The Case for Hyperplane Fitting Rotations in Factor Analysis: A Comparative Study of Simple Structure

James S. Fleming

Southwest Psychometrics and Psychology Resources

Abstract: Hyperplane fitting factor rotations perform better than conventional rotations in attaining simple structure for complex configurations. Hyperplane rotations are reviewed and then compared using familiar examples from the literature selected to vary in complexity. Included is a new method for fitting hyperplanes, hypermax, which updates the work of Horst (1941) and Derflinger and Kaiser (1989). Hypercon, a method for confirmatory target rotation, is a natural extension. These performed very well when compared with selected hyperplane and conventional rotations. The concluding sections consider the pros and cons of each method.

Key words: Factor analysis, factor rotation, hyperplane fitting, simple structure.

1. Introduction: Hyperplane Fitting and Simple Structure

Hyperplane fitting rotations (HFRs) are special cases of the more general class of procrustean rotations in factor analysis that attempt to fit a given factor pattern matrix to a partially specified target matrix. With HFRs the target matrix contains zero loadings in specified positions. This article reviews and then compares the HFRs one to the other and to other popular rotations using well-known examples to assess simple structure. Two new methods, the exploratory hypermax and confirmatory hypercon, update the methods of Horst (1941) and Derflinger and Kaiser (1989). It is argued that hyperplane fitting rotations are superior to familiar analytic rotations when factor structures are complex (variables load on more than one factor). Fortunately too, these also usually perform well with low complexity data; hence their applicability is fairly general.

Numerous studies have compared factor rotations; to cite just a few, there are the classic studies of Hakstian (1971) and Hakstian and Abell (1974), and more recently, of Jennrich (2006, 2007), Asparouhov and Muthén (2009), and Sass and Schmidt (2010). Examples for this study were chosen to be: (a) familiar to psychologists, (b) of interest to behavioral scientists, and (c) varied in

complexity. The concepts of simple structure, complexity, and hyperplane fitting are considered next.

1.1 Simple Structure as the Goal of Rotation in Factor Analysis

In the early days of factor analysis rotations were not always conducted. Spearman (1904) believed that a single factor provided evidence for a general factor of intelligence, accounting for most of the correlations between his mental tests. Other early researchers found multiple factors of intelligence, yet failed to rotate their solutions. But Thurstone (1947) argued for multiple factors and insisted that these are more interpretable when rotated, for only thus could simple structure be attained. His criteria were:

1. Each row of the factor pattern matrix should have at least one zero.

2. For m common factors, each column should have at least m zeros.

3. Every pair of columns should have several entries that vanish in one but not the other.

4. For every column pair a large proportion of variables should have vanishing entries in both columns, if there are four or more factors.

5. For every column pair there should only be a small number of non-vanishing entries in both columns.

The first condition is the most crucial, whereas the others provide stability in over determining the model (Yates, 1987; Browne, 2001). Under fairly general conditions it is possible to find $m - 1$ zeros for each column (Browne, 2001). In practice researchers can focus mainly on the first point. After this, all else being equal, a decision can be made as to which of two or more rotations works best for a given problem by subjective judgment (Asparouhov and Muthén, 2009; Browne, 2001; Rozeboom, 1992).

Multidimensional test constructors seek "perfect" or complexity one structure, in which each variable loads on only one factor. Rotated solutions, when they show that variables have high loadings on their appropriate factors and low loadings on the rest, are often presented in support of a test's validity. But drawing the line between high and low loadings (.30 or .40, for instance) is subjective (Cudeck and O'Dell, 1994).

Complexity one structure is difficult to achieve. Setting too many loadings to zero in confirmatory factor analysis (CFA) can have a distorting effect on the solution, producing, for example, factor correlations that are too high (Asparouhov and Muthén, 2009). But the problem is not unique to CFA, as the magnitude

of the correlations between factors in exploratory factor analysis (EFA) depends in part on the method of rotation. A common misunderstanding of Thurstonian principles is to assume that finding the simplest structure is synonymous solely with reducing complexity, as it is sometimes measured by simplicity indexes (Kaiser, 1974; Bentler, 1977; and Lorenzo-Seva, 2003). These are useful when complexity one structure is sought (Fleming, 2003). But simply lowering mid-range loadings can elevate these. Rather, assessment should be based on locating the near zero elements and minimizing these: "It is generally felt, however, that a loading matrix with a fair number of small values is simpler than one with mostly intermediate values" (Jennrich, 2006, p. 180).

Popular rotations designed to minimize complexity do produce good simple structure solutions when the underlying structure is even moderately complex (Browne *et al.*, 2010). However, they perform poorly when many complexities are present.

1.2 Hyperplane Resolution and Simple Structure

Planar rotations of axes in m dimensions attempt to maximize the hyperplane count, or number of zero loadings, by careful placement of the factor axes close to the boundaries of the hyperplanes. This is done with HFRs by forming target matrices containing zeros in specified positions, with remaining elements free to vary. A discrepancy function based on the difference between the unrotated and target rotated matrices is then minimized. This can be expressed as: $\mathbf{\Lambda} = \mathbf{AT} - \mathbf{B}$, where $\mathbf{A}(p \times m)$ is the factor pattern matrix to be transformed, $\mathbf{T}(m \times m)$ the transformation matrix, $\mathbf{\Lambda}$ the new pattern, and $\mathbf{B}(p \times m)$ the partially specified target. The function to be minimized is:

$$Q(\mathbf{\Lambda}) = ||\mathbf{AT} - \mathbf{B}||^2. \tag{1}$$

$\mathbf{A}$ itself may be an orthogonal rotation of an initial factor matrix $\mathbf{F}$, in which case the scaling of $\mathbf{F}$ differs among the methods; and some HFRs require that $\text{diag}(\mathbf{T}^{-1}\mathbf{T}^{-1'}) = \mathbf{I}$. Methods thus differ in (a) the scaling of the variables prior to a pre-rotation, (b) the specification of $\mathbf{B}$, and (c) whether or not $\mathbf{T}$ is restricted; thus we have either a restricted or an unrestricted least-squares problem, the former requiring multiple iterations. Details are given in Section 2 for hypermax and other HFR methods are reviewed in 3.1 and 3.2.

1.3 Criterion for Comparing Rotations

There is no recognized "gold standard" for assessing simple structure (Jennrich, 2007, p. 323), but the method proposed here, comparing averages (root-mean-squares, or RMS) of loadings in the hyperplane, seems appropriate for

this task. In all examples there is a clear break between hyperplane and salient elements for hypermax and for most other rotations as determined by sorted absolute loadings (SAL) plots (Jennrich, 2006, 2007). Those below the break are considered to be in the hyperplane, and those with the lowest RMS for these are considered to be the best rotations (see Figure 1 for SAL examples).
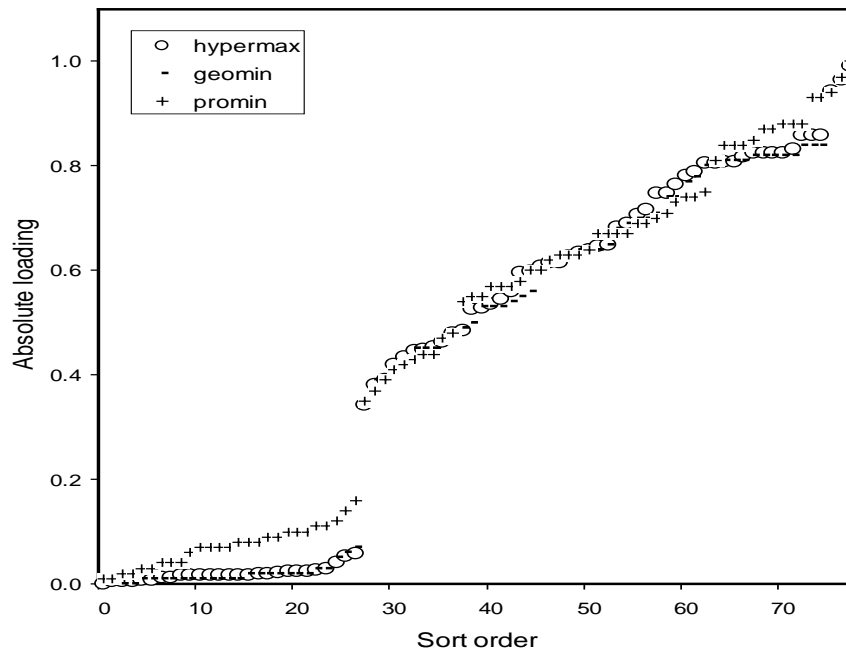

Figure 1: Sorted absolute loading plot illustrating selected rotations based on Thurstone's 26-variable box problem

2. The Hypermax Rotation

Psychometric work in hyperplane fitting dates back to Tucker (1940) and Horst (1941). Tucker sought to minimize potential hyperplane loadings whereas Horst's method, as modified by Derflinger and Kaiser (1989; called "Horst-Hilsch", or HH), minimizes instead the ratio of hyperplane to salient (non-hyperplane) loadings. (Horst maximized the ratio of the salient to total elements, but this is equivalent to minimizing the preceding ratio). Both presume that there is a way to identify the hyperplane elements in the target matrix. When achieved by an automated algorithm, the rotation is exploratory; when the zeros are pre-specified, it is confirmatory. We shall refer to the rotation part of the HH algorithm as the Horst-Derflinger-Kaiser (HDK) method. Let $\mathbf{A}(p \times m)$ be a row normalized pattern matrix resulting from a varimax rotation. Then find for each factor:

$\mathbf{K}_j = \mathbf{A}_o'\mathbf{A}_o$, a matrix $(m \times m)$, where $\mathbf{A}_o$ consists only of the rows of $\mathbf{A}$ that are in the hyperplane for factor j.

$\mathbf{L}_j = \mathbf{A}_x'\mathbf{A}_x$, a matrix $(m \times m)$, where $\mathbf{A}_x$ consists of the remaining rows of $\mathbf{A}$.

$\delta_j = (\boldsymbol{t}_j'\mathbf{K}_j\boldsymbol{t}_j)/(\boldsymbol{t}_j'\mathbf{L}_j\boldsymbol{t}_j)$, the ratio of two quadratic forms, is the quantity to be minimized. Min(δ_j) is the resulting smallest eigenvalue, and $\boldsymbol{t}_j$ is its corresponding eigenvector. Differentiating leads to the eigen equation:

$$(\mathbf{K}_j - \delta_j\mathbf{L}_j)\boldsymbol{t}_j = \mathbf{0}. \tag{2}$$

After determining each $\boldsymbol{t}_j$ the matrix $\mathbf{T}(m \times m) = [\boldsymbol{t}_1, \cdots, \boldsymbol{t}_m]$ is assembled and the transformation can proceed as $\boldsymbol{\Lambda}^* = \mathbf{AT}$. To obtain the factor pattern $\boldsymbol{\Lambda}$, a final rescaling is needed. Letting $\mathbf{D} = \text{diag}(\mathbf{T}'\mathbf{T})^{-1}$, compute the pattern as:

$$\boldsymbol{\Lambda} = \boldsymbol{\Lambda}^*\mathbf{D}^{-(1/2)}. \tag{3}$$

Tucker's method was employed by Lawley and Maxwell (1964), Jöreskog (1966), and Kaiser and Cerny (1978). But Derflinger and Kaiser found that this algorithm did not work well with certain problems. Also, as the ratio of two quadratic forms, their procedure (unlike Tucker's) is a scale-free method. For the Tucker method the minimization requires solving for the smallest eigenvalue and corresponding eigenvector of $(\mathbf{K}_j - \delta_j\mathbf{I})\boldsymbol{t}_j = \mathbf{0}$. But unlike (2), this equation cannot be solved uniquely without a further restriction on the eigenvector length; so the usual restriction is that $\boldsymbol{t}_j'\boldsymbol{t}_j - 1 = 0$.

2.1 The Hypermax Algorithm in Four Steps

The HH rotation updated Horst by defining a method for identifying hyperplane elements (step 2 below) and introducing the HDK rotation (discussed above and in step 3). For hypermax, the weighted varimax pre-rotation was introduced (step 1) as well as the double-iteration, which was added to provide greater accuracy (step 4).

1. *Weighted varimax pre-rotation and row normalization.* Both the Kaiser-Cerny (KC) and HH methods use varimax as a preliminary rotation of an initial factor matrix before applying the HDK transformation. Although HH was an improvement over the KC rotation, it still did not work in the presence of high complexity. Thurstone's (1947) 26-variable box problem is one example for which this, as do most rotations, fails to recover the known simple structure. The hypermax fix for this problem is to use weighted varimax (Cureton and Mulaik, 1975) as the preliminary rotation. An alternative to Kaiser's row normalization, this gives greater weight to variables

closer to the boundary of the hyperplanes. The effect is to better handle variable complexity.

Following this initial rotation, the variables should again be Kaiser (row) normalized before applying the final rotation; otherwise the results can be quite distorted.

2. *Identifying hyperplane elements.* To identify hyperplane elements, Derflinger and Kaiser compute a function for each column j of this rotated factor pattern $\mathbf{A}$ as:

$$f_j = \left[\sum_i (a_{ij}^2)^{1/b}/p \right]^{b/2}, \tag{4}$$

where f_j is an absolute power mean that can be varied with the parameter b, and p is the number of variables. The a_{ij}^2 are squared elements of $\mathbf{A}$. They stated that selecting $b = 4$ was consistently the most effective in their experience. The computations in this case are: $f_j = $ (mean of square-roots of absolute loading values)^2. The KC, HH, and hypermax rotations all employ this method for identifying hyperplane elements.

3. *Applying the HDK rotation in hypermax, and row and column rescaling.* When Kaiser normalization is applied to the rows, it is customary following the HDK rotation to return them to the original variable metric by denormalizing. But column rescaling must also be done, so the rotation is completed by applying (3).

4. *Double iteration.* This is an important addition to the KC, HH, and hypermax rotations. If, following the final transformation, f_j is recomputed via (4), occasionally it will identify additional elements in the hyperplane for one or more of the factors. When this is the case, the rotation should be repeated to include these elements.

3. Other Rotations Used in This Study

3.1 An Overview of Exploratory Hyperplane Rotations

Promax (Hendrickson and White, 1964) is a straightforward unrestricted least-squares solution to (1) in which the target matrix $\mathbf{B}$ is formed by powering elements of a varimax or other orthogonally rotated matrix (usually powers of 2, 3, or 4) while retaining the signs of the original loadings. Rows of the target are normed, and columns are rescaled so that the largest loading is 1.0. Note

that with promax the target contains approximate rather than exact zeros. The final step is a column rescaling using (3).

Promaj (Trendafilov, 1994), a method similar to promax, uses the technique of vector majorization (see Marshall and Olkin, 1979, for theoretical rationale), which slightly lowers the size of the salient loadings, but replaces the lowest with exact zeroes, purportedly for a better fit. Let $\boldsymbol{a}_j$ be a column of loading matrix $\mathbf{A}$ to be majorized, having elements $\boldsymbol{a}_{ij}$. We then find corresponding vector $\boldsymbol{b}_j$ for the target matrix $\mathbf{B}$ as: $\boldsymbol{b}_{ij} = \mathrm{sign}(\boldsymbol{a}_{ij})[\max(\boldsymbol{a}_{ij}^2 - c_j, 0)]^{1/2}$. The constant c_j is computed as follows. Let μ_j be the mean of the $\boldsymbol{a}_{ij}^2$, then let c_j be the value of these squared loadings that is closest to, but less than, μ_j.

Promin (Lorenzo-Seva, 1999) is similar to hypermax because it begins with a preliminary weighted varimax rotation, but salient elements are determined as the mean of the squared loadings plus 1/4 of the standard deviation of each column of the rotated pattern. This is followed by applying Browne's (1972) restricted procrustean rotation (see 3.3). Lorenzo-Seva viewed promin as an improvement over promax, and also as a simpler approximation to Kier's (1994) simplimax, a method that is considered next.

With simplimax (1) is minimized by restricted least-squares, so that $\mathrm{diag}(\mathbf{T}^{-1} \mathbf{T}^{-1'}) = \mathbf{I}$ at each step. Simplimax proceeds by alternation, first minimizing the transformation matrix $\mathbf{T}$ for a given $\mathbf{B}$, then vice versa, by using many random orthogonal starts for $\mathbf{T}$ with different numbers of zeros for each. Thus it uses more random starts than other methods. Each start is rotated by Browne's method, as with promin. One then decides on the final solution by selecting over the range of solutions with differing hyperplane counts. This is not necessarily the function with the minimum value; instead one may plot the functions to find a break, similar to the scree test for the number of factors.

3.2 Confirmatory Hyperplane Fitting Rotations: Browne's Method and Hypercon

Browne (1972) showed how to rotate a factor matrix $\mathbf{A}$ to fit a partially specified target matrix $\mathbf{B}$ directly while restricting the transformation vectors to unit length. He rotated $\mathbf{A}$ column by column in a sequence of pairwise elementary rotations, transforming one factor into the plane of the other at each step.

Let $\mathbf{B}$ be the partially pre-specified target matrix and $\mathbf{C}$ consists of zeros and ones, to match the target, with c_{ij} and b_{ij} as elements of $\mathbf{C}$ and $\mathbf{B}$, respectively, and λ_{ij} an element of rotated matrix $\mathbf{\Lambda}$. The function to be minimized can be expressed as:

$$f(\mathbf{\Lambda}) = \sum_{i=1}^{p} \sum_{j=1}^{m} c_{ij}(\lambda_{ij} - b_{ij})^2$$

using Newton-Raphson iterations, with the restriction on unit vector length maintained at each step. As a confirmatory rotation in the CEFA program (Browne *et al.*, 2010), Browne's method is called TARROT (TARget ROTation).

The hypercon rotation is simply the application of the HDK rotation to a target matrix in which the zero elements are pre-specified. Unlike TARROT, hypercon is not a direct rotation; thus the rotated matrix must be rescaled column-wise using (3). It is less general than TARROT because it only applies to hyperplane fitting, whereas TARROT is a general procrustes method. But also unlike TARROT, hypercon is non-iterative, and is therefore very fast. As will be seen, hypermax/hypercon and Browne's method in simplimax/TARROT, give virtually identical results in the ensuing examples.

3.3 Other (Non-Hyperplane Fitting) Rotations Used for Comparison

Additional rotations, selected from the most popular, include (1) the orthogonal varimax (Kaiser, 1958); (2) direct oblimin (Jennrich and Sampson, 1966), for which two values of the obliqueness parameter were used: 0 (quartimin) and -3; (3) the Harris-Kaiser (1964) orthoblique rotation with eigenvalue power parameter set to 0 ("independent clusters") and to .5 ("proportional", see p. 361) for their "Class II" methods; (4 and 5) weighted promax (Cureton and Mulaik, 1975); and weighted oblimin (Lorenzo-Seva, 2000), both of which use the weighted varimax pre-rotation; and (6) geomin (Yates, 1987; Browne, 2001).

Though not an HFR (a target matrix is not formed), geomin nevertheless is also useful in detecting near zero elements. It is an extension of Thurstone's idea that, for a given row, the product of the loadings would be zero if at least one element is zero, which satisfies his most important condition for simple structure. The function to be minimized is (Bernaards and Jennrich, 2005): $f(\mathbf{\Lambda}) = \mathbf{u}^{'} \exp\{m^{-1}[\log_e(\mathbf{\Lambda}^2 + \varepsilon)]\}\mathbf{v}$, where ε is a small constant (e.g., .01) that can vary with the number of factors, $\mathbf{\Lambda}$ the transformed loadings, and $\mathbf{u}^{'}$ and $\mathbf{v}$ are conformable summing vectors.

4. Examples

4.1 Thurstone's Box Problems as Exemplars of Complexity

Because the test of the adequacy of hyperplane fitting rotations is how well they can uncover the simplest structure when variables are highly complex, Thurstone (1947) devised his 26-variable box problem as a challenge. This problem, consisting of a series of non-linear functions of the dimensions of 30 boxes, has a clear simple structure in three dimensions: length, width, and height. The ability to uncover this structure can only be met by a few analytic rotations. The upper

portion of Table 1 compares these. The best results are shaded. All results except for hypermax and geomin were taken from their original published articles. Methods of factor extraction, which differed among published studies, and programs used for rotation, are also given. These include CEFA, Factor (Lorenzo-Seva and Ferrando, 2006), and Explorer (Fleming, 2011). (Different programs were needed for the examples because each had some rotations not found in others.)

Table 1: Summary measures for comparing box problem rotations

Rotation	No. in hyper-plane	RMS of hyperplane loadings	Extraction method	Program for rotation
26-Variable Box Problem[a]				
Hypermax	27	.023	ULS	Explorer
Simplimax	27	.024	PAFA	unknown
Geomin	27	.025	ULS	CEFA
Weighted promax	27	.036	PAFA	unknown
Weighted Oblimin	21	.070	PAFA	unknown
Promin	22	.082[b]	Centroid[c]	unknown
20-Variable Box Problem				
Hypermax	27	.026	PCA	Explorer
Horst-Hilsch	27	.026	PCA	Explorer
Simplimax	27	.026	PCA	Factor
Geomin	27	.026	PCA	CEFA
Weighted oblimin	27	.029	PCA	Factor
Orthoblique (power = .5)	27	.035	PCA	Explorer
Oblimin (parameter = -3)	27	.044	PCA	Explorer
Promax (power = 2)	27	.045	PCA	Explorer
Kaiser-Cerny	25	.050	PCA	Explorer
Weighted Promax	25	.051	PCA	Explorer
Promaj	26	.053	PCA	Explorer
Oblimin (power = 0)	27	.062	PCA	Explorer
Promin	23	.075	PCA	Factor
Varimax	13	.114	PCA	Explorer
Promax (power = 4)	6	.130	PCA	Explorer
Orthoblique (power = 0)	6	.133	PCA	Explorer

Note: RMS = root-mean-square. PCA = principal component analysis, PAFA = principal axis factor analysis, ULS = unweighted (ordinary) least-squares. Best results are shaded in these tables.
[a]All results except for hypermax and geomin are from previously published studies.
[b]Includes two misspecified loadings that should have been in the hyperplane.
[c]Centroid is an older method of factor extraction, used before PAFA became practical with the advent of computers. The factor analysis here was based on such an earlier study.

Hypermax, simplimax, and geomin gave the very best results, each correctly identifying the 27 hyperplane elements from the known solution within a very narrow $0. \pm .10$ band. Weighted promax was not far behind. If this critical band is extended to $0. \pm .20$, then weighted oblimin also identified all 27 correctly; but promin misspecified two elements as being in the hyperplane that were not (see Lorenzo-Seva, 1999). (The hypermax factor solutions for these examples may be found on the author's website.)

SAL plots in Figure 1 illustrate the method by which hyperplane elements are identified. Selected comparisons of three rotations are shown. Note the similarity between hypermax and geomin, but also how these differ from promin.

A second example tests the relative effectiveness of several rotations using Thurstone's 20-variable box problem, an exemplar frequently used to compare rotations. This example also has many complexities, but unlike the 26-variable problem, most conventional rotations can do a fair job at uncovering the structure.

Results are shown in the lower part of Table 1. Here again the same three, hypermax, simplimax, and geomin, gave the best, and essentially identical, results, but so did HH; and weighted oblimin was close behind. Other results varied. It is not surprising that varimax and the orthoblique independent clusters method did poorly by comparison; the former is orthogonal rather than oblique, and the latter is most suitable when perfect structure is expected.

4.2 WAIS-R Subtest Factors

Eleven subtests of the WAIS-R intelligence test (Wechsler, 1981) were factor analyzed using unweighted least-squares (ULS). These factors are general-verbal, numerical, and spatial, respectively. The WAIS-R presents a problem of moderate complexity. Three subtests have loadings $> .25$ on two factors.

For this problem, oblimin with obliquity parameter of -3 was not included in the results because it produced a first, general factor and did not closely resemble any of the other solutions given here. The better rotations found 16-18 loadings $\leq .15$ in absolute value and 17-19 $\leq .20$. The rotations are ranked in ascending order in Table 2, based on the RMS for the latter nineteen. While most of these exhibited simple structure, two that did not do well were KC and varimax. Hypermax fared best with RMS $= .091$. This is not to quibble, as others are on a par as well, but nonetheless it demonstrates that hypermax is again among the best. But notice also that promax with power $= 4$ also did very well, and better, in fact, than its competitors, promin and promaj.

4.3 Two Classic Examples

Table 2: Summary measures for comparing rotations for WAIS-R subtests

Rotation	Absolute loadings $\leq .15$	Absolute loadings $\leq .20$	RMS of 19 lowest
Hypermax[e]	18	19	.091
Promax[e] (power = 4)	17	18	.096
Oblimin[e] (parameter = 0)	17	17	.097
Simplimax[f]	17	18	.097
Promaj[e]	17	19	.098
Weighted oblimin[e]	16	18	.099
Orthoblique[e] (power = 0)	18	18	.102
Geomin[c]	16	18	.102
Promin[f]	16	19	.104
Weighted promax[e]	16	19	.120
Horst-Hilsch[e]	16	19	.136
Orthoblique[e] (power = .5)	15	17	.144
Promax[e] (power = 2)	15	17	.147
Kaiser-Cerny[e]	8	16	.171
Varimax[e]	0	3	.289

[c]CEFA program. [f]Factor program. [e]Explorer program.

Two classic examples from the literature are: Thurstone's (1938) Ten Primary Mental Abilities, and Eight Physical Variables (Harman, 1976, p. 22). Three factors were extracted from the first, and two for the second via ULS. Ten PMA has factors of verbal, arithmetic, and spatial abilities. There were cross-loadings $\geq .40$ for two variables in this example. The better rotations maximized the number of loadings $\leq .20$. As seen in Table 3 (top), many rotations produced pattern matrices with 16-17 loadings $\leq .20$, but the top five had comparable low RMSs.

The variables loading on the physical factors exhibited low complexity for most solutions, with hypermax again among the best (Table 3, bottom). But three did less well on RMS: orthoblique (power = .05), oblimin (-3), and varimax.

4.4 A Confirmatory Illustration: Twenty-Four Psychological Tests

This example illustrates hypercon as a confirmatory method, given a predefined target matrix from Mulaik (1972, p. 307). He presented a promax target matrix that was used with Jöreskog's (1966) procedure. This target was used to rotate a principal factors solution from Harman (1976, p. 124). Comparisons between Mulaik's solution and the one here are imperfect as his was derived from

Table 3: Summary measures for comparing rotations for ten primary mental abilities and eight physical variables

Example Rotation	No. in hyper-plane	RMS of hyperplane loadings
Ten PMA		
Promax[e] (power = 2)	17	.073
Hypermax[e]	17	.075
Horst-Hilsch[e]	17	.075
Simplimax[f]	16	.076
Promin[f]	16	.076
Promaj[e]	17	.101
Oblimin[e] (power = 0)	17	.105
Geomin[c]	17	.106
Weighted oblimin[f]	16	.108
Weighted promax[e]	16	.110
Orthoblique[e] (power = 0)	15	.114
Oblimin[e] (parameter = −3)	16	.117
Orthoblique[e] (power = .5)	15	.118
Promax[e] (power = 4)	15	.117
Kaiser-Cerny[e]	9	.136
Varimax[e]	12	.169
Eight physical		
Simplimax[f]	8	.053
Promin[f]	8	.053
Hypermax[e]	8	.054
Horst-Hilsch[e]	8	.054
Promax[e] (power = 4)	8	.054
Promax[e] (power = 2)	8	.054
Oblimin[e] (power = 0)	8	.054
Orthoblique[e] (power = 0)	8	.054
Promaj[e]	8	.054
Kaiser-Cerny[e]	8	.054
Weighted promax[e]	8	.055
Geomin[c]	8	.059
Weighted oblimin[f]	7	.059
Orthoblique[e] (power = .5)	4	.072
Oblimin[e] (parameter = −3)	8	.105
Varimax[e]	0	.210

Note: Hyperplane width for Ten PMA was $0. + .20$; for Eight physical it was $0. + .10$.
[c]CEFA program. [f]Factor program. [e]Explorer program.

a centroid factoring. (It is not appropriate to apply weighted varimax to centroid solutions.) But present results could more meaningfully be compared to Browne's TARROT. For hypercon and TARROT, all targeted coefficients were $\leq .25$ in absolute value. Likewise, for Mulaik's factor pattern based on Jöreskog's method, as well as TARROT, all targeted elements were $\leq .25$.

Evaluating the similarities between TARROT and hypercon was further fa-

cilitated by computing congruence coefficients (Tucker, 1951) for corresponding factors as well as their RMS differences. Coefficients for corresponding factors were .997, .989, .998, and .999; the RMS differences between were .028, .046, .026, and .025. Interestingly, when one solution was rotated to approximate the other by procrustes matching (Korth and Tucker, 1976; Fleming, 1991), the resulting congruence coefficients all were exactly 1.000 to 3 decimals and all RMS differences were < .02. TARROT and hypercon, then, gave virtually identical results. The factor loading matrix for hypermax is shown in Table 4.

Table 4: Factor pattern for hypercon target rotation of 24 psychological tests

No.	Factor			
	I	II	III	IV
1	**-0.04**	**0.09**	0.70	**0.02**
2	**-0.01**	**0.01**	0.50	**-0.04**
3	**0.00**	**-0.13**	0.64	**0.01**
4	**0.11**	**-0.01**	0.60	**-0.07**
5	0.77	**0.11**	**0.05**	**-0.05**
6	0.79	**-0.06**	**0.04**	**0.06**
7	0.86	**0.05**	**0.06**	**-0.13**
8	0.54	**0.12**	0.28	**-0.05**
9	0.83	**-0.10**	**0.04**	**0.05**
10	**0.11**	0.79	**-0.23**	**0.08**
11	**0.06**	0.50	**-0.05**	0.35
12	**-0.11**	0.72	**0.13**	**-0.02**
13	**0.02**	0.51	0.40	**-0.08**
14	**0.12**	**-0.07**	**-0.10**	0.62
15	**0.02**	**-0.05**	**-0.02**	0.57
16	**-0.12**	**-0.11**	0.36	0.52
17	**0.02**	**0.09**	**-0.10**	0.64
18	**-0.16**	**0.24**	**0.20**	0.45
19	**0.00**	**0.05**	**0.17**	0.37
20	0.28	**-0.05**	0.36	**0.19**
21	**0.01**	0.34	0.37	**0.09**
22	0.24	**-0.03**	0.34	0.20
23	0.23	**0.09**	0.48	**0.07**
24	0.26	0.40	**0.04**	0.19

Factor
Intercorrelations:

1.00			
0.32	1.00		
0.48	0.33	1.00	
0.48	0.49	0.49	1.00

Note: The 62 elements corresponding to Mulaik's (1972) targeted zeros are bolded. Iterated principal factors was the method used for factor extraction.

5. How Hypermax Can Fail−and What to Do about It

HFRs that depend on weighted varimax can sometimes fail, as in Browne's (2001, pp. 138-141) example. Here 2 of 12 variables had low communalities in a three factor solution, but they are nonetheless of complexity 2 after weighting. Some factoring methods−notably alpha and image analysis−tacitly assume that variables are sampled from a well defined domain; hence they should have reasonably high communalities. So a realistic requirement for hypermax might be that variables with low communalities be eliminated. But this is also why it is a good idea to try more than one rotation.

There are also cases in which (4) can fail to find the correct cutoff for salient loadings. Asparouhov and Muthén (2009) give an example for ten variables with the largest five loadings for each defining two factors. But the second factor also has two cross-loadings of .25−not large, yet not insignificant. Using simulation, they found that geomin consistently produced the correct solution, but quartimin did not. With this same problem (factored by ULS for the recreated model covariance matrix), hypermax assigned these loading to the hyperplane. The solution was "not bad" (Table 5) yet it was not perfectly optimal; but note that the misspecification still did not force those loadings to zero as in CFA. Note also that simplimax fared even worse for both 8 specified zeros (simple structure) and 10 (complex structure). But when the 8 hyperplane values were correctly specified, hypercon exactly produced the optimal solution. Thus, when targeted elements appear to be misspecified (perhaps $> .15$ in a large sample), researchers can repeat the rotation by re-specifying the hyperplane elements, and then apply hypercon.

Table 5: Two-factor Example Based on Asparouhov and Muthén (2009)

	Model Factors		Hypermax Factors		Simplimax Factors (8 zeros)		Simplimax Factors (10 zeros)		Uniqueness
1	.80	.00	.84	-.08	.85	-.13	.85	-.10	.36
2	.80	.00	.84	-.08	.85	-.13	.85	-.10	.36
3	.80	.00	.84	-.08	.85	-.13	.85	-.10	.36
4	.80	.25	.84	.17	.90	.10	.86	.15	.36
5	.80	.25	.84	.17	.90	.10	.86	.15	.36
6	.00	.80	.00	.80	.15	.72	.01	.79	.36
7	.00	.80	.00	.80	.15	.72	.01	.79	.36
8	.00	.80	.00	.80	.15	.72	.01	.79	.36
9	.00	.80	.00	.80	.15	.72	.01	.79	.36
10	.00	.80	.00	.80	.15	.72	.01	.79	.36

Note: Factor correlations were .50 for the model, .57 for hypermax, .48 for simplimax with 8 zeros specified, and .58 for simplimax with 10 zeros specified. The factors are not standardized.

Finally, it is possible for the column rank of the target matrix to be $< m$. Though unlikely in practice, this would call for some modification of the target.

6. Summary of Findings and Limitations

It was not possible to evaluate all methods of rotations in a single study and the number of examples was of necessity limited. Further research should compare the best of these rotations with others using more real and constructed examples, or by simulation. However, certain definite conclusions can be drawn from the present examples.

1. Except when the factor structure is complex, none of the oblique methods considered here are really bad rotations. If simple structure is truly the criterion, then all were better choices than varimax for these examples.

2. With the highly complex 26-variable box problem, few rotations are acceptable; and hypermax was equal to the best of these.

3. With moderately complex structure (Thurstone's 20-variable box problem; WAIS-R; and 10 PMA), most rotations are acceptable, although some are clearly better than others. Again, hypermax was equal to the best.

4. With Eight Physical Variables, the least complex example, all the rotations worked very well, with the exception of varimax and orthoblique with power parameter $= .5$, and oblimin with parameter -3. But orthoblique independent clusters exhibited very simple structure. On the other hand, the reverse was true for the 20-variable box problem. These results confirmed expectation (Harris and Kaiser, 1966); namely, that degree of complexity should moderate the choice of the power parameter for this class of rotations.

5. Consistently, hypermax and simplimax gave the very best solutions in the empirical examples. However, simplimax is computationally much more intensive.

6. Simplimax and geomin both not only require selecting from several solutions, but optimizing their fitting functions does not necessarily minimize the RMS for loadings in the hyperplane; and finding the best fit for geomin may require trying different values of the epsilon parameter. Also, geomin can sometimes fail when variables have three or more nonzero loadings (Asparouhov and Muthén, 2009).

7. In some examples KC did noticeably worse than other rotations.

8. Promax is still not a bad choice. However, it is important to look at more than one solution−raising loadings to the second power works best for some solutions, but the fourth power can work with others. Also, weighted promax should be tried for highly complex problems. In all, promax held up surprisingly well compared with its newer competitors, promin and promaj; but further study comparing these methods is needed.

9. Although computer speed is no longer the concern that it once was (as computers are now so fast), it is worth mentioning that, being non-iterative, hypermax is very fast. A colleague who tested them independently reported that hypermax was 38 times faster than geomin for Thurstone's 26-variable box problem.

10. Browne's algorithm, used by both simplimax and TARROT, is an effective tool for hyperplane fitting. However, hypercon performed as well as Browne's TARROT with twenty-four psychological tests, and as was seen, simplimax (also using Browne's algorithm) and hypermax produced virtually identical results. But unlike the HDK procedure, Browne's method requires multiple pairwise passes though the columns of the loading matrix, subject to the restriction that the transformation vectors be of unit length at each step. How is it, then, that these methods resulted in virtually identical factor matrices?

    One clue is that once a sufficient number of zero elements is determined then one has, in essence, a model that is well over-identified; hence any linear transformation that preserves the zeros should be unique up to a column rescaling. That is why not only hypermax but also HH and KC sometimes give, upon column rescaling, identical results to Browne's method. But more importantly, because hypermax and hypercon are scale-free methods, optimizing the reference structure is equivalent to optimizing the pattern.

    Browne's target rotation is, however, a more general procedure for procrustes rotation, whereas hypermax and hypercon are limited to hyperplane fitting.

11. SAL plots are analogous to the scree test for the number of factors, in that there is often a clear break between the smallest and larger loadings. The points below the break are presumably in the hyperplane; and therefore do not differ substantially from zero. Fortunately, the break points were very clear in these examples. Unfortunately, the breaks are not always so crisp (see Jennrich, 2006, p. 182, Figure 3). But even in those cases, one may still judge one rotation as better than another in a relative sense, based on the closeness (RMS) of the smallest points to the baseline. Despite this

weakness, RMS measures based on SAL plots may be the best method for comparing rotations.

7. Conclusions

One significant finding is that results so far suggest that, as long as the elements in the hyperplane are properly identified, the HDK rotation is very effective in producing an optimal solution, as seen with HH, hypermax and hypercon. Thus it seems surprising that Horst's 1941 article has received scant attention in the factor rotation literature.

Hypermax showed great promise in the empirical examples, in which hypermax and simplimax consistently performed best in locating hyperplanes. Hypermax, however, is computationally much more efficient. Hypermax and HH often gave identical results, but the addition of weighted varimax improves the otherwise effective HH for complex structures. Geomin, which did best in the constructed example (Table 5), is a hyperplane locating rotation, though not a hyperplane fitting one as defined here. It was slightly less effective in some examples and it has been known to fail with more than two cross-loadings for a given variable, but it did as well as any for the complex box problems. Hypermax, simplimax, geomin, and confirmatory target rotation should be high on the researcher's list of effective tools, especially with complex data. Target rotations included hypercon and Browne's method, but again, hypercon using HDK was more efficient. Other hyperplane and analytic rotations were less effective in some examples.

It should be obvious that no rotation is perfect for every occasion. Conventional analytic rotations are satisfactory for recovering simple structure in many cases. But typically these do not work well with highly complex data, for which Thurstonian simple structure requires identifying those often elusive zeros by locating hyperplane boundaries, thus either eliminating rotational indeterminacy — or at least reducing it—as some data will still require careful scrutiny in order to determine the best solution(s).

Acknowledgements

References

Asparouhov, T. and Muthén, B. (2009). Exploratory structural equation modeling. *Structural Equation Modeling* **16**, 397-438.

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

Bernaards, C. A. and Jennrich, R. I. (2005). Gradient projection algorithms and software for arbitrary rotation criteria in factor analysis. *Educational and Psychological Measurement* **65**, 676-696.

Browne, M. W. (1972). Oblique rotation to a partially specified target. *British Journal of Mathematical and Statistical Psychology* **25**, 207-212.

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. and Mels, G. (2010). *CEFA: Comprehensive Exploratory Factor Analysis, Version 3.04.* [Computer software and manual.] http://quantrm2.psy.ohio-state/edu/browne/

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

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

Derflinger, G. and Kaiser, H. F. (1989). Method Horst-Hilsch (HH) for the transformation problem in exploratory factor analysis. *Psychological Reports* **65**, 211-220.

Fleming, J. S. (1991). *COMPARE: An Interactive Program for Computing Measures of Similarity for Factor Loadings.* [Computer software.] http://swppr.org.

Fleming, J. S. (2003). Computing measures of simplicity of fit for loadings in factor-analytically derived scales. *Behavior Research Method, Instruments, and Computers* **34**, 520-524.

Fleming, J. S. (2011). *Explorer: A Program for Common Factor Analysis and Related Models. (Release 3.2).* [Computer software.] http://swppr.org.

Hakstian, A. R. (1971). A comparative evaluation of several prominent methods of oblique factor transformation. *Psychometrika* **36**, 175-193.

Hakstian, A. R. and Abell, R. A. (1974). A further comparison of oblique factor transformation methods. *Psychometrika* **39**, 429-444.

Harman, H. H. (1976). *Modern Factor Analysis*, 3rd. edition. University of Chicago Press, Chicago.

Harris, C. W. and Kaiser, H. F. (1964). Oblique factor analytic solutions by orthogonal transformations. *Psychometrika* **29**, 347-362.

Hendrickson, A. E. and White, P. O. (1964). Promax: a quick method for rotation to oblique simple structure. *British Journal of Statistical Psychology* **17**, 65-70.

Horst, P. (1941). A non-graphical method for transforming an arbitrary factor matrix into a simple structure factor matrix. *Psychometrika* **6**, 79-99.

Jennrich, R. I. (2006). Rotation to simple loadings using component loss functions: the oblique case. *Psychometrika* **71**, 173-191.

Jennrich, R. I. (2007). Rotation methods, algorithms, and standard errors. In *factor analysis at 100*: *historical developments and future directions* (Edited by R. Cudeck and R. C. MacCallum), 315-335. Erlbaum.

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

Jöreskog, K. G. (1966). Testing a simple structure hypothesis in factor analysis. *Psychometrika* **31**, 165-178.

Kaiser, H. F. (1958). The varimax criterion for analytic rotation in factor analysis. *Psychometrika* **23**, 187-200.

Kaiser, H. F. (1974). An index of factor simplicity. *Psychometrika* **39**, 31-36.

Kaiser, H. F. and Cerny, B. A. (1978). Casey's method for fitting hyperplanes from an intermediate orthomax solution. *Multivariate Behavioral Research* **13**, 395-401.

Kiers, H. A. L. (1994). Simplimax: oblique rotation to an optimal target with simple structure. *Psychometrika* **59**, 567-579.

Korth, B. and Tucker, L. R. (1976). Procrustes matching by congruence coefficients. *Psychometrika* **41**, 531-535.

Lawley, D. N. and Maxwell, A. E. (1964). Factor transformation methods. *British Journal of Statistical Psychology* **17**, 97-103.

Lorenzo-Seva, U. (1999). Promin: a method for oblique rotation. *Multivariate Behavioral Research* **34**, 347-365.

Lorenzo-Seva, U. (2000). The weighted oblimin rotation. *Psychometrika* **65**, 301-318.

Lorenzo-Seva, U. (2003). A factor simplicity index. *Psychometrika* **68**, 49-60.

Lorenzo-Seva, U. and Ferrando, P. J. (2006). Factor: a computer program to fit the exploratory factor analysis model. *Behavior Research Methods, Instruments, and Computers* **38**, 88-91.

Marshall, A. W. and Olkin, I. (1979). *Inequalities: Theory of Majorization and its Applications.* Academic Press, New York.

Mulaik, S. A. (1972). *The Foundations of Factor Analysis.* McGraw-Hill, New York.

Rozeboom, W. (1992). The glory of suboptimal factor rotation: why local minima in analytic optimization of simple structure are more blessing than curse. *Multivariate Behavioral Research* **27**, 585-599.

Sass, D. A. and Schmitt, T. A. (2010). A comparative investigation of rotation criteria within exploratory factor analysis. *Multivariate Behavioral Research* **45**, 73-103.

Spearman, C. (1904). General intelligence, objectively determined and measured. *American Journal of Psychology* **15**, 201-293.

Trendafilov, N. T. (1994). A simple method for procrustean rotation in factor analysis using majorization theory. *Multivariate Behavioral Research* **29**, 385-408.

Thurstone, L. L. (1938). *Primary Mental Abilities.* Psychometric Monographs, No. 1, University of Chicago Press, Chicago.

Thurstone, L. L. (1947). *Multiple Factor Analysis.* University of Chicago Press, Chicago.

Tucker, L. R. (1940). A rotational method based upon the mean principal axis of a subgroup of tests. *Psychological Bulletin* **37**, 578.

Tucker, L. R. (1951). *A Method of Synthesis for Factor Analysis Studies.* Personnel Section Research Report No. 984. Department of the Army, Washington DC.

Wechsler, D. (1974). *Manual for the Wechsler Intelligence Scale for Children − Revised.* Psychological Corporation, New York.

Yates, A. (1987). *Multivariate Exploratory Data Analysis*: *A perspective on Exploratory Factor Analysis.* State University of New York Press, Albany, New York.

James S. Fleming
Southwest Psychometrics and Psychology Resources
2275 W. Pine Drive, Prescott, AZ 86305, USA
jsf.swppr@gmail.com