ROTATION TO SIMPLE LOADINGS USING COMPONENT LOSS FUNCTIONS: THE OBLIQUE CASE

Robert I. Jennrich

University of California at Los Angeles

July 30, 2004

Component loss functions (CLFs) similar to those used in orthogonal rotation are introduced to define criteria for oblique rotation in factor analysis. It is shown how the shape of the CLF effects the performance of the criterion it defines. For example it is shown that monotone concave CLFs give criteria that are minimized by loadings with perfect simple structure when such loadings exist. Moreover, if the CLFs are strictly concave, minimizing must produce perfect simple structure whenever it exists. Examples show that methods defined by concave CLF perform well much more generally. While it appears important to use a concave CLF the specific CLF used is less important. For example the very simple linear CLF gives a rotation method that can easily outperform the most popular oblique rotation methods promax and quartimin and is competitive with the more complex simplimax and geomin methods.

Key words: Component loss criteria, factor analysis, geomin, gradient projection, loading polish, promax, quartimin, simplimax, sorted absolute

loading plots.

Requests for reprints should be sent to Robert I. Jennrich, Department of Mathematics, University of California, Los Angeles, CA 90095.

1 Introduction

The rotation problem in factor analysis arises from a desire to find a simple and contextually meaningful relation between items and factors. At first this search was restricted to the use of orthogonal factors, but frequently better results can be obtained by allowing the factors to be oblique which is the case considered here. Oblique rotation methods that rotate factors to produce simple loading matrices are called direct methods by Harman (1976). Indirect methods based on rotating factors to produce simple reference structures are mostly of historical interest. Only direct methods are considered.

Unfortunately simple loading matrices are not well defined. Thurstone (1935, p.156) has set forth a number of general principals which vaguely stated say a large number of small loadings are what one should attempt to achieve. Actually Thurstone's conditions are precise, but in general unattainable and hence can only be approximated.

At first attempts were made to approximate Turnstone's conditions by visually rotating hyperplanes in two dimensional plots in an effort to maximize the number of items close to the hyperplanes. This number is called a hyperplane count. Eber (1966) attempted to implement this procedure analytically, but the hyperplane count criterion has serious discontinuities that make analytic rotation difficult. A break through came when Katz and Rohlf (1974) replaced the zero-one hyperplane count for each item by a smooth function of its hyperplane distance. They considered a two parameter family of such functions. Rozeboom (1991) introduced a more flexible four parameter family and applied it directly to the loadings rather than to hyperplane distances. He also allowed the possibility that the function be an arbitrary growth function. We begin with this degree of generality and, as Rozeboom did, apply the functions directly to the loadings.

More specifically we consider a class of criteria that are defined by an arbitrary component loss function (CLF) similar to that considered by Jennrich (2004(a)) for orthogonal rotation. This is evaluated at the absolute value of each component λ_{ir} of a loading matrix Λ. The sum of these losses is the value of the corresponding CLF criterion at Λ.

This CLF approach, or what might be called a generalized hyperplane count approach, has been largely overlooked which is unfortunate because a method so historical, natural, and simple needs to be more carefully considered if for no other reason than to understand why it should not be pursued.

A basic question is how the shape of a CLF affects the performance of the corresponding method. A number of theoretical results address this question. For example a loading matrix is said to have perfect simple structure if each row has at most one nonzero element. One might argue that a minimum requirement for any proper rotation method is that it is optimized by perfect simple structure when it exists. It is shown assuming only that a CLF is concave (i.e., curved downward) and nondecreasing is sufficient to guarantee this. Moreover, if the CLF is strictly concave optimization must produce perfect simple structure whenever it exists. Results like these are important because they provide some clear simple guidance for constructing CLF criteria.

Numerical comparisons demonstrate the theoretical results and show how CLF rotation relates to other forms of oblique rotation. Very simple forms of CLF rotation can handily outperform the most popular methods of oblique rotation, promax (Hendrickson and White, 1964) and direct quartimin (Jennrich and Sampson, 1966), and compete well with some of the best methods including geomin (Yates, 1987, p. 46) and simplimax (Kiers, 1994).

The theory and computing in the oblique case is somewhat more difficult than in the orthogonal case, but as mentioned oblique applications are generally of greater interest.

2 Rotation to simple loadings

Let Λ be a factor loading matrix and let $Q(\Lambda)$ be the value of an oblique rotation criterion at Λ. Consider minimizing $Q(\Lambda)$ over all oblique rotations of an initial loading matrix A, that is over all

$$\Lambda = A(T')^{-1}$$

where T is an arbitrary non-singular matrix with columns of unit length. A minimizing Λ is called an oblique rotation of A corresponding to Q. Different criteria Q produce different rotations of A. The factor correlation matrix for oblique rotation is

$$\Phi = T'T$$

To minimize Q over all oblique rotations of A the derivative free gradient projection (GP) algorithm of Jennrich (2004(b)) was used. The only problem specific information required by this algorithm is a formula for the value of Q at an arbitrary loading matrix Λ. A number of criteria considered have local minima. We dealt with this by arbitrarily defining the best rotation produced from 100 random starts of the GP algorithm to be the operational minimizer of the rotation criterion used. Random starts were also used by Kiers (1994) and Browne (2001), but they restricted them to be orthogonal. This is not sufficient because it is possible that random orthogonal starts, even an infinite number, will not produce a global minimizer. Following Roseboom (1991) oblique random starts have been used. By an oblique random start we mean a rotation matrix T whose columns are independently generated and randomly selected from a unit sphere of appropriate dimension.

Unless otherwise noted, in every application, including the simulations, the GP algorithm, which is strictly monotone, converged to a stationary point. GP algorithms for oblique and orthogonal rotation together with examples of their use, may be downloaded from `http://www.address`. For the derivative free case Matlab and R(=S) versions are given.

3 Component loss rotation criteria

Let Λ be a p by k loading matrix with components λ_{ir}. A rotation criterion Q of the form

$$Q(\Lambda) = \sum \sum h(|\lambda_{ir}|)$$

will be called a component loss criterion (CLC). The function h is its defining component loss function (CLF). From this point of view $Q(\Lambda)$ is the total loss for the components of Λ and the rotation problem is to minimize this total loss. At this point h is any real valued function whose domain is all nonnegative values. As noted, Katz and Rohlf (1974) and Rozeboom (1991) considered criteria of this form.

The component loses are functions of $|\lambda_{ir}|$ rather than of λ_{ir}^2 as they were in the orthogonal case (Jennrich, 2004(a)). This choice is motivated by a desire to simplify the statements of theorems below.

The simplest example of a CLF is the linear CLF defined by

$$h(|\lambda|) = |\lambda|$$

Other examples include the cubic CLF

$$h(|\lambda|) = |\lambda|^3$$

the basic concave CLF

$$h(|\lambda|) = 1 - e^{-|\lambda|}$$

and the quadratic right constant CLF

$$h(|\lambda|) = \begin{cases} |\lambda|^2/.3 & |\lambda| \leq .3 \\ 1 & |\lambda| > .3 \end{cases}$$

These are plotted in Figure 1. The basic concave CLF is called a monomolecular growth curve and is a member of the Katz and Rohlf (1974) family of CLFs. The quadratic right constant CLF is designed to encourage loadings less than 0.3 and as will be seen is related to the simplimax criteria. Note that the linear and basic concave CLFs are concave and nondecreasing. The basic concave CLF is also strictly concave. These, as will be shown, are desirable properties. There are many concave CLFs. Calling the strictly concave CLF above the basic concave CLF is an arbitrary choice, but one that will be used extensively.

CLFs that differ only by an additive constant or a positive constant multiplier give CLC that are equivalent because they have the same minimizers.

4 Concavity and perfect simple structure

A loading matrix has perfect simple structure if it has at most one nonzero loading in each row. What constitutes simple structure is widely debated, but it is clear that perfect simple structure is the simplest possible structure. A desirable property of an analytic rotation criterion is that it be optimized by a loading matrix with perfect simple structure when such a loading matrix exists, that is when there is a rotation of A with perfect simple structure. It will be shown that if a CLF is concave and nondecreasing, the corresponding CLC has this property.

Lemma 1: If Λ and $\hat{\Lambda}$ are rotations of A and all elements in the i-th row of $\hat{\Lambda}$ are zero except possibly for $\hat{\lambda}_{is}$, then

$$|\hat{\lambda}_{is}| \leq |\lambda_{i1}| + \cdots + |\lambda_{ik}|$$

Proof: Because $\Lambda = A(T')^{-1}$, $A = \Lambda T'$ and the i-th row a_i of A is given by

$$a_i = \lambda_{i1} t'_1 + \cdots + \lambda_{ik} t'_k$$

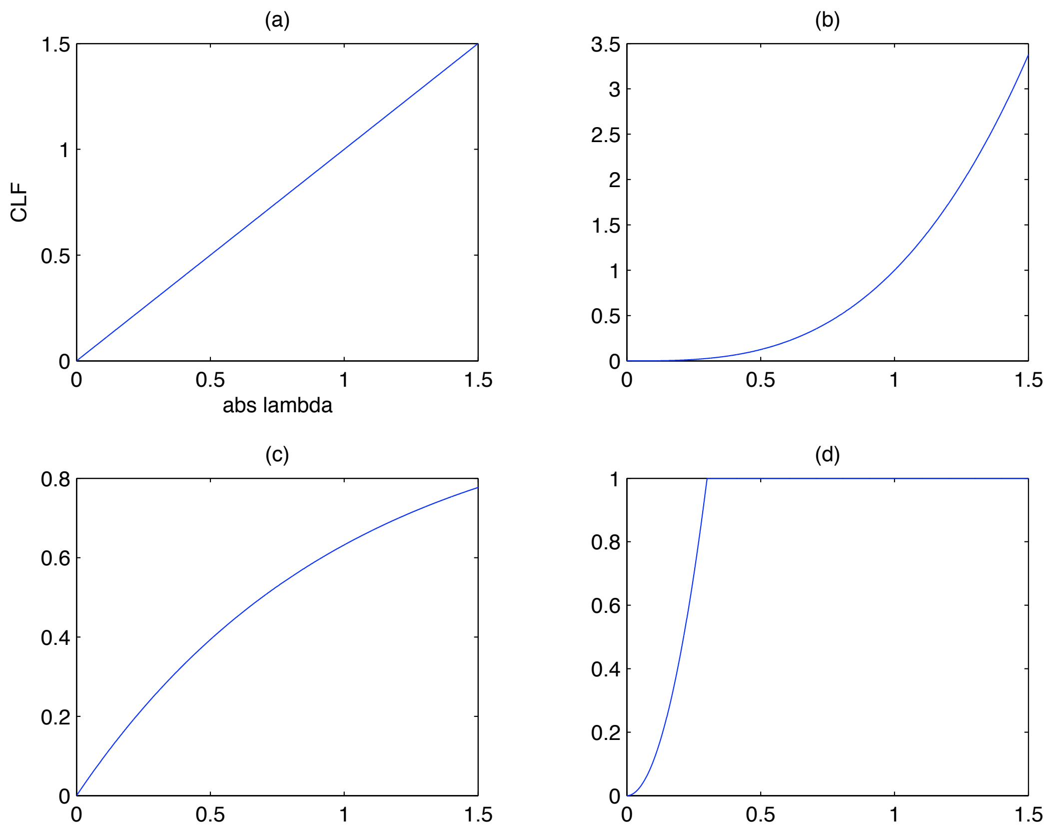

Figure 1: (a) Linear CLF. (b) Cubic CLF. (c) Basic concave CLF. (d) Quadratic right constant CLF.

where t_r is the r-th column of T. Since the t_r have unit length,

$$\|a_i\| \le |\lambda_{i1}| + \cdots + |\lambda_{ik}| \tag{1}$$

Similarly $A = \hat{\Lambda}\hat{T}'$ and because all but the s-th element in the i-th row of $\hat{\Lambda}$ are zero

$$a_i = \hat{\lambda}_{is}\hat{t}'_s$$

Hence $\|a_i\| = |\hat{\lambda}_{is}|$ and the assertion of the lemma follows from this and (1).

Lemma 2: If h is concave, $h(0) = 0$, and $u_1, \cdots, u_k$ are nonnegative, then

$$h(\sum u_r) \le \sum h(u_r)$$

Moreover, if h is strictly concave and at least two u_r are not zero, the inequality is strict.

Proof: Clearly the first assertion holds when all $u_r = 0$. Assume now that at least one $u_r \ne 0$. Then

$$u_r = (1 - \frac{u_r}{\sum u_r})0 + \frac{u_r}{\sum u_r}\sum u_r$$

is a convex combination of 0 and $\sum u_r$. Because h is concave

$$h(u_r) \ge (1 - \frac{u_r}{\sum u_r})h(0) + \frac{u_r}{\sum u_r}h(\sum u_r) = \frac{u_r}{\sum u_r}h(\sum u_r)$$

Assume now that at least two u_r are not zero and h is strictly concave. Then for some r, $u_r/\sum u_r$ is not equal to zero or one and the inequality is strict. Summing on r completes the proof.

Theorem 1: If there is an oblique rotation $\hat{\Lambda}$ of A that has perfect simple structure and if h is a concave and nondecreasing CLF, then $\hat{\Lambda}$ minimizes the corresponding CLC over all oblique rotations of A. Moreover, if h is strictly concave and any minimizer must have perfect simple structure.

Proof: By adding a constant if necessary we may assume without loss of generality that $h(0) = 0$. Then given i, for some s,

$$\sum_r h(|\hat{\lambda}_{ir}|) = h(|\hat{\lambda}_{is}|)$$

Let Λ be any rotation of A. Using Lemma 1 and the monotonicity of h

$$h(|\hat{\lambda}_{is}|) \leq h(\sum_r |\lambda_{ir}|)$$

Because h is concave and $h(0) = 0$ it follows from Lemma 2, that

$$h(\sum_r |\lambda_{ir}|) \leq \sum_r h(|\lambda_{ir}|)$$

Using the first equality and the last two inequalities

$$\sum_r h(|\hat{\lambda}_{ir}|) \leq \sum_r h(|\lambda_{ir}|)$$

Summing on i proves the first assertion.

Assume now that h is strictly concave and Λ does not have perfect simple structure. Then some row i, Λ must have at least two nonzero values and from Lemma 2 for this i the second inequality above must be strict. As a consequence Λ cannot minimize the corresponding CLC. This proves the second assertion.

The second assertion of Theorem 1 says that when the CLF is strictly concave CLF rotation must produce perfect simple structure whenever it exists.

As we will see shortly the concavity assumption in Theorem 1 is necessary. The nondecreasing assumption is also necessary. Consider, for example, the CLF, $h(|\lambda|) = -|\lambda|^3$ which is concave but decreasing rather than nondecreasing. The corresponding CLC approaches minus infinity as the factor loading matrix approaches infinity which happens as the rotation matrix T becomes singular. Thus any perfect simple structure rotation of A, if it exists, will fail to minimize this CLC. This shows that the conclusion of Theorem 1 can fail when the nondecreasing CLF assumption is not satisfied.

Theorem 1 identifies some CLC that are optimized by loading matrices with perfect simple structure. Other criteria also have this property. Carroll's

(1953) quartimin criterion which is not a CLC can be expressed in the form

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

If Λ has perfect simple structure, $Q(\Lambda) = 0$ which is the smallest value $Q(\Lambda)$ can have. Thus if Λ is also a rotation of A, it minimizes the quartimin criterion over all rotations of A.

Before turning to some perfect simple structure examples we need to deal with an algorithm problem that arises with the use of concave nondecreasing CLFs. Note that if h is a concave nondecreasing CLF the only way $h(|\lambda|)$ can be differentiable function of λ at $\lambda = 0$ is if h is constant which is a very uninteresting case. Thus if h is not constant the corresponding CLC is not differentiable at Λ if Λ has a zero component.

This raises two problems when attempting to minimize a CLC defined by a concave nondecreasing CLF if the minimizing Λ has one or more zero components. Minimization becomes numerically difficult and especially so with algorithms that use gradients. Moreover, stationarity is not defined making it difficult to monitor convergence and be assured the algorithm has at least reached a stationary point.

One way to deal with these problems is to round corners. Figure 2(a) is a plot of $h(|\lambda|)$ on λ when h is a linear CLF. To round the corner at $\lambda = 0$ one can approximate $h(|\lambda|)$ on a small interval $-\epsilon \leq \lambda \leq \epsilon$ by a parabola. Moreover, this may be done so the values and derivatives of the parabola and $h(|\lambda|)$ agree at $\lambda = \pm\epsilon$. Such an approximation is shown in Figure 2(b) using $\epsilon = .1$ and Figure 2(c) using $\epsilon = .01$. In the latter case the approximation is almost invisible. For any $\epsilon > 0$, however, the approximation makes the corresponding CLC differentiable at values of Λ having one or more zero components. We call this modification an epsilon modification. The details for this in the general case are given in the Appendix. For our

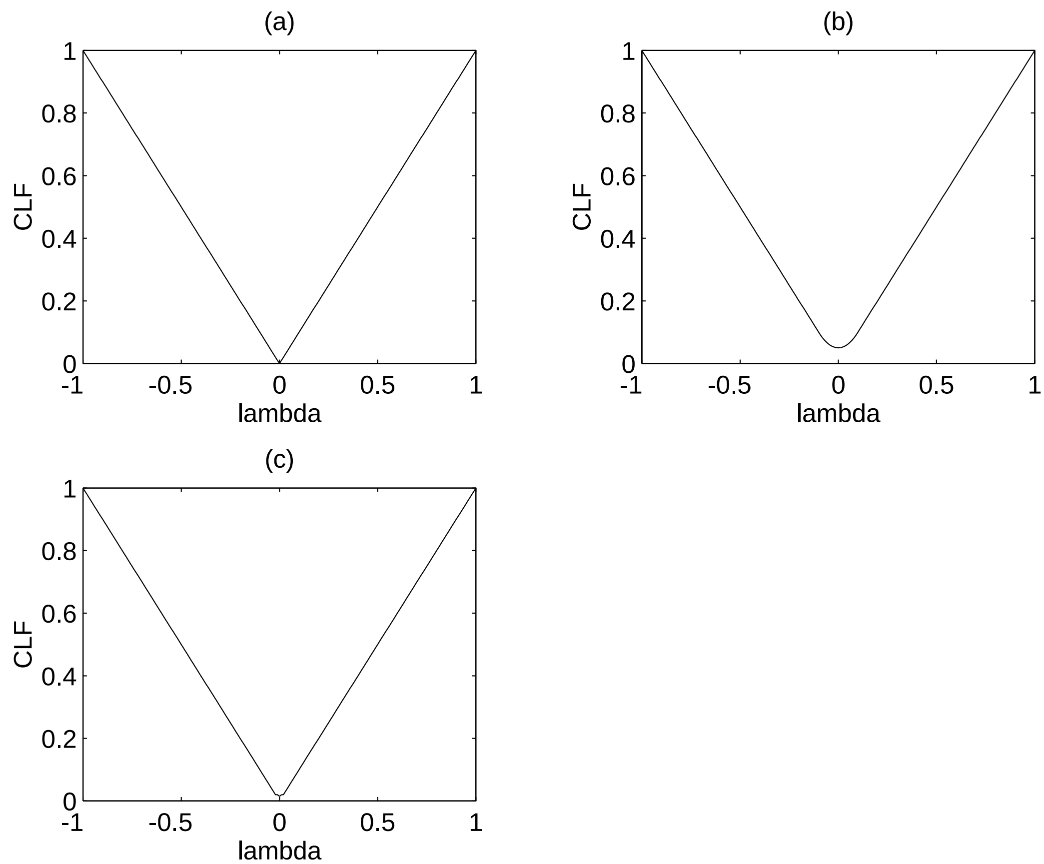

Figure 2: A plot of $h(|\lambda|)$ on λ when h is a linear CLF (a), a modified linear CLF with $\epsilon = .1$ (b), and a modified linear CLF with $\epsilon = .01$ (c).

applications an epsilon modification with $\epsilon = .01$ was used whenever required for computational purposes.

The matrix A in Table 1 has perfect simple structure. Because the linear CLF is concave and non-decreasing, it follows from Theorem 1 that A minimizes the corresponding CLC. The second matrix in Table 1 is the computed rotation of A using a linear CLF. It agrees exactly with A to the precision displayed. The third matrix in Table 1 was computed using the cubic CLF. This CLF is strictly convex rather than concave. The poor result shows how badly the conclusion of Theorem 1 can fail when the concavity assumption does not hold.

Table 1: A perfect simple structure loading matrix A, a CLF rotation of A using a linear CLF, and a CLF rotation of A using a cubic CLF.

A			linear			cubic		
1	0	0	1.00	.00	.00	.60	.60	.37
1	0	0	1.00	.00	.00	.60	.60	.37
1	0	0	1.00	.00	.00	.60	.60	.37
0	1	0	.00	1.00	.00	-.69	.31	.63
0	1	0	.00	1.00	.00	-.69	.31	.63
0	0	1	.00	.00	1.00	-.44	.77	-.69

The second matrix in Table 1 is also the computed quartimin rotation of A. As noted this is expected for quartimin rotation because A has perfect simple structure.

5 Thurstone type simple structure

As precisely defined a loading matrix with Thurstone (1935, p156) simple structure has a fair number of exact zeros. We will say a loading matrix has simple structure of Thurstone type if it has one or more exact zeros. The matrix A in Table 2 does not have perfect simple structure, but it does have Thurstone simple structure and simple structure of the Thurstone type.

The second matrix in Table 2 is a rotation of A using a linear CLF. To the precision displayed it very nearly re-produces the Thurstone simple structure of A. The third matrix in Table 2 is a quartimin rotation of A. It produces a much cruder approximation to the Thurstone simple structure of A. The fourth matrix in Table 2 is a rotation of A using the basic concave CLF. Like the linear CLF it very nearly re-produces the Thurstone simple structure of

Table 2: Linear CLF, quartimin, and basic concave CLF rotations of A.

A			linear			quartimin			basic concave		
1	0	0	1.00	-.01	-.01	1.05	-.10	-.10	1.00	-.00	-.00
0	1	0	-.01	1.00	-.01	-.01	1.02	-.15	-.00	1.00	-.00
0	0	1	-.01	-.01	1.00	-.01	-.15	1.02	-.00	-.00	1.00
.89	.45	0	.89	.44	-.01	.89	.37	-.15	.89	.45	-.01
.89	0	.45	.89	-.01	.44	.89	-.15	.37	.89	-.01	.45
0	.71	.71	-.01	.70	.70	-.14	.62	.62	-.01	.71	.71

A. Note that although this is not a perfect simple structure example the two concave CLF methods motivated by Theorem 1 work very well.

The following theorem shows that A is at least a local minimum of a CLC defined by an appropriate CLF.

Theorem 2 Let $\hat{\Lambda}$ be a rotation of A that has Thurstone type simple structure and let $c > 0$ be less than the smallest nonzero absolute loading in $\hat{\Lambda}$. If $h(u)$ is constant for all $u \geq c$ and $h(0) \leq h(u)$ for all $u \geq 0$, then $\hat{\Lambda}$ is a local minimizer of the CLC defined by h.

Proof: We may assume without loss of generality that $h(0) = 0$. Note that $|\hat{\lambda}_{ir}| > 0$ if and only if $|\hat{\lambda}_{ir}| > c$. Let Λ be another rotation of A. If Λ is sufficiently close to $\hat{\Lambda}$, $|\lambda_{ir}| > c$ if and only if $|\hat{\lambda}_{ir}| > c$. Thus for Λ sufficiently close to $\hat{\Lambda}$

$$\sum_i \sum_r h(|\hat{\lambda}_{ir}|) = \sum\sum_{|\hat{\lambda}_{ir}|>c} h(|\hat{\lambda}_{ir}|) = \sum\sum_{|\lambda_{ir}|>c} h(|\lambda_{ir}|) \leq \sum_i \sum_r h(|\lambda_{ir}|)$$

This completes the proof.

Theorem 2 applies to almost any CLF that becomes constant sufficiently soon. The Theorem is weak in the sense that it guarantees only a local rather than a global minimum. The local minimum may of course also be global.

Note that .3 is less than the smallest nonzero absolute loading in A of Table 2. Thus the quadratic right constant CLF in Figure 1 satisfies the assumptions of Theorem 2. The theorem implies that A is a local minimum of the corresponding CLC. Using this CLC, the GP algorithm produced A to 6.61 decimal places. This provides empirical support for the theorem.

6 The more general case

Perfect simple structure and Thurstone simple structure don't occur in practice. They are at best idealizations. Unfortunately, there is no generally accepted broadly applicable definition of simple structure. It is generally felt, however, that a loading matrix with a fair number of small values is simpler than one with mostly intermediate values. Motivated by this, we consider methods designed to produce as many small loadings as possible.

Because many rotation criteria, including all mentioned thus far, are influenced more by large rows of A than by small rows, and this generally does not seem desirable, one is motivated to normalize the rows of A before rotation begins. This is called Kaiser normalization (Kaiser, 1958). Such a modification makes the resulting method invariant with respect to row scaling. While we will not demand this be done, in order to avoid the normalization issue, in our examples we use initial loading matrices A with normalized or nearly normalized rows.

Note, however, if a rotation of an initial loading matrix A has perfect simple structure this is also true for a row scaled form of A. Thus if the hypotheses of Theorems 1 and 2 hold for an initial loading matrix A they also hold for a row normalized form of A. As far as these theorems are concerned one can use A or a row normalized form of A. This should not, however, motivate one to ignore row scaling in practice.

6.1 Some other rotation methods

Because of their wide availability in statistical software promax and quartimin are by far the most popular methods of oblique rotation. Two newer attractive methods are geomin and simplimax. These methods are used for empirical comparison with CLF methods. There definitions follow.

Promax may be viewed as orthogonal rotation with oblique polish. The version considered here begins with an orthogonal varimax rotation of an initial loading matrix A. The components of the result are raised to the third power and the resulting matrix is used as a target for an oblique procrustes rotation of A. The result of this second rotation is the promax rotation of A.

Quartimin is defined by the criterion given in (1).

Geomin is defined by the criterion

$$Q(\Lambda) = \sum_i (\prod_r \lambda_{ir}^2)^{1/k}$$

Like the component loss criterion for a concave CLF, this criterion is not a differentiable function of a loading λ_{ir} when $\lambda_{ir} = 0$. In practice this problem is resolved by replacing λ_{ir}^2 by $\lambda_{ir}^2 + \epsilon$ for some ϵ. Like Browne (2001) we use $\epsilon = .01$.

Simplimax is defined by the criterion

$$K_m(\Lambda) = \sum_i \sum_r [\lambda_{ir}^2 \leq \lambda_{(m)}^2] \lambda_{ir}^2$$

where $[\cdot]$ is one when its argument is true and zero otherwise and $\lambda_{(m)}^2$ is the m-th smallest value of the λ_{ir}^2. The simplimax criterion is minimized when the sum of squares of the m smallest loadings is as small as possible. The parameter m may be viewed as the number of target zeros for the criterion.

Although at first it appears to be, the simplimax criterion is not a CLC because

$$[\lambda_{ir}^2 \leq \lambda_{(m)}^2]$$

is a function of all the components of Λ and not just of λ_{ir}. The standard way to minimize K_m, however, is to compute $\lambda_{(m)}$ from the current value of Λ and holding this fixed minimize, or at least reduce, the value of the CLC with CLF

$$h(|\lambda|) = [\lambda^2 \leq \lambda^2_{(m)}]\lambda^2$$

Thus K_m might be called an iteratively re-defined CLC criterion.

As we will see next, although they are different criteria, simplimax and CLC using quadratic right constant CLFs can and often do produce the same loadings.

Let Q_b be the CLC defined by the quadratic right constant CLF

$$h(|\lambda|) = \begin{cases} (\lambda/b)^2 & |\lambda| \leq b \\ 1 & |\lambda| > b \end{cases}$$

When $b = .3$ this is the quadratic right constant CLF in Figure 1.

The following theorem may be found in Jennrich (2004(a)).

Theorem 3: If for a loading matrix $\hat{\Lambda}$, b is strictly between $\hat{\lambda}^2_{(m)}$ and $\hat{\lambda}^2_{(m+1)}$, then $\hat{\Lambda}$ is a local minimum of K_m if and only if it is a local minimum of Q_b.

Theorem 3 says that the local minima of K_m for various m can be found among the local minima of Q_b for various b and conversely. In this sense simplimax and quadratic right constant CLF methods are equivalent even though the rotation criteria used are not.

7 Comparisons using some familiar data

To continue it will be helpful to at look at some familiar more realistic problems then the simple examples considered thus far. These will be used to compare the linear and basic concave CLF methods with the promax, quartimin, geomin, and simplimax methods identified in the previous section.

7.1 Twenty four psychological tests

Harman (1976, Table 10.9) gave an initial loading matrix A obtained from a maximum likelihood extraction of four factors from a subset of psychological tests data collected by Holzinger and Swineford (1939). This matrix has been used extensively by many authors to demonstrate and compare rotation methods. Because the use of simplimax differs somewhat from that of the other four methods, it will be discussed in a separate subsection.

7.1.1 Promax, quartimin, geomin, and basic concave CLF

For the promax, quartimin, and geomin methods the identity start and all 100 random starts gave the same criterion value. Apparently for these data and methods random starts are not required. That was almost the case for the basic concave CLF method. It gave the minimum criterion value using the identity and 96 of the 100 random starts.

The rotations produced are quite similar. This is suggested by the similarity of the sorted absolute loading plots (Jennrich, 2004(a)) in Figure 3. Table 3 displays the promax and basic concave CLF rotations. The simplimax rotation in Table 3 will be discussed later. In the table large values are boldfaced. A value in a row is large if its magnitude (absolute value) is greater than half the magnitude of the largest value in the row. In Table 3 the large values for the promax and basic concave CLF rotations are, except for two, all in the same positions. Actually the promax, basic KR, and quartimin, and qeomin rotations differ in at most two positions. Thus all four of these methods give quite similar results.

7.1.2 Simplimax

Using simplimax is more complicated than using the methods of the previous subsection. One must specify the number of target zeros m and the number

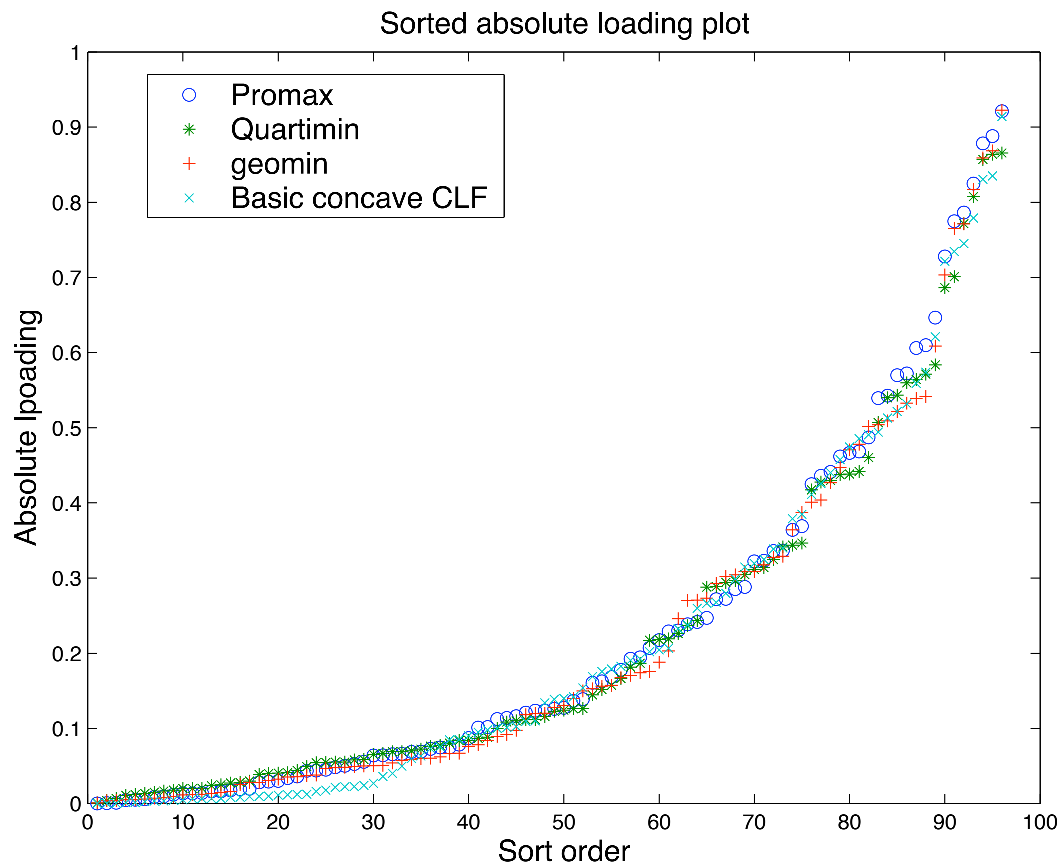

Figure 3: SAL plots for the promax, quartimin, geomin, and basic concave CLF rotations of the 24 psychological tests data.

Table 3: Promax, basic concave CLF, and simplimax $m = 60$ rotations of the 24 psychological tests data.

Promax				Basic concave CLF				Simplimax m=60			
.77	-.03	-.02	.00	**.72**	-.01	.06	.01	**.75**	-.01	.03	-.02
.49	.00	-.05	.00	**.46**	.02	.00	.00	**.47**	.02	-.02	-.02
.65	.01	-.21	.01	**.62**	.04	-.14	.00	**.60**	.02	-.15	-.02
.57	.12	-.07	-.06	**.53**	.14	.00	-.05	**.54**	.14	-.02	-.08
-.01	**.79**	.11	-.04	-.01	**.73**	.17	.00	.02	**.78**	.10	-.05
-.02	**.82**	-.07	.08	.02	**.78**	-.01	.10	.01	**.82**	-.07	.05
.00	**.89**	.05	-.14	.00	**.83**	.11	-.08	.01	**.88**	.05	-.14
.23	**.54**	.12	-.07	.21	**.51**	.18	-.02	.25	**.55**	.12	-.07
-.03	**.88**	-.10	.07	.01	**.83**	-.04	.09	-.01	**.87**	-.10	.04
-.24	.06	**.92**	.01	-.30	.00	**.91**	.10	-.09	.10	**.83**	.07
.01	.04	**.47**	**.29**	.01	.01	**.49**	**.31**	.12	.07	**.43**	**.29**
.23	-.16	**.73**	-.09	.13	-.18	**.74**	-.01	.32	-.12	**.68**	-.04
.47	.02	**.44**	-.13	**.38**	.01	**.49**	-.06	**.51**	.05	**.44**	-.11
-.13	.11	-.03	**.61**	-.02	.10	.00	**.56**	-.04	.14	-.03	**.55**
-.01	.01	-.04	**.57**	.07	.01	-.01	**.52**	.06	.04	-.04	**.52**
.36	-.12	-.14	**.54**	**.43**	-.10	-.08	**.49**	**.41**	-.09	-.10	**.48**
-.10	.02	.12	**.61**	-.02	.00	.15	**.57**	.01	.05	.11	**.57**
.25	-.18	**.22**	**.42**	**.27**	-.18	**.26**	**.41**	**.33**	-.14	**.22**	**.40**
.17	.03	.04	**.34**	**.20**	.03	.09	**.32**	**.22**	.05	.05	**.31**
.32	**.29**	-.06	**.19**	**.34**	**.28**	.01	**.19**	**.34**	**.30**	-.03	.16
.37	.01	**.34**	.07	**.32**	.00	**.39**	.11	**.42**	.04	**.33**	.08
.32	**.27**	-.05	**.19**	**.34**	**.27**	.02	**.19**	**.34**	**.29**	-.02	.16
.46	**.24**	.07	.07	**.44**	**.24**	.14	.10	**.48**	**.26**	.09	.06
.02	**.27**	**.44**	.16	.01	.23	**.47**	.20	.11	**.30**	**.40**	.17

of random starts to be used. Kiers (1994) suggested making a scree plot that displays the operationally optimal values of the simplimax criterion on a range of values of m. The hope is that the scree plot will show a clear jump immediately following some value of m. The recommendation then is to use this value of m for simplimax rotation. Unfortunately there is not always a clear jump.

Figure 4 shows a scree plot for the 26 psychological tests data. The range of m is from 46 to 74 in steps of 4. The plot is very smooth with no sign of a jump. Clearly some method other than the scree plot must be used to choose m. In the plot the lowest value of $m = 46$ corresponds to a simplimax target with average row complexity two and the largest value $m = 74$ corresponds to a target with average row complexity one.

One alternative to the scree plot is to look directly at the loading matrices produced by a selection of values of m. Table 4 displays the simplimax loadings for $m = 50, 60, 70$. The value $m = 60$ corresponds to an average target row complexity of 1.5 which is the average row complexity for the rotations in the previous subsection. The $m = 50$ rotation in Table 4 is unsatisfactory because of the many large loadings in the first column. The $m = 60$ and $m = 70$ rotations are quite similar, the author at least would have difficulty choosing between them.

Table 3 displays the the simplimax $m = 60$ rotation together with the promax and basic concave CLF rotations. These are all quite similar. This together with observations from the previous subsection show that promax, quartimin, geomin, basic concave CLF, and simplimax with $m = 60$ give very similar rotations for the 24 psychological tests problem and in particular that the CLF method worked as well as the alternative methods considered.

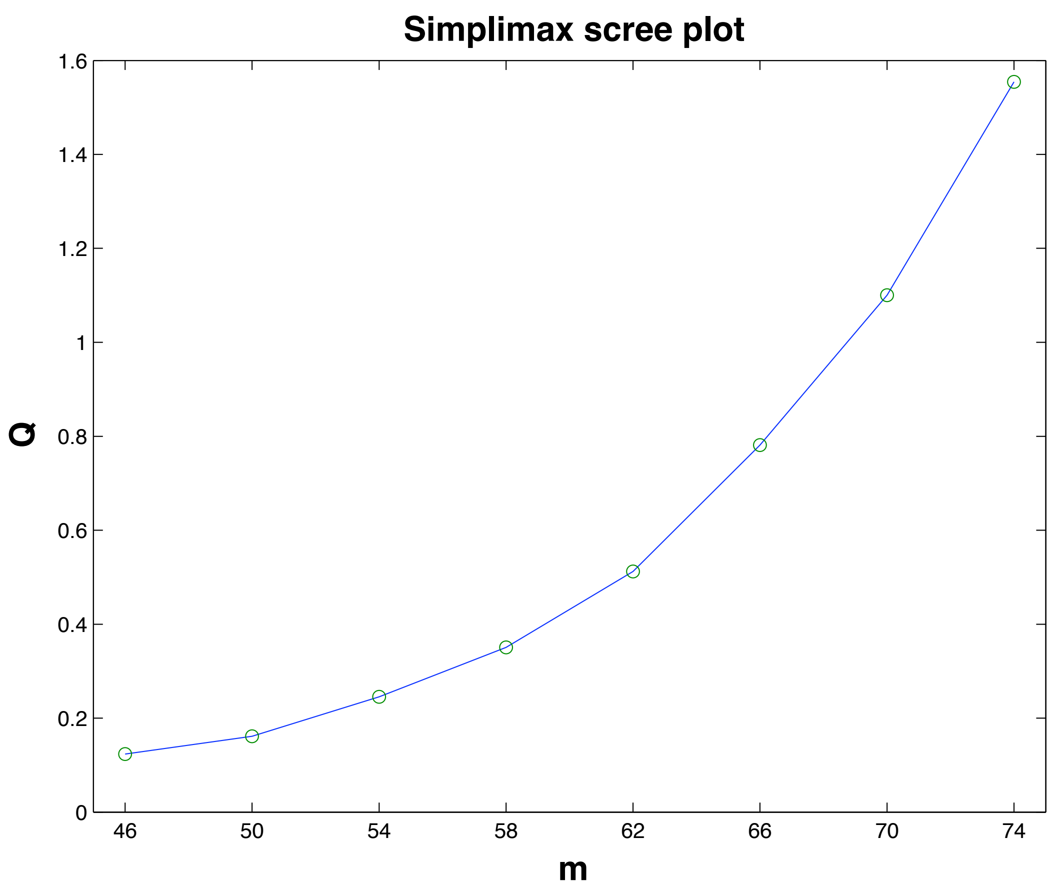

Figure 4: Simplimax scree plot for the 24 psychological tests data

Table 4: Simplimax rotations of the 24 psychological tests data using $m = 50, 60, 70$

m=50				m=60				m=70			
.24	**.47**	-.05	**.32**	**.75**	-.01	.03	-.02	**.72**	-.02	.05	.06
.13	**.33**	-.01	**.17**	**.47**	.02	-.02	-.02	**.45**	.01	.00	.03
.03	**.54**	-.02	.18	**.60**	.02	-.15	-.02	**.57**	.02	-.14	.05
.18	**.38**	.09	.18	**.54**	.14	-.02	-.08	**.53**	.13	-.01	-.03
.68	-.12	**.58**	-.26	.02	**.78**	.10	-.05	.02	**.77**	.12	-.03
.63	.02	**.56**	**-.39**	.01	**.82**	-.07	.05	-.03	**.81**	-.05	.09
.61	-.07	**.68**	-.29	.01	**.88**	.05	-.14	.03	**.87**	.07	-.13
.57	.02	**.40**	-.06	.25	**.55**	.12	-.07	.24	**.53**	.14	-.04
.62	.03	**.61**	**-.42**	-.01	**.87**	-.10	.04	-.04	**.86**	-.08	.08
.98	**-.86**	.07	.21	-.09	.10	**.83**	.07	-.09	.05	**.85**	.04
.82	-.37	-.05	.06	.12	.07	**.43**	**.29**	.03	.02	**.45**	**.33**
.67	**-.43**	-.08	**.45**	.32	-.12	**.68**	-.04	.33	-.16	**.71**	-.05
.55	-.07	.05	**.38**	**.51**	.05	**.44**	-.11	**.52**	.02	**.46**	-.08
.59	-.06	-.10	**-.32**	-.04	.14	-.03	**.55**	-.20	.09	-.03	**.64**
.50	.02	-.17	-.23	.06	.04	-.04	**.52**	-.10	.00	-.04	**.61**
.41	**.32**	**-.27**	-.04	**.41**	-.09	-.10	**.48**	.25	-.13	-.09	**.60**
.69	-.16	-.17	-.22	.01	.05	.11	**.57**	-.16	.00	.11	**.65**
.60	-.02	-.26	.10	**.33**	-.14	**.22**	**.40**	.20	-.19	.23	**.48**
.45	.06	-.09	-.05	**.22**	.05	.05	**.31**	.12	.02	.06	**.38**
.46	.22	.13	-.07	**.34**	**.30**	-.03	.16	**.27**	**.28**	-.01	**.23**
.58	-.05	-.02	.24	**.42**	.04	**.33**	.08	**.38**	.00	**.35**	.12
.46	.21	.12	-.07	**.34**	**.29**	-.02	.16	**.27**	**.26**	-.01	**.24**
.51	.21	.14	.09	**.48**	**.26**	.09	.06	**.43**	**.24**	.12	.12
.85	-.34	.15	.00	.11	**.30**	**.40**	.17	.06	**.26**	**.42**	.20

7.1.3 Computation

Minimizing the simplimax criterion was considerably more difficult than minimizing the criteria in Section 7.1.1. Rather than producing the same criterion value from all or almost all 100 random starts, many different values were produced. Table 5 gives convergence results for the 9 values of m that were considered. The minimum count value is the number of random starts that produced the minimum observed value of the criterion. A minimum count value of 1 is disturbing because it suggests that more starts may produce a smaller minimum criterion value. The failure count value is the number of random starts from which the rotation algorithm failed to converge within 1000 iterations. Only $m = 48$ and $m = 50$ produced algorithm failures. In contrast to simplimax, the algorithms of the Section 7.1.1 all had minimum counts of at least 96 and failure counts of zero.

7.2 Thurstone's box problem

Starting with a collection of 20 boxes Thurstone (1947) constructed 26 variables defined by simple linear and nonlinear functions of the box dimensions x, y and z. The definitions of these variables are given in the first column of Table 6. A three factor initial loading matrix A was extracted from data generated using these variables. For this we used the A given by Cureton and Mulaik (1975). Like the 24 psychological tests data in Harman this matrix has been used extensively to demonstrate and compare rotation methods.

Table 6 contains promax, quartimin and geomin rotations of A. The promax and quartimin rotations in Table 6 can be considered failures (Kiers, 1994; Browne, 2001) because the factors do not appear to have a clear relation to the box dimensions x, y and z used in the formulas to generate the variables. In this regard quartimin does somewhat better than promax.

Table 5: Simplimax random start summary for the 24 psychological tests data. "m" is the simplimax parameter. "Qmin" is the minimum criterion value out of 100 random starts. "min count" is the number of random starts that had the minimum criterion value. "failure count" is the number is the number times the algorithm failed to converge within 1000 iterations.

m	Qmin	min count	failure count
46	0.1236	1	1
50	0.1613	1	0
54	0.2456	1	0
58	0.3508	2	0
60	0.4189	14	0
62	0.5120	23	0
66	0.7814	2	0
70	1.1001	7	0
74	1.5543	20	0

For quartimin in every row but two, the smallest absolute loadings can be associated with the missing dimensions in the formula for the corresponding variable. This is far from the case for promax. A second reason for calling these rotations failures is that one can do much better. For example for the geomin rotation in Table 6 the factors are clearly related in an appropriate way to the box dimensions x, y and z.

Table 7 contains the basic concave CLF, simplimax m=27, and linear CLF rotations for the box problem. Like the geomin rotation these rotations are far superior to the promax and quartimin rotations in Table 6. In each case the factors are clearly related in an appropriate way to the box dimensions x, y and z. Indeed the geomin, basic concave CLF, simplimax m=27, and linear CLF rotations are very similar.

The parameter value $m = 27$ for simplimax rotation was used by Kiers who obtained it from a scree plot that displayed a large jump following $m = 27$. It is also the value suggested by the formulas in the first column of Table 7, but one cannot expect to have this form of information in general practice.

While the geomin, basic concave CLF, linear CLF, and simplimax rotations are very similar in this example there are reasons for preferring the first three of these to simplimax. One is that when using simplimax one must choose a value for the parameter m and this requires a scree plot or some other method. Another is the local minimum problem. This is displayed in Table 8. Simplimax had 80 local minima. That means that 80 distinct minima were generated from the 100 random starts. The other methods considered had at most 16. For simplimax only 2 of the 100 random starts produced the operational minimum while for the other methods at least 23 did this. One feels it is quite likely these methods have reached a global minimum. While this is probably true for simplimax as well, without the

Table 6: Promax, quartimin, and geomin rotations for Thurstone's box data.

formula	promax			quartimin			geomin		
x	.60	-.24	.73	1.33	.35	.61	.99	-.02	-.01
y	.67	.65	-.05	.68	1.36	.73	.06	.94	.05
z	.85	-.36	-.44	.35	.41	.80	-.00	.06	.97
xy	.79	.33	.42	1.25	1.14	.84	.64	.64	-.01
xz	.92	-.41	.14	1.02	.45	.88	.60	.00	.65
yz	.91	.18	-.36	.55	1.07	.92	-.02	.61	.64
x^2y	.77	.08	.58	1.37	.86	.80	.84	.38	.01
xy^2	.78	.49	.20	1.03	1.30	.84	.39	.81	.03
x^2z	.83	-.37	.39	1.19	.41	.81	.79	-.02	.42
xz^2	1.02	-.44	-.07	.89	.50	.97	.44	.03	.86
y^2z	.85	.37	-.28	.59	1.22	.88	-.02	.77	.45
yz^2	.92	-.00	-.41	.49	.87	.90	-.03	.44	.78
x/y	-.04	-.78	.60	.50	-.88	-.10	.75	-.83	.01
y/x	.04	.78	-.60	-.50	.88	.10	-.75	.83	-.01
x/z	-.18	.16	.97	.84	.04	-.13	.82	.01	-.83
z/x	.18	-.16	-.97	-.84	-.04	.13	-.82	-.01	.83
y/z	-.15	.96	.29	.25	.92	-.04	-.01	.85	-.80
z/y	.15	-.96	-.29	-.25	-.92	.04	.01	-.85	.80
$2x + 2y$	.77	.40	.36	1.17	1.20	.83	.55	.71	-.02
$2x + 2z$	.92	-.44	.10	.96	.41	.87	.56	-.02	.69
$2y + 2z$	.92	.19	-.34	.58	1.08	.92	-.01	.62	.63
$(x^2 + y^2)^{1/2}$	.77	.39	.34	1.16	1.18	.82	.54	.70	-.01
$(x^2 + z^2)^{1/2}$	.90	-.42	.08	.92	.42	.86	.53	-.01	.68
$(y^2 + z^2)^{1/2}$	.90	.20	-.31	.60	1.08	.91	.02	.62	.60
xyz	.98	.05	.08	1.06	1.00	.98	.45	.48	.47
$(x^2 + y^2 + z^2)^{1/2}$	.95	.10	-.01	.94	1.03	.96	.34	.53	.49

Table 7: Basic concave CLF, simplimax, and linear CLF rotations of Thurstone's box data.

formula	concave CLF			simplimax			linear CLF		
x	.99	-.01	-.01	.99	-.01	-.01	.99	-.01	.00
y	.07	.94	.05	.06	.94	.04	.07	.94	.06
z	.00	.07	.96	.01	.05	.97	.01	.07	.96
xy	.64	.64	-.01	.63	.65	-.02	.64	.64	.00
xz	.60	.01	.65	.60	.00	.65	.60	.01	.65
yz	-.02	.61	.65	-.02	.61	.64	-.01	.61	.65
x^2y	.84	.39	.01	.83	.39	.01	.84	.39	.01
xy^2	.39	.81	.04	.38	.82	.03	.39	.81	.04
x^2z	.79	-.01	.42	.79	-.02	.42	.79	-.01	.42
xz^2	.45	.03	.86	.45	.02	.86	.45	.04	.86
y^2z	-.01	.77	.46	-.02	.76	.45	-.01	.77	.46
yz^2	-.02	.44	.79	-.02	.43	.78	-.02	.44	.79
x/y	.74	-.82	.01	.75	-.83	.02	.74	-.82	.00
y/x	-.74	.82	-.01	-.75	.83	-.02	-.74	.82	.00
x/z	.81	.01	-.82	.81	.02	-.82	.81	.01	-.82
z/x	-.81	-.01	.82	-.81	-.02	.82	-.81	-.01	.82
y/z	-.00	.84	-.79	-.02	.86	-.80	.00	.84	-.79
z/y	.00	-.84	.79	.02	-.86	.80	.00	-.84	.79
$2x+2y$	.55	.71	-.01	.55	.72	-.02	.55	.71	-.01
$2x+2z$	.56	-.02	.69	.56	-.03	.69	.56	-.02	.69
$2y+2z$	.00	.62	.63	-.00	.61	.63	.00	.62	.63
$(x^2+y^2)^{1/2}$	.54	.70	-.00	.54	.71	-.01	.54	.70	.00
$(x^2+z^2)^{1/2}$	.53	-.00	.68	.53	-.01	.68	.53	.00	.68
$(y^2+z^2)^{1/2}$	.02	.62	.60	.02	.61	.60	.03	.62	.60
xyz	.46	.48	.47	.45	.48	.46	.46	.48	.47
$(x^2+y^2+z^2)^{1/2}$	.35	.53	.49	.34	.53	.48	.35	.53	.49

Table 8: Random start summary for the box data. "min count" is the number of random starts out of 100 that had the minimum criterion value. "local minima" is the number of distinct local minima generated. "failure count" is the number of times the algorithm failed to converge in 1000 iterations. "identity start" is the ability "Y" of the identity start to produce the operational minimum.

Method	min count	local minima	failure count	identity start
promax	100	1	0	Y
quartimin	100	1	0	Y
qeomin	30	4	0	N
basic concave CLF	23	13	0	N
linear CLF	24	16	0	N
simplimax	2	80	3	N

results from the other methods or the formulas in Table 6, facing 80 local minima and an operational minimum count of 2 would not be too reassuring. One can always take more random starts, but how many?

One can use CLF rotation to produce starting values for other forms of rotation or equivalently use other forms of rotation to polish CLF rotation. Figure 5 is a sorted absolute loading plot for the basic concave CLF rotation of the box data. Because of the large gap between the 27th and 28th absolute loadings, the plot suggests using $m = 27$ for simplimax rotation. This provides an alternative to the scree plot for choosing m. Using this value of m and the basic concave CLF rotation matrix T as a starting value for the simplimax algorithm gives the simplimax rotation in Table 7 without additional random starts. Thus the simplimax rotation in Table 7 is also a

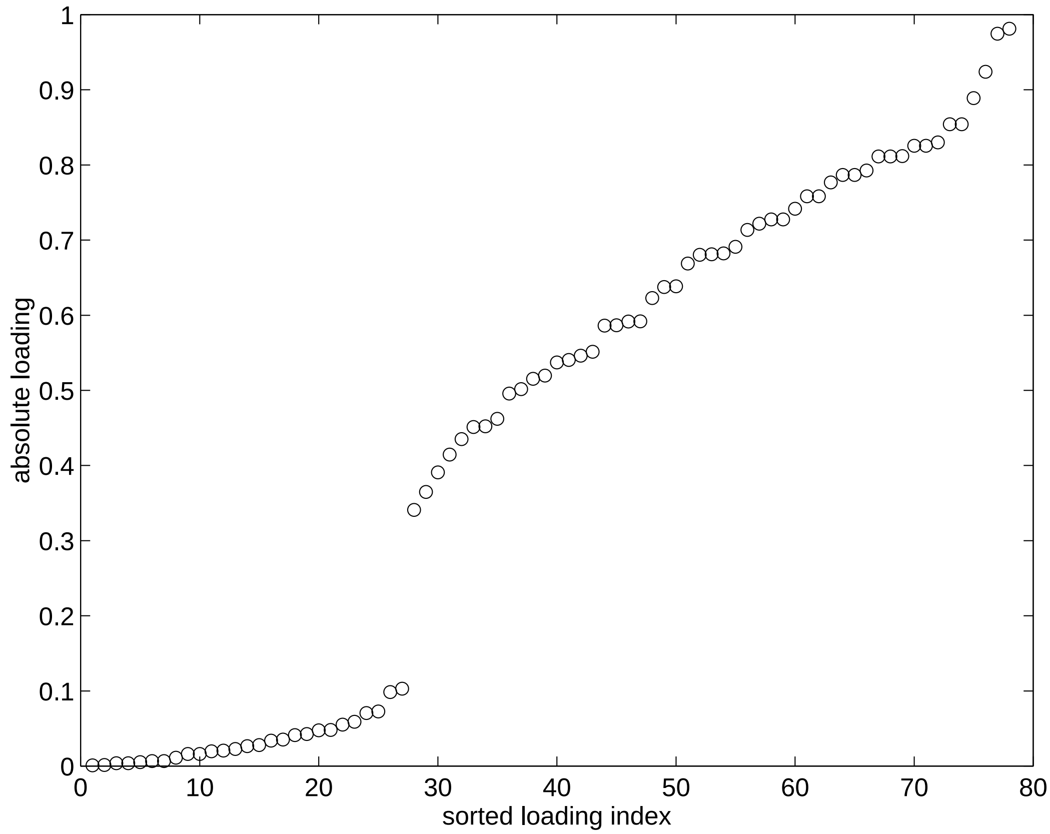

Figure 5: Sorted absolute loading plot for the basic concave CLF rotation of the box data.

basic concave CLF rotation with simplimax polish.

We have not to this point illustrated Theorem 3. This theorem suggests an appropriate quadratic right constant CLF rotation may produce the same result as simplimax. To illustrate this note that using Figure 5, the quadratic right constant CLF in Figure 1 satisfies Theorem 3 when $m = 27$. Using this CLF to polish the basic concave CLF rotation produced the simplimax rotation as suggested by the theorem.

8 Choosing a CLF

It is interesting to note in how similar the basic concave and linear CLF rotations are in the Tables 1 and 7 that display both. While not displayed this similarity was also true for the 24 psychological tests data in Table 3. Consider the parameterized family of strictly concave CLFs defined by

$$h_b(|\lambda|) = 1 - e^{-|\lambda|/b}$$

When $b = 1$ this is the basic concave CLF. Because CLFs that differ by a positive multiplier define equivalent criteria, when comparing CLFs it helps to re-scale them so they all have the value one when $|\lambda| = 1$. This eliminates irrelevant differences when making comparisons. Figure 6 displays re-scaled versions of the CLFs, h_b for $b = .5, 1, 2, \infty$. The $b = \infty$, CLF is the linear CLF. This labeling is motivated by the fact that when re-scaled $h_b(|\lambda|)$ approaches $|\lambda|$ as $b \to \infty$. The parameter b may be viewed as a measure of convexity or more precisely an inverse measure of convexity.

The criteria in Figure 6 were compared using the box data. A rotation was obtained for each value of b and a measure of agreement computed for each pair. Given a pair of rotations Λ_1 and Λ_2 the number of decimal places of agreement was computed using

$$\text{agree} = -\log_{10}(\|\Lambda_1 - \Lambda_2\|/(pk)^{1/2})$$

where $\|M\|$ denotes the Frobenius norm of the matrix M. Table 9 displays the agreement. The agreement between the basic concave and linear CLF rotations observed in Table 7 seems to hold for a variety of values of the parameter b. This suggests that when using concave CLFs the degree of convexity is not a critical factor.

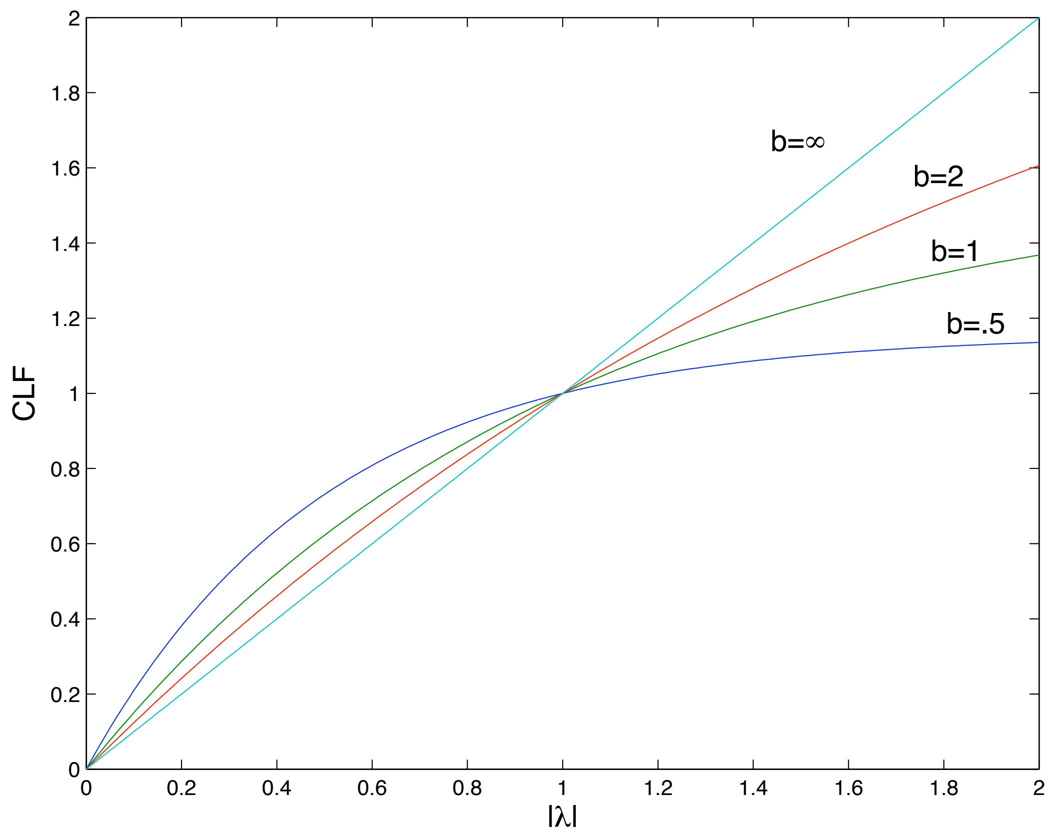

Figure 6: Comparison of scaled concave CLFs, h_b

Table 9: The number of decimal places of agreement between rotations of the box data using several concave CLFs, h_b.

	$b = 1$	$b = 2$	$b = \infty$
$b = .5$	3.43	3.03	2.70
$b = 1$		3.22	2.77
$b = 2$			2.95

9 Discussion

We have investigated the use of CLFs for oblique rotation extending the work of Jennrich (2004(a)) for orthogonal rotation. These functions assign a loss to each component of a loading matrix and oblique rotation is used to minimize the total loss. Given the simplicity and directness of this approach it is surprising that it has received so little attention.

The CLF approach has some nice theoretical properties. The most important of these is that when a CLF is concave and nondecreasing the corresponding criterion is minimized by a loading matrix with perfect simple structure if there is such a loading matrix. This motivates using concave CLFs. These appear to work well not only for perfect simple structure applications, but in the examples considered for Thurstone simple structure and more general applications as well. Indeed this also applies to the simplest of all concave CLFs which is the linear CLF. The corresponding CLC is arguably the simplest of all rotation criteria that are functions of the absolute loadings. These include all criteria known to the author. That linear CLF rotation works as well as some of the best methods available in the cases considered is quite surprising. Why, one wonders, has this not been observed earlier? One possible reason is that until now there has been no theory to point to the importance of concave CLFs. Another may be a computational problem associated with using concave CLFs.

The computational problem is that CLC defined by concave CLFs are not differentiable at a loading matrix with one or more zero values. We have dealt with this by rounding concave CLFs at the origin. This provides a good approximation to a truly concave CLF, but puts a strain on the optimization algorithm used because the resulting CLC, while differentiable, is not very smooth at a loading matrix with one or more small loadings which are the loading matrices of greatest interest. We have not discussed this computing issue because our GP algorithm worked well enough. This issue, however, deserves further investigation.

In our presentation we have focused on the basic concave CLF rather than on the linear CLF. In hind site this may have been a mistake because the linear CLF appears to work as well as the basic CLF and is simpler. It seems important to use a concave CLF, but the choice of concave CLF seems less important. To avoid what at present seems to be unnecessary complexity, the author recommends using the simple linear CLF.

In the examples the geomin, simplimax, and the concave CLF methods worked quite well. It is perhaps worth noting that geomin, which is not a CLF method, has a concave CLF flavor. To see this note that the geomin criterion can be written in the form

$$Q(\Lambda) = \sum_i \exp(\sum_r h(|\lambda_{ir}|)$$

where

$$h(|\lambda|) = 2\log(|\lambda|)/k$$

is a concave CLF.

10 Appendix: Epsilon modification

If h has a nonzero right derivative at zero, then $h(|\lambda|)$ is not a differentiable function of λ at $\lambda = 0$. To fix this choose an $\epsilon > 0$, let

$$b = \frac{h'(\epsilon)}{2\epsilon} \qquad \text{and} \qquad a = h(\epsilon) - b\epsilon^2$$

and let

$$h_\epsilon(u) = \begin{cases} a + bu^2 & 0 \le u \le \epsilon \\ h(u) & u > \epsilon \end{cases}$$

For small values of λ

$$h_\epsilon(|\lambda|) = a + b\lambda^2$$

and this is differentiable at $\lambda = 0$. Moreover, h and h_ϵ have the same value and derivative at ϵ. To see this note that

$$h_\epsilon(\epsilon) = a + b\epsilon^2 = h(\epsilon)$$

The left derivative of h_ϵ at ϵ is

$$2b\epsilon = h'(\epsilon)$$

which is also the right derivative of h_ϵ at ϵ. Thus h_ϵ is differentiable at ϵ and $h'_\epsilon(\epsilon) = h'(\epsilon)$.

11 References

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

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

Cureton, E. E. & Mulaik, S. A. (1975). The weighted varimax rotation and the promax rotation. *Psychometrika, 40*, 183-185.

Eber, H. W. (1966). Toward Oblique Simple Structure: Maxplane. *Multivariate Behavioral Research, 1*, 112-125.

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

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

Jennrich, R.I. (2004(a)). Rotation to simple loadings using component loss functions: The orthogonal case. *Psychometrika*, in press.

Jennrich, R.I. (2004(b)). Derivative free gradient projection algorithms for rotation. *Psychometrika*, in press.

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

Katz, J. O. & Rohlf, F. J. (1974). FUNCTIONPLANE—A new approach to simple structure rotation. *Psychometrika, 39*, 37-51.

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

Rozeboom, W. W. (1991). Theory and practice of analytic hyperplane optimization. *Multivariate Behavioral Research, 26*, 179-197.

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

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