimin and geomin.

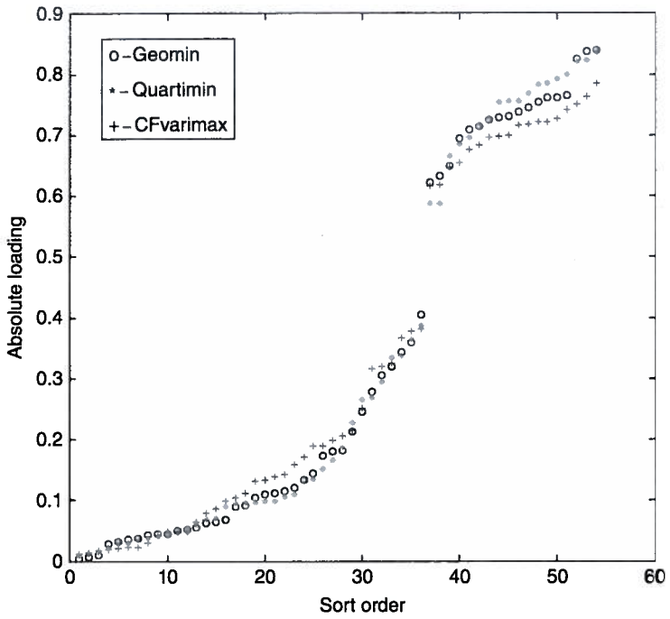

Figure 10.5 Comparison using SAL plots.

At present, the only simple thing one can say about comparing rotation methods in the complex case is that it is complex. Clearly, more work on this case is needed.

Where to start

The first problem faced by a factor analyst is where to start. For this, one can give fairly reasonable advice.

- A number of popular EFA programs use a normal deviance function to extract an initial loading matrix. This seems to work well and is probably a reasonable way to begin. The normal deviance function is a proper deviance function and a reasonable choice even when one is not sampling from a normal distribution.
- Because it is more general it seems reasonable to begin with oblique rotation and reserve orthogonal rotation for applications where orthogonal rotation is required or desired.
- Theory suggests that quartimin rotation works well when one is seeking simple structure and some limited computer experience suggests that geomin may work well when seeking rotations with simple and more complex structure.

A number of EFA programs use oblique quartimin or oblique geomin as default choices for a rotation method, and presently at least these seem appropriate first choices.

Standard Errors for Rotated Loadings

Measured in terms of actual computer usage, EFA is one of the most frequently used methods of statistical analysis. Among statistical methods it is somewhat unique in that standard computer implementations fail to produce standard errors for the parameter estimates. For many forms of statistical analysis standard errors are a by-product of the fitting procedure. This is true, for example, for regression analysis and the many forms of analysis related to it. It is also true for confirmatory factor analysis, but because of the rotation problem it is not true for EFA. Other methods for producing standard errors are required. Here we outline a number of methods that may be used to produce standard errors for rotated loadings. For the most part these are asymptotic methods based on linear approximations. As discussed below, little progress seems to have been made using less linear methods such as the jackknife, the bootstrap, and Markov chain Monte Carlo (MCMC) methods. Cudeck and O'Dell (1994) discuss a variety of uses for standard errors in EFA.

Historical note

An analytic rotation of an initial loading matrix A has the form

$$\hat{A} = AT^{-1}.$$

Because both A and T are random, Lawley and Maxwell (1971) believed that:

> It would be almost impossible to take sampling errors in the elements of T into account. The only course is, therefore, to ignore them in the hope they are relatively small.

Wexler (1968) provided some evidence that one cannot always ignore sampling errors in T. Archer and Jennrich (1973) and Jennrich (1973) showed that the Lawley and Maxwell approximation is not needed. This is discussed in the next subsection.

Methods using the asymptotic distribution of the initial loadings

Under appropriate assumptions an initial loading matrix A is asymptotically normally distributed. More precisely, if n is the sample size used to compute A, then as $n \to \infty$,

$$\sqrt{n}(a - a_0) \to N(0, \text{acov}(a)),$$

where a is A written in vector form, a_0 is a constant vector, and acov(a) is the asymptotic covariance matrix for a.

Results of this form have been given by

- Anderson and Rubin (1956) for principal component EFA and normal sampling;
- Lawley (1967) for canonical loadings and normal sampling;
- Jöreskog (1969) for confirmatory factor analysis and normal sampling;
- Browne (1984) for confirmatory factor analysis and non-normal sampling;
- Girshick (1939) for principal component analysis and normal sampling.

Actually, Lawley's formula for acov($\hat{a}$) has an error that was corrected by Jennrich and Thayer (1973).

The confirmatory factor analysis approach is particularly attractive. It assumes that the upper diagonal part of A is zero. A standard confirmatory factor analysis then produces an estimate for acov($\hat{a}$) as a by-product of the analysis.

Consider any analytic rotation. Let A be the initial loading matrix and let $\hat{\Lambda}$ be the corresponding rotated loading matrix, which is a function of A. In vector form let

$$\hat{\lambda} = h(a).$$

Assuming that a is asymptotically normally distributed, it follows from the delta method that $\hat{\lambda}$ is also, and that

$$\text{acov}\hat{\lambda} = \frac{dh}{da}\text{acov}(a)\frac{dh'}{da},$$

where dh/da is the Jacobian of h at a. Thus, the problem of finding the asymptotic distribution for rotated loadings given that this is known for the corresponding initial loadings reduces to finding the Jacobian of the function that defines the rotation. Jennrich (1973) has shown how to compute dh/da using implicit differentiation.

Constrained maximum likelihood methods

For the case when the observed responses are a normal sample Jennrich (1974) used a result of Silvey (1971) on constrained maximum likelihood estimation to find the asymptotic distribution of the rotated loadings $\hat{\Lambda}$.

Let λ be Λ written as a vector, ϕ be the upper-diagonal part of Φ written as a vector, ψ be the diagonal of Ψ written as a vector, $\theta = (\lambda', \phi', \psi')'$ be the complete parameter vector, and $\ell(\theta)$ be the likelihood of θ given the observed responses.

This likelihood is over-parameterized. Jennrich (1973) has shown, however, that for oblique rotation, Λ and Φ must satisfy the stationary condition

$$\text{ndg}\left(\Lambda'\frac{dQ}{d\Lambda}\Phi^{-1}\right) = 0.$$

Write this as

$$\varphi(\theta) = 0.$$

Let $\dot{\varphi}$ be the Jacobian of φ at θ and $\mathcal{I}$ be the information matrix at θ. Using Silvey's result on constrained maximum likelihood estimation,

$$\begin{pmatrix} \mathcal{I} & \varphi' \\ \varphi & 0 \end{pmatrix}^{-1} = \begin{pmatrix} \text{acov}(\hat{\theta}) & * \\ * & * \end{pmatrix}.$$

That is, when the augmented information matrix on the left is inverted, what is in the upper left-hand corner of the inverse is the asymptotic covariance matrix for the constrained maximum likelihood estimate $\hat{\theta}$. If $\mathcal{I}$ and $\dot{\varphi}$ are replaced by sample estimates, what is in the upper left-hand corner of the inverse is a consistent estimate of the asymptotic covariance matrix of $\hat{\theta}$. This can be used to assign standard errors to the factor loading estimates λ_{ir}. If $\dot{\varphi}$ is computed numerically, this is a very simple approach when using maximum likelihood factor analysis.

A method based on the asymptotic distribution of $s = \text{vech}(S)$

A difficulty with the previous approach is the assumption that one is sampling from a normal distribution. In practice this is at best only approximately true. To relax this assumption, consider the use of a method based on the asymptotic distribution of $s = \text{vech}(S)$, where vech(S) is a column vector containing the diagonal and upper diagonal components of the sample co- variance matrix S.

More specifically, let θ be the parameter vector defined in the previous subsection. An EFA estimate $\hat{\theta}$ of these parameters is a function of s. That is,

$$\hat{\theta} = h(s).$$

Let Γ be the asymptotic covariance matrix for s. By the delta method, the asymptotic covariance matrix for $\hat{\theta}$ is given by

$$\widehat{\text{acov}}(\hat{\theta}) = \dot{h}(\sigma)\Gamma\dot{h}(\sigma)',$$

where $\sigma = \text{vech}(\Sigma)$ and $\dot{h}(\sigma)$ is the Jacobian of h at σ. This can be consistently estimated using

$$\widehat{\text{cov}}(\hat{\theta}) = \dot{h}(s)\hat{\Gamma}\dot{h}(s)',$$

where $\hat{\Gamma}$ is a consistent estimator of Γ. The diagonal elements of $\widehat{\text{acov}}(\hat{\theta})$ can be used to assign standard errors to the components of $\hat{\theta}$, including the factor loadings $\hat{\lambda}_{ij}$.

To use this apparently very simple method one needs a formula for $\dot{h}$(s) and $\hat{\Gamma}$. In EFA there is no formula for $h(s)$. The estimate $\hat{\theta}$ is implicitly defined as a vector that minimizes

$$f(\theta, s) = D(S, \Lambda\Phi\Lambda' + \Psi)$$

and satisfies $\varphi(\theta) = 0$, where

$$\varphi(\theta) = \text{ndg}\left(\Lambda'\frac{dQ}{d\Lambda}\Phi^{-1}\right)$$

and Q is the criterion used to define the rotation. Let

$$g(\theta, s) = \begin{pmatrix} \dot{f}_1(\theta, s) \\ \varphi(\theta) \end{pmatrix},$$

where $\dot{f}(\theta,s)$ is the Jacobian of $f(\theta,s)$ viewed as a function of its first argument. It follows that

$$g(\hat{\theta},s)=0.$$

One can use this and implicit differentiation to find a formula for $\dot{h}(s)$. Since $\hat{\theta}=h(s)$,

$$g(h(s),s)=0.$$

Differentiating this with respect to s gives

$$\dot{g}_1(h(s),s)\dot{h}(s)+\dot{g}_2(h(s),s)=0.$$

Solving for $\dot{h}(s)$ gives

$$\dot{h}(s)=-\dot{g}_1(h(s),s)^{-1}\dot{g}_2(h(s),s).$$

When using a least squares or normal deviance function the derivatives $\dot{g}_1(\theta,s)$ and $\dot{g}_2(\theta,s)$ can be computed easily using numerical differentiation and, with some effort, analytically as well.

A convenient estimator for Γ is the sample covariance matrix of the n vectors

$$d_i=\text{vech}\big((x_i-\bar{x})(x_i-\bar{x})'\big);$$

see, for example, Satorra and Bentler (1990, Formula 2.4).

One trouble with the method of this subsection is that it has not been published. The author suspects, however, it is the method used in CEFA.

Pseudo-value methods

In practice researchers analyze sample covariance matrices S that probably were not generated by a factor analysis model. In fact, this is probably the usual case. These analyses violate the assumptions made by the standard error methods discussed above. One might wonder if it is nevertheless possible to provide standard errors for the estimates $\hat{\Lambda},\hat{\Phi}$, and $\hat{\Psi}$ produced by these analyses. This can in fact be done using the infinitesimal jackknife (IJK).

Jennrich (2008) showed how to use the IJK to produce standard errors for a covariance structure analysis of nonnormal data. This proceeds as follows.

Let S be a sample covariance matrix for a sample $x_1,\ldots,x_n$ and let

$$\hat{\theta}=h(S),$$

where h is an arbitrary function. For each x_i the IJK produces a pseudo-value

$$\theta_i^+=\mathrm{d}h_S((x_i-\bar{x})(x_i-\bar{x})),$$

where $\mathrm{d}h_S$ is the differential of h at S. The sample covariance matrix of these values is a consistent estimate of the asymptotic covariance matrix for $\hat{\theta}$.

The IJK method has several advantages.

- It is a non-parametric method. One can sample from any distribution.
- The covariance structure $\Sigma(\theta)$ need not be correctly specified. That is, there may be no θ such that $\Sigma(\theta)$ is equal to the covariance matrix for the population sampled. This is important because in practice it almost never is.
- The third and fourth sample moments of the sample values x_i are not required.
- The only real difficulty when using this method is finding the derivatives of D and Σ, which is simple if numerical derivatives are used, but can be a bit messy otherwise.

In the context of EFA, Zhang, Preacher, and Jennrich (2012) have shown how to use the IJK methods of Jennrich (2008) to produce standard errors for an EFA of a sample correlation matrix obtained from a nonnormal sample. This is important because in practice one often analyzes sample correlation matrices rather than sample covariance matrices.

The authors give explicit formulas for ordinary least squares and maximum likelihood extraction and for arbitrary CF rotation. They indicate the modifications required for the analysis of a sample covariance matrix rather than a sample correlation matrix.

Because these IJK methods are nonparametric and do not require data generated by a correctly specified factor analysis model they may provide an attractive option for general purpose factor analysis software such as SAS, SPSS, Stata, CEFA, and Mplus.

Less linear methods

By less linear methods we mean things like the jackknife, the bootstrap, and for Bayesian estimation MCMC methods. Pennell (1972) and Clarkson (1979) have used the jackknife to produce standard errors for rotated loadings. The basic jackknife uses n jackknife values each of which requires carrying out a factor analysis extraction and rotation. Doing this n times makes this an expensive procedure. Jackknifing by groups helps to reduce this expense, but the main problem with using the jackknife concerns alignment. In EFA the rotated loading matrix Λ is determined only up to column permutation and column sign change. To make the jackknife work the generated loading matrices must be aligned. Doing this in an automated way that is good enough for jackknifing is pretty much an unsolved problem. Alignment failures can have devastating effects on jackknife standard error estimates.

There is also considerable expense when using bootstrap methods, and again alignment is the main problem. MCMC methods face these problems as well, and require a parametric form for the sampling distribution of x. At present the only feasible estimates for standard errors of rotated loadings seem to be linearization methods like those in the previous subsections.

Some Examples Using Real Data

Chen, West, and Sousa (2006) used a 17-variable quality of life data set to compare two methods of analysis. The covariance matrix for their investigation is given in their paper and will be used here.

Example 1: Comparing orthogonal and oblique rotation

Using the covariance matrix of Chen et al., maximum likelihood EFA was used to extract initial loadings A and unique variances $\hat{\Psi}$. The initial loadings are given on the left-hand side of Table 10.4. To simplify comparisons the loadings have been rounded to two decimal places and the decimal points removed. These loadings are canonical loadings, which means $A'\hat{\Psi}^{-1}A$ is diagonal. They are the loadings initially produced by standard maximum likelihood EFA programs. It is these loadings that are rotated to produce various forms of rotated loadings.

Many EFA programs do not display the initial loadings they produce. We obtained these by requesting "no rotation." To compare orthogonal and oblique rotation we have used the quartimin criterion because it can be used for either orthogonal or oblique rotation. The orthogonal and oblique quartimin rotations of A are also given in Table 10.4. Clearly the oblique rotation is much closer to simple structure than the orthogonal rotation. This is also displayed by the SAL plots given in Figure 10.6. Clearly the oblique loadings are much closer to simple structure than the orthogonal loadings. One expects this in general, because there are many more oblique rotations of A than orthogonal rotations.

Example 2: Comparing quartimax and varimax rotation

The earliest analytic rotations were orthogonal and based on the quartimax criterion. Varimax rotation was introduced to reduce the tendency of quartimax to weight too heavily on the first factor.

Table 10.5 contains quartimax and varimax rotations of the initial loading matrix A from the previous subsection.

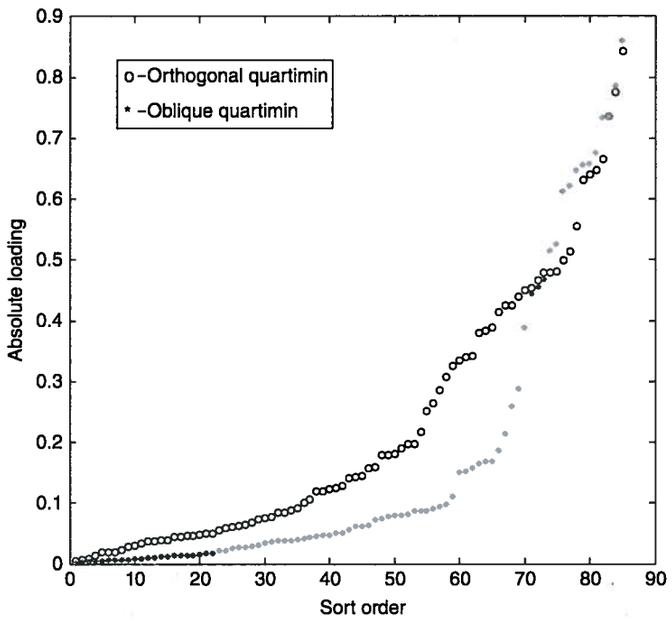

Figure 10.6 Comparison using SAL plots.

Table 10.5 Example 2 rotations

Quartimax					*Varimax*				
29	63	4	9	-5	17	65	12	18	-0
26	45	8	0	-2	17	47	15	9	3
33	66	4	-4	5	21	69	13	8	11
22	65	2	-3	2	12	66	9	5	6
33	64	4	6	1	21	66	13	17	6
38	18	48	6	3	20	18	55	17	9
42	13	39	5	1	27	14	47	16	8
41	20	48	4	-6	25	21	56	15	1
44	14	43	3	20	25	15	50	16	26
48	16	7	12	38	31	18	16	27	44
56	18	9	31	16	38	20	19	45	21
45	12	18	12	34	28	13	25	26	40
50	34	6	19	7	35	36	17	33	13
51	25	14	47	7	30	25	23	59	10
84	9	-10	-4	-12	82	18	10	17	2
74	11	9	-5	-2	66	17	25	14	11
78	6	-5	-14	-1	75	14	14	6	13

Varimax does seem to load the first factor more lightly and spread the loadings across factors more than quartimax does, which is as expected. The differences in this example, however, are not great, and neither is close to simple structure.

Quartimax rotation is the same as orthogonal quartimin rotation (Harman, 1976, p. 284). Thus the orthogonal quartimin rotation in Table 10.4 is the same as the quartimax rotation in Table 10.5.

Example 3: Exploratory bifactor rotation

Chen et al. used their data to compare confirmatory bifactor and two-stage factor analysis models for their quality of life data.

Building a bifactor model is a bit of a project. Chen et al. used a standard bifactor model for quality of life data based on a number of earlier studies and goodness of fit testing to help identify structural zeros. We will use exploratory bifactor analysis to see what bifactor model their data might suggest without any prior knowledge about the quality of life data.

An exploratory bifactor analysis of their data using the biquartimin criterion defined in Theorem 10.1 gave the loading matrix in Table 10.6. One can use Table 10.6 to suggest a bifactor model by setting all loadings with absolute value less than 0.02 equal to zero. This gives the loading matrix in Table 10.7.

The bifactor model suggested by Table 10.7 agrees exactly with Chen et al.'s standard model except for the three loadings on the third factor denoted by an "x." These were free loadings in Chen et al.'s model. In Chen et al.'s confirmatory analysis the loading estimates in these three positions were "insignificant." This motivated Chen et al. to suggest that the third group factor might be absorbed by the general factor and dropped. A less extreme alternative suggested by Table 10.7 would be to retain the third group factor, but add structural zeros in the positions containing an x. In any event, exploratory bifactor analysis has effortlessly discovered a bifactor model that is at least close to that found

Table 10.6 Exploratory bifactor rotation loading matrix.

0.46	0.52	0.02	0.09	0.00
0.38	0.36	0.05	0.00	0.02
0.51	0.53	-0.02	-0.06	0.00
0.41	0.55	-0.01	-0.05	-0.03
0.51	0.51	-0.00	0.04	0.01
0.53	-0.01	0.35	-0.01	-0.07
0.51	-0.04	0.29	0.00	0.02
0.53	0.03	0.41	0.00	0.01
0.60	-0.09	0.22	-0.09	-0.08
0.61	-0.08	-0.19	-0.03	-0.04
0.64	-0.03	-0.07	0.23	0.06
0.60	-0.12	-0.08	-0.03	-0.07
0.59	0.17	-0.03	0.15	0.08
0.64	0.04	0.00	0.39	-0.01
0.62	0.02	-0.02	0.06	0.60
0.63	-0.02	0.07	-0.02	0.39
0.60	-0.04	-0.02	-0.09	0.51

Table 10.7 Adjusted exploratory bifactor rotation loading matrix.

0.46	0.52	0	0	0
0.38	0.36	0	0	0
0.51	0.53	0	0	0
0.41	0.55	0	0	0
0.51	0.51	0	0	0
0.53	0	0.35	0	0
0.51	0	0.29	0	0
0.53	0	0.41	0	0
0.60	0	0.22	0	0
0.61	0	0	x	0
0.64	0	0	0.23	0
0.60	0	0	x	0
0.59	0	0	x	0
0.64	0	0	0.39	0
0.62	0	0	0	0.60
0.63	0	0	0	0.39
0.60	0	0	0	0.51

by Chen et al., and this suggests that exploratory bifactor analysis, which is a form of EFA, may be a valuable tool for constructing confirmatory bifactor models.

Discussion

We have given a primarily historical overview of rotation methods in EFA. This included a discussion of rotation, rotation criteria, rotation algorithms, choosing a rotation method, producing standard errors for rotated loadings, and some real data applications.

The problem of extracting initial loadings is fairly well resolved. Least squares and maximum likelihood extraction seem to work well.

We seem to have done very well on the algorithm problem. In a sense the problem is solved. The very general and simple pairwise line search algorithm of Browne and Cudeck and the faster gradient projection algorithm of Jennrich allow one to optimize almost any rotationcriterion. There is always room for improvement, but the need at present is not pressing.

The problem of assigning standard errors to factor loading estimates is also fairly well resolved using the linearization methods discussed, and in particular the infinitesimal jackknife method.

Choosing a good rotation method is a difficult problem. EFA is an exploratory method that suggests a relation between observed variables and unobserved factors. But different rotation methods suggest different relations. A rotation method that makes sense in one application may not be satisfactory in another. It is ultimately the user who must decide which rotation, if any, is best in a particular application.

For those seeking loading matrices with simple structure, quartimin and geomin seem to work well when it is possible to have such rotations.

When it is not possible to obtain rotations with simple structure, or structures with greater complexities are desired, little progress has been made to recommend a specific choice. The best we have at present are a few examples for which desirable rotation methods have been identified, for example geomin on Thurstone's box problem. In the complex case the problem of comparing rotation methods or choosing a good rotation method appears to be stalled. Some new ideas are needed.

We have recommended the use of numerical derivatives in several places. Psychometricians and more generally statisticians seem to fear numerical derivatives. What they should fear are the consequences of failing to try new methods because the derivatives required are too complex. The work described here suggests that numerical derivatives can produce accurate results and can greatly simplify what would otherwise be very complicated methods.

References

Anderson, T. W. & Rubin, H. (1956). Statistical inference in factor analysis. *Proc. Third Berkeley Symp. Math. Statist. Probability, 5*, 111–150.

Archer, C. O. & Jennrich, R. I. (1973). Standard errors for rotated factor loadings. *Psychometrika, 38*, 581–592.

Bentler, P. M. (1977). Factor simplicity index and transformations. *Psychometrika, 42*, 277–295.

Browne, M. W. (1984). Asymptotically distribution-free methods for the analysis of covariance structures. *British Journal of Mathematical and Statistical Psychology, 37*, 62–83.

Browne, M. W. (2001). An overview of analytic rotation in exploratory factor analysis. *Multivariate Behavioral Research, 36*, 111–150.

Browne, M. W., Cudeck, R., Tateneni, K. & Mels, G. (2002). CEFA: Comprehensive Exploratory Factor Analysis, Version 1.10 [Computer software and manual]. Retrieved from http://quantrm2.psy.ohio-state.edu/browne/

Carroll, J. B. (1953). An analytical solution for approximating simple structure in factor analysis. *Psychometrika, 18*, 23–28.

Carroll, J. B. (1960). IBM 704 program for generalized analytic rotation solution in factor analysis. Harvard University, unpublished.

Chen F. F., West S. G. & Sousa K. H. (2006). A comparison of bifactor and second-order models of the quality of life. *Multivariate Behavioral Research, 41*, 189–225.

Clarkson, D. B. (1979). Estimating the standard errors of rotated factor loadings by jackknifing. *Psychometrika, 44*, 297–314.

Crawford, C. B. & Ferguson, G. A. (1970). A general rotation criterion and its use in orthogonal rotation. *Psychometrika, 35*, 321–332.

Cudeck, R. & O'Dell, L. L. (1994). Application of standard error estimates in unrestricted factor analysis: Significance tests for factor loadings and correlations. *Psychological Bulletin, 115*, 475–487.

Cureton, E. E. & Mulaik, S. A. (1975). The weighted varimax rotation and the promax rotation. *Psychometrika, 40*, 183–195.

Girshick, M. A. (1939). On the sampling theory of roots of determinantal equations. *Annals of Mathematical Statistics, 10*, 203–224.

Harman, H. H. (1976). *Modern factor analysis* (3rd ed.). Chicago, IL: University of Chicago Press.

Jennrich, R. I. (1973). Standard errors for obliquely rotated factor loadings. *Psychometrika, 38*, 593–604.

Jennrich, R. I. (1974). Simplified formulae for standard errors in maximum likelihood factor analysis. *British Journal of Mathematical and Statistical Psychology, 27*, 122–131.

Jennrich, R. I. (2002). A simple general procedure for oblique rotation. *Psychometrika, 66*, 289–306.

Jennrich, R. I. (2004). Derivative free gradient projection algorithms for rotation. *Psychometrika, 69*, 475–480.

Jennrich, R. I. (2008). Nonparametric estimation of standard errors in covariance analysis using the infinitesimal jackknife. *Psychometrika, 73*, 579–594.

Jennrich, R. I. & Bentler, P. M. (2011). Exploratory bifactor analysis. *Psychometrika, 76*, 537–549.

Jennrich, R. I. & Sampson, P. F. (1966). Rotation for simple loadings. *Psychometrika, 31*, 313–323.

Jennrich, R. I. & Thayer, D. T. (1973). A note on Lawley's formulas for standard errors in maximum likelihood factor analysis. *Psychometrika, 38*, 571–580.

Jöreskog, K. G. (1969). A general approach to confirmatory maximum likelihood factor analysis. *Psychometrika, 34*, 183–202.

Lawley, D. N. (1967). Some new results on maximum likelihood factor analysis. *Proceedings of the Royal Society of Edinburgh, A67*, 256–264.

Lawley, D. N. & Maxwell, A. E. (1971). *Factor analysis as a statistical method.* New York, NY: Elsevier.

McCammon, R. B. (1966). Principal component analysis and its application in large-scale correlation studies. *Journal of Geology, 74*, 721–733.

McKeon, J. J. (1968). *Rotation for maximum association between factors and tests.* Unpublished manuscript, Biometric Laboratory, George Washington University.

Pennell, R. (1972). Routinely computable confidence intervals for factor loadings using the "jackknife." *British Journal of Mathematical and Statistical Psychology, 25*, 107–114.

Satorra, A. & Bentler, P. M. (1990). Model conditions for asymptotic robustness in the analysis of linear relations. *Computational Statistics and Data Analysis, 10*, 235–249.

Schmitt, T. A. & Sass, D. A. (2011). Rotation criteria and hypothesis testing for exploratory factor analysis: Implications for factor pattern loadings and interfactor correlations. *Educational and Psychological Measurement, 71910*, 95–113.

Silvey, D. S. (1971). *Statistical inference.* Baltimore, MD: Penguin Books.

Thurstone, L. L. (1935) *Vectors of the mind.* Chicago, IL: University of Chicago Press.

Thurstone, L. L. (1947) *Multiple factor analysis.* Chicago, IL: University of Chicago Press.

Wexler, N. (1968). *An evaluation of an asymptotic formula for factor loading variance by random methods.* Unpublished doctoral dissertation. Rutgers University.

Yates, A. (1987). *Multivariate exploratory data analysis: A perspective on exploratory factor analysis.* Albany, NY: State University of New York Press.

Zhang, G., Preacher, K. J. & Jennrich, R. I. (2012). The infinitesimal jackknife with exploratory factor analysis. *Psychometrika, 77*, 634–648.