UCLA

Department of Statistics Papers

Title

Rotation to Simple Loadings using Component Loss Functions: The Orthogonal Case

Permalink

<https://escholarship.org/uc/item/0f84000d>

Author

Robert Jennrich

Publication Date

2011-10-25

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

Robert I. Jennrich

May 24, 2003

Component loss functions (CLFs) are used to generalize the quartimax criterion for orthogonal rotation in factor analysis. These replace the fourth powers of the factor loadings by an arbitrary function of the second powers. Criteria of this form were introduced by a number of authors primarily Katz and Rohlf (1974) and Rozeboom (1991), but there has been essentially no follow up to this work. A method so simple, natural, and general deserves to be investigated more completely. A number of theoretical results are derived including the fact that any method using a concave CLF will recover perfect simple structure whenever it exists and there are methods that will recover Thurstone simple structure whenever it exists. Specific CLFs are identified and it is shown how to compare these using standardized plots. Numerical examples are used to illustrate and compare CLF and other methods. Sorted

absolute loading plots are introduced to aid in comparing results and setting parameters for methods that require them.

Key words: Factor analysis, component loss criteria, gradient projection, hyperplane count methods, quartimax, minimum entropy, simplimax, sorted absolute loading plots, varimax. ¹

¹The author is very indebted to a reviewer for pointing him to the generalized hyperplane count literature and to all the reviewers for valuable comments and suggestions.

1 Introduction

The rotation problem in factor analysis arises from a desire to find a simple and contextually meaningful relation between items and factors. Rotation methods attempt to achieve this by rotating factors to produce simple loading matrices. Unfortunately simple loading matrices are not well defined. Thurstone has set forth a number of general principals which vaguely stated say a large number of small loadings and a few large ones are what one should attempt to achieve. Actually Thurstone's (1935, p.156) conditions are precise, but in general unattainable and hence at best can only be approximated. At first attempts were made to approximate Thurstone'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 without the growth function requirement.

More specifically we consider a class of criteria that may be viewed as a

generalization of the quartimax criterion (Newhaus and Rigley, 1954). These are defined by an arbitrary component loss function (CLF) that is evaluated at the square 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 . These criteria include the Katz and Rohlf criteria, the Rozeboom criteria, and others that will be introduced. The CLF or what might be called the generalized hyperplane count approach has been largely overlooked which is unfortunate because a method so simple, natural, and general needs to be more carefully considered if for no other reason than to understand why it should not be pursued. Actually, in the orthogonal case considered, an appropriate CLF method can handily outperform quartimax and varimax (Kaiser, 1958). Also of interest is that Browne's (1972) partially specified target method and its generalization to Kiers' SIMPLIMAX method may be viewed as weighted and iteratively re-weighted CLF methods respectively.

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. This is also called a perfect cluster configuration. One might argue that a minimum requirement for any proper rotation method is that it finds perfect simple structure when it exists. It is shown that assuming only that a CLF is concave (i.e., curved downward) is sufficient to guarantee this. This is important because it provides clear simple guidance for constructing CLFs. A one parameter family of criteria is introduced. Like the Katz and Rohlf and Rozeboom families it is designed to

produce as many small loadings as possible. It is shown that the local minima of this new family are identical to those of Kiers' (1994) SIMPLIMAX family. Moreover, when a Thurstone simple structure exists, it is a local minimizer of all criteria from the new family that have a sufficiently small parameter value.

In the orthogonal case considered, CLFs have three unnecessary degrees of freedom which can make them difficult to compare by looking at their plots. It is shown how to use standardized plots to eliminate this problem. A simple new entropy criterion that works well is introduced. Its advantage is that it is parameter free so when using it there is no need to find appropriate parameter values. A variety of examples are used to compare CLF and other methods. Sorted absolute loading plots are introduced to aid in comparing results and choosing parameters for methods that require them.

We consider the orthogonal case not because it is more important, but because the theory and some of the computing is considerably simpler than that for the oblique case which makes it a natural place to introduce basic ideas and results.

2 Rotation to simple loadings

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

$$\Lambda = AT$$

where T is an arbitrary orthogonal matrix. A minimizing value of Λ is called a rotation of A corresponding to Q . Different criteria Q produce different rotations of A . Some standard rotation methods are formulated as maximization problems. When necessary we will re-formulate them as equivalent minimization problems.

To minimize Q over the orthogonal rotations of A we use the gradient projection algorithm of Jennrich (2001, 2002). The GP algorithm converged to a stationary point in every application considered in this paper, including the random searches. The GP algorithm, together with examples of its use, may be downloaded from <http://www.stat.ucla.edu/research/gpa/>. Matlab, R(=S), SAS, and SPSS versions are given.

3 Component loss rotation criteria

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

$$Q(\Lambda) = \sum \sum h(\lambda_{ir}^2)$$

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 may be any real valued function with a domain large enough to allow the evaluation of Q at any possible Λ . As noted, Katz and Rohlf (1974) and Rozeboom (1991) considered criteria of this form.

Figure 1: (a) The quartimax CLF. (b) The entropy CLF.

The quartimax criterion is the simplest example of a CLC. Formulated as a minimization problem, quartimax rotation minimizes a criterion of the form

$$Q(\Lambda) = - \sum \sum \lambda_{ir}^4$$

This is a CLC with CLF

$$h(\lambda^2) = -(\lambda^2)^2$$

This is plotted in Figure 1(a).

As an example of a non-standard, but interesting CLC, consider

$$Q(\Lambda) = - \sum \sum \lambda_{ir}^2 \log \lambda_{ir}^2 \quad (1)$$

Its CLF is

$$h(\lambda^2) = -\lambda^2 \log \lambda^2$$

This is plotted in Figure 1(b). As will be seen shortly, it is natural to call this an entropy criterion. This entropy criterion differs from that of McCammon

(1966) which is more complex, is the ratio of two entropies, and is not a CLC.

Note that both the quartimax and the entropy CLFs are concave. This property alone leads to some interesting results that are discussed in the next section.

One would like to compare CLC by comparing their CLFs. For this purpose the plots in Figure 1 are unnecessarily different. Note that adding a constant to a rotation criterion or multiplying it by a positive constant produces an equivalent criterion in the sense that the new criterion has the same minima as the original criterion. Because $\sum \sum \lambda_{ir}^2$ is constant under orthogonal rotation, adding multiples of this to a rotation criterion also produces an equivalent criterion. Thus if h is a CLF, then the CLF $\tilde{h}$ defined by

$$\tilde{h}(\lambda^2) = \alpha + \beta\lambda^2 + \gamma h(\lambda^2) \quad , \quad \gamma > 0 \quad (2)$$

is equivalent to h in the sense that it defines an equivalent CLC.

Thus the plots in Figure 1, and CLF plots in general, have 3 unnecessary degrees of freedom. These make comparisons of CLFs more difficult than necessary. To remove this difficulty one can standardize CLFs to make them unique. We say a CLF, h has standard form with respect to some $c > 0$ if

$$h(0) = 0 \quad , \quad h(c) = 1 \quad , \quad h'(c) = 0$$

where $h'(c)$ denotes the derivative of h at c . The values of the λ_{ir}^2 may range from zero to the largest communality for the items x_i . This suggests plotting h over this range and using the largest communality for c . Doing this puts the forced stationary point for h at the right hand side of the plot. The following theorem tells how to express CLFs in standard form.

Theorem 1: If h is a CLF differentiable at $c > 0$ and

$$h(c) - h(0) \neq ch'(c) \tag{3}$$

then

$$\tilde{h}(\lambda^2) = \frac{h(\lambda^2) - h(0) - \lambda^2 h'(c)}{h(c) - h(0) - ch'(c)}$$

is a standard form CLF equivalent to h .

Proof: Note that $\tilde{h}$ is of the form (2) and hence is equivalent to h . Note also that

$$\tilde{h}(0) = 0 \quad , \quad \tilde{h}(c) = 1 \quad , \quad \tilde{h}'(c) = 0$$

Thus $\tilde{h}$ has standard form. This completes the proof.

It is shown in the Appendix that assumption (3) is satisfied if h is concave and nonlinear on the interval $[0, c]$. Using Theorem 1 and $c = 1$ the quartimax and entropy CLFs in Figure 1 may be expressed in the standard forms displayed in Figure 2(a).

Note that the CLFs in Figure 1 look very different, but after the alignment displayed in Figure 2(a) they are quite similar. The similarity is real because the CLFs in Figure 2(a) are equivalent to the corresponding CLFs in Figure 1. While the similarity of the Figure 2(a) plots leads one to expect similar performance there is an important difference that will be noted in Section 3.2.

Figure 2: (a) The quartimax (lower) and entropy (upper) CLFs expressed in the standard format defined by Theorem 1 using $c = 1$. The quartimax curve is $2\lambda^2 - \lambda^4$ and the entropy curve is $\lambda^2(1 - \log \lambda^2)$. (b) A concave CLF with two squared loadings marked by asterisks.

3.1 Concave CLFs and perfect simple structure

Let $w_1, \dots, w_n$ be a mass distribution, that is a sequence of non-negative numbers.

$$E(w) = - \sum_{i=1}^n w_i \log w_i$$

is called its entropy. As is well known, and shown below, among all n -component mass distributions with a specified total mass t , the uniform distribution with $w_i = t/n$ for all i has maximum entropy and any distribution with its total mass at a single point has minimum entropy. This result may be generalized as follows.

Theorem 2: Let $w = (w_1, \dots, w_n)$ be a mass distribution with $\sum w_i = t$

and let

$$f(w) = \sum h(w_i)$$

where h is a concave function. Then f is maximized by the uniform distribution $w_i = t/n$ for all i and minimized by any distribution with $w_i = t$ for some i and $w_j = 0$ for all $j \neq i$.

Proof: By Jensen's inequality (see e.g. Ferguson, 1967, p76)

$$\frac{1}{n} \sum h(w_i) \leq h\left(\frac{1}{n} \sum w_i\right)$$

Thus

$$\sum h(w_i) \leq \sum h(t/n)$$

This proves the first assertion. To prove the second note that we may assume without loss of generality that $h(0) = 0$ and note that

$$w_i = \left(1 - \frac{w_i}{t}\right)0 + \frac{w_i}{t}t$$

Because this is a convex combination of 0 and t and because h is concave

$$h(w_i) \geq \left(1 - \frac{w_i}{t}\right)h(0) + \frac{w_i}{t}h(t) = \frac{w_i}{t}h(t)$$

If $\tilde{w}_i = t$ for some i and $\tilde{w}_j = 0$ for all $j \neq i$,

$$\sum h(w_i) \geq h(t) = \sum h(\tilde{w}_i)$$

This proves the second assertion and the theorem.

Note that the entropy criterion (1) is the entropy of the squared loadings λ_{ir}^2 and hence is appropriately called an entropy criterion.

A well known result is that when there is a perfect simple structure rotation, it minimizes the quartimax criterion. The following theorem proves and generalizes this result to all CLC with concave CLF.

Theorem 3: If a rotation $\hat{\Lambda}$ of an initial loading matrix A has perfect simple structure, then $\hat{\Lambda}$ minimizes any CLC that has a concave CLF.

Proof: Let c_i be the squared length of the i -th row of A . This is called its communality. Let Λ be any rotation of A . Because the rotations are orthogonal

$$c_i = \sum_r \lambda_{ir}^2 = \sum_r \hat{\lambda}_{ir}^2$$

Because each row of $\hat{\Lambda}$ has at most one nonzero element, it follows from Theorem 2 that

$$\sum_r h(\hat{\lambda}_{ir}^2) \leq \sum_r h(\lambda_{ir}^2)$$

Thus

$$\sum_i \sum_r h(\hat{\lambda}_{ir}^2) \leq \sum_i \sum_r h(\lambda_{ir}^2)$$

which completes the proof.

Because quartimax and minimum entropy have concave CLFs it follows from Theorem 3 that they are minimized by rotated loading matrices with perfect simple structure whenever they exist.

As a second application of Theorem 2 consider the following question. When will quartimax and varimax rotation produce the same result? The following theorem gives a simple sufficient condition for this. Let Λ^2 be the element-wise square of Λ and let λ_{+r}^2 be the sum of the elements in the r -th column of Λ^2 .

Theorem 4: If $\hat{\Lambda}$ is a quartimax rotation of A and if

$$\hat{\lambda}_{+1}^2 = \cdots = \hat{\lambda}_{+k}^2$$

then $\hat{\Lambda}$ is also a varimax rotation of A .

Proof: The varimax criterion may be expressed in the form

$$V(\Lambda) = \sum \sum \lambda_{ir}^4 - \frac{1}{p} \sum_r (\sum_i \lambda_{ir}^2)^2$$

Let

$$h(\lambda^2) = -(\lambda^2)^2$$

Then $V(\Lambda)$ may be expressed in the form

$$V(\Lambda) = \sum \sum \lambda_{ir}^4 + \frac{1}{p} \sum_r h(\lambda_{+r}^2) \quad (4)$$

Because the first term on the right defines the quartimax criterion and $\hat{\Lambda}$ is a quartimax rotation of A , $\hat{\Lambda}$ maximizes the first term on the right of (4).

Note that for any rotation Λ of A

$$\sum_r \lambda_{+r}^2 = t$$

where t is the sum of the squared components of A . Thus for any rotation Λ of A

$$\lambda_{+1}^2, \cdots, \lambda_{+k}^2$$

may be viewed as a distribution on the integers $1, \cdots, k$ with total mass t .

Because h is concave, it follows from Theorem 2 that

$$\sum_r h(\lambda_{+r}^2)$$

is maximized if

$$\lambda_{+1}^2 = \cdots = \lambda_{+k}^2$$

Because this, by assumption, is true for the quartimax rotation $\hat{\Lambda}$ of A , $\hat{\Lambda}$ also maximizes the second term on the right in (4), and hence maximizes V . Thus $\hat{\Lambda}$ is a varimax rotation of A . This completes the proof.

3.2 Less than perfect simple structure

The matrix A on the left in Table 1 satisfies Thurstone's (1935, p.156) three conditions for simple structure:

1. Each row of the factor structure should have at least one zero.
2. Each column should have at least k zeros.
3. For every pair of columns there should be at least k variables whose entries vanish in one column but not the other.

The second and third matrices in Table 1 are quartimax and minimum entropy rotations of A . To the precision displayed, the minimum entropy rotation has Thurstone simple structure and by almost any definition of simplicity, it is significantly simpler than the quartimax rotation. Because the columns of the quartimax rotation have equal length, Theorem 4 suggests that the varimax rotation of A may be identical to the quartimax rotation and this is true to the precision displayed. Thus in this example at least, there is a CLF method that performs significantly better than the quarti-

Table 1: Quartimax, varimax, and minimum entropy rotations of A .

A			quartimax varimax			entropy		
1	0	0	.67	.67	-.33	1.00	.00	.00
0	1	0	.67	-.33	.67	.00	1.00	.00
0	0	1	-.33	.67	.67	.00	.00	1.00
.71	.71	0	.95	.24	.24	.71	.71	.00
.71	0	.71	.24	.95	.24	.71	.00	.71
0	.71	.71	.24	.24	.95	.00	.71	.71
.71	.71	0	.95	.24	.24	.71	.71	.00
.71	0	.71	.24	.95	.24	.71	.00	.71
0	.71	.71	.24	.24	.95	.00	.71	.71

max and popular varimax methods. The example provides a simple clear motivation for an interest in CLF methods.

As pointed out by the associate editor, another problem with the quartimax and varimax methods for this problem is that small changes in A produce large changes in its rotation. This is discussed briefly in the Appendix.

Note that when plotted in standard form as in Figure 2(a) the slope near the origin for the entropy CLF is greater than that for the quartimax CLF and is in fact infinite at zero. This means that for values of λ^2 near the origin a move toward the origin will produce a greater reduction in $h(\lambda^2)$ for the entropy CLF than for the quartimax CLF and this may explain why the

minimum entropy rotation produced significantly better results.

4 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 many small values and a small number of larger values is simpler than one with mostly intermediate values. Motivated by this, we will consider methods that produce small or large loadings and hopefully not too many intermediate loadings. This is also the goal of the generalized hyperplane count methods.

Because many rotation criteria, including all mentioned thus far, are influenced more by large rows of Λ 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 will use initial loading matrices A with normalized rows or at least nearly normalized rows.

Another issue we would like to avoid at least for now is that of local minima. We will do this by arbitrarily defining the best rotation produced from 20 random starts our GP algorithm to be an operational minimizer of the rotation criterion used. This strategy was used by Kiers (1994) and

Browne (2001). We will comment further on the local minimum problem in Section 7. By a random start we mean an orthogonal matrix T that is randomly selected from the uniform distribution on the group of orthogonal matrices. These were generated as described in the Appendix.

4.1 Concave CLF are desirable

The quartimax and entropy CLFs in Figure 1 are both concave. We would like to argue that this is generally a good idea. We already know from Theorem 3 that using a concave CLF will produce perfect simple structure whenever it exists.

Consider the general case when there is no rotation with perfect simple structure. For orthogonal rotation the sum of the squared loadings is constant. Thus the only way one squared loading can be reduced is by increasing at least one other. In Figure 2(b) two squared loadings are indicated by asterisks below a strictly concave CLF plot. If the smaller squared loading is made smaller and the larger squared loading is made larger by an equal amount, the sum of the two CLF values is decreased because the CLF is strictly concave. Thus in general a concave CLF encourages making small loadings smaller and large loadings larger which is what one wants to produce simple structure. This last argument is very heuristic, but again it suggests that concave CLFs are a good idea, at least in the context of orthogonal rotation. Rozeboom recommends the use of concave CLFs for the oblique case at least away from a neighborhood of the origin.

4.2 CLFs designed to produce small loadings

We consider two families of CLFs specifically designed to produce small loadings.

The first is the family proposed by Katz and Rohlf (1974). It is defined by CLFs of the form

$$h(\lambda^2) = 1 - \exp(-(\lambda/b)^2)$$

Actually Katz and Rohlf allow the exponent 2 on the right to be replaced by some other power, but in their examples they use only the power 2. This CLF is plotted in Figure 3(a) for $b = .3$. The parameter b is called a bandwidth parameter. For any bandwidth b , when $\lambda = 2b$, $h(\lambda^2) = .982$. Thus b is half the distance to the value of λ at which $h(\lambda^2)$ comes close to its maximum value. In the figure this is at $\lambda^2 = .36$. Note finally that Katz and Rohlf CLFs are concave.

Because it has some nice theoretical properties we will also consider the family of linear right constant criteria defined by CLFs of the form

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

This is plotted in Figure 3(b) for $b = .6$. Again b is a bandwidth parameter. It is the value of λ at which $h(\lambda^2)$ becomes constant. Thus a linear right constant CLF with bandwidth $2b$ approximates the Katz and Rohlf CLF with bandwidth b . Note, finally, that linear right constant CLFs are also concave.

Figure 3: (a) Katz and Rohlf CLF with $b = .3$. (b) Linear right constant CLF with $b = .6$.

Linear right constant rotation has the following exact zero property. Consider the matrix A in Table 1 and a linear right constant CLF h with any bandwidth $b < .71$. A small change in A will increase, or at least cannot decrease, the value of the corresponding CLC. To see this note that h is equal to one for all 9 nonzero loadings in A and zero for all 9 zero loadings. A small change in A will not change the value of h for the nonzero loadings, and will increase it, or at least not decrease it, for the zero loadings. Thus A is a local minimum of any right constant CLC with $b < .71$. More generally if Λ is a rotation of an initial loading matrix A and Λ has Thurstone simple structure, then Λ is a local minimizer of any linear right constant CLC with bandwidth less than the smallest nonzero element in Λ . This local minimizer result does not apply to the other CLF methods we have considered, but it is reasonable to expect it to hold approximately for methods that use CLFs

Table 2: Quartimax, minimum entropy, and linear right constant (linrc) with $b = .3$ rotation of a matrix A with Thurstone simple structure.

A			quartimax			entropy			linrc		
1	0	0	.99	-.11	-.11	1.00	-.03	-.03	1.00	.00	.00
0	1	0	.11	.99	-.01	.03	1.00	.00	.00	1.00	.00
0	0	1	.11	-.01	.99	.03	.00	1.00	.00	.00	1.00
.89	.45	0	.93	.35	-.10	.90	.42	-.03	.89	.45	.00
.89	0	.45	.93	-.10	.35	.90	-.03	.42	.89	.00	.45
0	.71	.71	.16	.70	.70	.05	.71	.71	.00	.71	.71
.89	.45	0	.93	.35	-.10	.90	.42	-.03	.89	.45	.00
.89	0	.45	.93	-.10	.35	.90	-.03	.42	.89	.00	.45
0	.71	.71	.16	.70	.70	.05	.71	.71	.00	.71	.71

that approximate linear right constant CLFs. For example it is reasonable to expect the result to hold approximately for the Katz and Rohlf family.

The first matrix in Table 2 is an initial loading matrix that satisfies Thurstone's conditions for perfect simple structure. The second matrix is a quartimax rotation of A and the third is a minimum entropy rotation of A . The final matrix is a CLF rotation of A using a linear right constant CLF with $b = .3$. To the precision displayed, linear right constant rotation produced Thurstone simple structure. Minimum entropy rotation came fairly close to doing this, and quartimax rotation did not do as well. Actually linear right constant rotation approximated A to 6.77 decimal places which suggests that

if it could have been computed exactly, it would have produced A exactly as is expected because $b = .3 < .45$. This provides empirical support for the Thurstone simple structure property of linear right constant methods. The number of decimal places of agreement was computed using

$$-\log_{10}(\max_{ir} |\lambda_{ir} - a_{ir}|)$$

As noted the Katz and Rohlf CLF with $b = .15$ approximates the linear right constant CLF with $b = .3$. When $b = .15$ is used for Katz and Rohlf rotation, the result is the same as the linear right constant result in Table 2 to the precision displayed. Thus it also recovered Thurstone simple structure to the precision displayed.

It should be noted that while families of CLFs provide flexibility, they have a serious drawback in that they require the specification of parameter values. In this regard the entropy CLF is attractive because it has no parameters and so far at least seems to perform reasonably well. We have not considered Rozeboom's four parameter family because it has more flexibility than we are prepared to discuss.

4.3 SAL plots for comparing results and choosing bandwidths

While it is fairly easy to compare the rotations in Table 2, even in this simple example one must look fairly hard to form an opinion. Something that helps, and helps in choosing bandwidth parameters, is a sorted absolute

loading (SAL) plot. This is defined by its name. More specifically let $n = pk$ and let

$$|\lambda_1| \leq \cdots \leq |\lambda_n|$$

denote the absolute values of the loadings in Λ sorted into increasing order. The SAL plot is a plot of $|\lambda_i|$ on i for $i = 1, \dots, n$.

Figure 4 displays the SAL plots for the quartimax and minimum entropy loading matrices in Table 2. Displayed in this way one is almost immediately motivated to prefer the minimum entropy rotation to the quartimax rotation. As before, minimum entropy seems to encourage small loadings more than quartimax.

The SAL plot is also useful for choosing bandwidths. Both plots in Figure 4 suggest choosing b for linear right constant rotation somewhere between .2 and .3 in the hope of further reducing the apparent small loadings. As noted in Section 4.2 using $b = .3$ and linear right constant rotation in fact makes them zero and gives Thurstone simple structure.

Note that the sorted absolute loading plot is very easy to make. In particular it does not require the often difficult to produce column permutations and sign changes required to align loading matrices for comparison. While easier to compare than a set of aligned matrices, SAL plots contain less information. They display only the distribution of the absolute loadings and not the sign and location of the loadings within a loading matrix. SAL plots are not meant to replace aligned matrices, but rather to provide an easy and useful first look when comparing a number of loading matrices and to aid in

Figure 4: Sorted absolute loading plot for the quartimax (o) and minimum entropy (*) rotations of the matrix A in Table 2.

choosing bandwidths for methods that require them. For the author at least, after first using them, it has become very difficult to work without them.

5 Some related methods

We consider here a method proposed by Browne (1972) and an orthogonal analog of an oblique rotation method proposed by Kiers (1994). While not CLF methods, these methods are closely related to them. Indeed, in a sense to be defined, Kiers’ method is equivalent to the linear right constant CLF method.

5.1 Partially specified target rotation

Browne (1972) proposed a method for rotating to a partially specified target. When the specified target values are all zero, his method minimizes the criterion

$$B(\Lambda) = \sum \sum w_{ir} \lambda_{ir}^2$$

where $w_{ir} = 1$ when the corresponding target value is zero and $w_{ir} = 0$ when the target value is unspecified. Browne’s criterion may be viewed as a weighted CLC with weights w_{ir} and CLF

$$h(\lambda^2) = \lambda^2$$

We have not discussed weighted CLC, but they represent a natural generalization of the CLC we have discussed. Browne suggested his method might be used to “improve” on a loading matrix produced by some other analytic

rotation procedure. For example the small loadings in a varimax rotation might be used to specify target zeros. Browne's method is then used to rotate the varimax solution toward this partially specified target.

5.2 SIMPLIMAX rotation

Kiers (1994) proposed the following method for oblique rotation. Given a loading matrix Λ let $\lambda_{(m)}^2$ be the m -th smallest of the squared loadings λ_{ir}^2 . A SIMPLIMAX rotation of an initial loading matrix A is an oblique rotation Λ of A that minimizes

$$K_m(\Lambda) = \sum \sum I(\lambda_{ir}^2 \leq \lambda_{(m)}^2) \lambda_{ir}^2$$

where $I(\cdot)$ is 1 when its argument is true and zero otherwise. This may be viewed as a weighted CLC with dynamic weights $w_{ir} = I(\lambda_{ir}^2 \leq \lambda_{(m)}^2)$ corresponding to the m smallest loadings. The orthogonal analog is obtained by replacing oblique rotation of A by orthogonal rotation.

The criterion K_m is not a CLC because $I(\lambda_{ir}^2 \leq \lambda_{(m)}^2)$ is a function of all the loadings in Λ and not just the component λ_{ir} . It may, however, be viewed as an iteratively re-weighted CLC. We will show that although they are different criteria, SIMPLIMAX and criteria using linear right constant CLFs can and often do produce the same loadings.

Let Q_b be a CLC using a linear right constant CLF with bandwidth b .

Theorem 5: If for a loading matrix $\hat{\Lambda}$, b^2 is strictly between $\hat{\lambda}_{(m)}^2$ and $\hat{\lambda}_{(m+1)}^2$, then $\hat{\Lambda}$ is a local minimum of K_m if and only if it is a local minimum of Q_b .

Proof: Because b^2 is strictly between $\hat{\lambda}_{(m)}^2$ and $\hat{\lambda}_{(m+1)}^2$, for all Λ in a sufficiently small neighborhood of $\hat{\Lambda}$, b^2 is strictly between $\lambda_{(m)}^2$ and $\lambda_{(m+1)}^2$. For such Λ

$$K_m(\Lambda) = b^2 Q_b(\Lambda) + m - pk$$

as can be seen by writing out the definitions of K_m and Q_b . Since this holds for all Λ in a neighborhood of $\hat{\Lambda}$, $\hat{\Lambda}$ is a local minimum of K_m if and only if it is a local minimum of Q_b .

Theorem 5 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 right constant linear CLF methods are equivalent even though the rotation criteria are not.

For the data in Table 2, the SIMPLIMAX method with $m = 9$ produced the same results as the linear right constant method with $b = .3$. This is predicted by Theorem 5 because for the linear right constant loadings $|\hat{\lambda}_{(9)}| \leq .3 \leq |\hat{\lambda}_{(10)}|$.

6 Thurstone's box problem

It will be helpful to look at a familiar more realistic problem than the simple examples considered thus far. Table 3 contains the quartimax and minimum entropy rotations for Thurstone's (1947, p.136) box problem together with a SIMPLIMAX rotation with $m = 27$.

The rotations in Table 3 seem similar. To investigate this further, Figure 5 displays the sorted absolute loading plots for the minimum entropy

Table 3: Quartimax, minimum entropy, and SIMPLIMAX rotation of Thurstone's box data.

formula	quartimax			min. entropy			SIMPLIMAX		
x^2	.99	.09	.01	.99	.10	.03	.99	.11	.05
y^2	.17	.97	.16	.15	.97	.14	.14	.98	.12
z^2	.10	.08	.98	.07	.10	.98	.05	.13	.98
xy	.60	.79	.13	.59	.80	.12	.58	.81	.12
xz	.47	.09	.87	.45	.11	.88	.43	.14	.89
yz	.14	.45	.88	.11	.47	.87	.09	.49	.86
$\sqrt{x^2 + y^2}$	.81	.59	.07	.80	.60	.08	.80	.61	.08
$\sqrt{x^2 + z^2}$	.91	.12	.41	.90	.13	.43	.89	.15	.44
$\sqrt{y^2 + z^2}$	.14	.81	.58	.12	.82	.56	.10	.83	.55
$2x + 2y$	.72	.69	.10	.71	.70	.10	.70	.71	.10
$2x + 2z$	.95	.05	.50	.94	.07	.52	.92	.09	.54
$2y + 2z$	.14	.66	.74	.12	.68	.73	.09	.70	.72
$\log x$	.98	.12	.01	.98	.13	.03	.98	.14	.04
$\log y$	.12	.95	.21	.11	.95	.20	.09	.96	.18
$\log z$	.11	.04	.96	.08	.06	.96	.06	.08	.96
xyz	.41	.44	.78	.38	.46	.78	.36	.48	.78
$\sqrt{x^2 + y^2 + z^2}$	.75	.55	.36	.74	.56	.37	.73	.58	.37
e^x	.97	.05	.02	.97	.06	.04	.96	.07	.06
e^y	.21	.93	.01	.20	.94	.09	.18	.94	.07
e^z	.09	.09	.97	.07	.11	.97	.05	.14	.97

Figure 5: Sorted absolute loading plot for minimum entropy (o) and SIMPLIMAX (*) rotations of Thurstone's box data.

and SIMPLIMAX rotations. They also appear quite similar. Note that the largest jump in the minimum entropy plot occurs between the 27th and 28th absolute loading. This motivated using $m = 27$ for SIMPLIMAX rotation and illustrates another use for SAL plots and minimum entropy rotation.

7 Local minima

We defined a rotation to be an operational minimizer of a rotation criterion if it has the smallest criterion value among 20 rotations produced by the GP algorithm from random starts. One, of course, hopes that the minimum criterion value will be produced by more than one random start.

Table 4 gives the number of random starts that produced the minimum criterion value for each of our examples. For example for the data in Table 1 quartimax rotation produced the same criterion value from all of 20 random starts and minimum entropy did so for 18 out of 20. For the data in Table 2 the criteria considered produced apparent local minima more often. In particular all but 6 of the random starts produced local minima using linear right constant CLF rotation with $b = .3$. For the box data in Table 3 only SIMPLIMAX encountered local minima. For our examples 20 random starts seem to have been sufficient because in every case an operational minimizer was obtained by at least 6 random starts.

As indicated in Table 4 an identity start produced an operational minimizer in every case except for SIMPLIMAX in Table 3. This is not surprising for Tables 1 and 2 because for them an identity start is clearly excellent. Table 4 also reports the results obtained using quartimax starts. These produced operational minimizers in every case. While there is little evidence for this in our examples, it seems likely that the use of a quartimax start will produce an operational minimizer more often than an identity start. One might also consider using starts produced by other methods that seem to be immune to

Table 4: The number out of 20 random starts that produced operational minima for all of the examples considered and the success (Y,N) of the identity and quartimax starts in doing this.

	Example	# at min.	identity start	quartimax start
Table 1	quartimax	20	Y	Y
	min. entropy	18	Y	Y
Table 2	quartimax	20	Y	Y
	min. entropy	10	Y	Y
	lin. rt. const. $b = .3$	6	Y	Y
	Katz and Rohlf $b = .15$	7	Y	Y
	SIMPLIMAX $m = 12$	8	Y	Y
Table 3	quartimax	20	Y	Y
	min. entropy	20	Y	Y
	SIMPLIMAX $m = 27$	13	N	Y

the starting value problem such as varimax.

8 Discussion

We have investigated an overlooked class of rotation criteria based on CLFs. CLFs of special interest have been identified. These include quartimax, minimum entropy, the Katz and Rohlf family, and the linear right constant fam-

ily. Browne's rotation to a partially specified target and Kiers' SIMPLIMAX methods may be viewed as weighted and iteratively re-weighted CLF methods respectively. Standardized plots for CLFs were introduced to remove artificial differences between these plots and simplify their comparison.

A number of theoretical results were obtained including the ability of concave CLFs to recover perfect simple structure, the ability of appropriate linear right constant CLFs to recover Thurstone simple structure, and an equivalence between the linear right constant CLF and SIMPLIMAX methods. We compared the performance of a variety of CLF and other methods by means of examples using SAL plots to aid in these comparisons.

A question of interest that has not been addressed is to what extent a specific CLF method can be used without random starts by using some other method to produce a single starting value. Kiers (1994) provides some evidence that this may not work for SIMPLIMAX. To what extent is this true for other methods?

Another question arises from the fact that some methods, Katz and Rohlf, linear right constant, and SIMPLIMAX produce not one, but many rotations, one for each value of the parameter that indexes them. On one hand this is good because it provides an investigator with multiple solutions to consider. On the other hand it provides too many and it might be useful to start with a default choice. For example for Katz and Rohlf and linear right constant CLF methods $b = .15$ and $b = .3$ respectively may be reasonable first guesses. It is less clear how one might choose a default value for m in the SIMPLIMAX case, but a preliminary minimum entropy rotation may suggest one.

In the view of the author the most important next step is to consider the oblique case. While this is reasonably straight forward, there are some problems. The main problem is that under oblique rotation the sum of the squared loadings in each row of a loading matrix is no longer constant and the loadings are no longer bounded. These differences need to be dealt with. To what extent, for example, can the theoretical results for the orthogonal case be reformulated in the oblique case? Another problem is that optimization of some interesting new criteria in the oblique case is much more difficult than the optimizations encountered in the orthogonal case. Also in the oblique case numerical comparisons are of greater interest and should be more extensive.

Appendix

Generating random orthogonal matrices

Let Z be a k by k matrix whose components are independent standard normal variables. Let $Z = TR$ be a QR factorization of Z or equivalently let T be a Gram-Schmidt orthonormalization of Z . Then T is uniformly distributed on the group of orthogonal matrices.

The Matlab code

$$[T,R]=qr(randn(k,k))$$

will generate a realization T of T . Similar code may be used in other computing environments.

Condition (3)

Lemma 1: If h is concave and nonlinear on an interval $[0, c]$, then (3) holds.

Proof: The line tangent to h at c is given by

$$\ell(u) = h(c) + h'(c)(u - c)$$

Assume (3) is false. Then $h(c) - h'(c)c = h(0)$ and

$$\ell(u) = h(0) + h'(c)u$$

Because h is concave

$$h(u) \leq \ell(u) \tag{5}$$

Also because h is concave

$$\begin{aligned} h(u) &= h\left(\left(1 - \frac{u}{c}\right)0 + \frac{u}{c}c\right) \\ &\geq \left(1 - \frac{u}{c}\right)h(0) + \frac{u}{c}h(c) \\ &= h(0) + \frac{h(c) - h(0)}{c}u \\ &= h(0) + h'(c)u = \ell(u) \end{aligned}$$

Using this and (5) gives

$$h(u) = \ell(u)$$

which implies h is linear. This contradiction implies (3) is true.

Small changes in A in Table 1

The value .71 in the matrix A in Table 1 is approximately $1/\sqrt{2}$. The associate editor observed that if .71 is replaced by $1/\sqrt{2}$ there are many different

quartimax rotations of A . Indeed there appear to be an infinite number including one that leaves A unchanged. He also observed that if .71 is replaced by .70 a value slightly smaller than $1/\sqrt{2}$ quartimax rotation of A leaves A unchanged to the precision displayed in Table 1. For A modified in this way quartimax and varimax work quite well. Of course one does not have the option of modifying one's data to make a method work well and even if one did this may not work.

If a copy of the last three rows of A is appended to the end of A , small modifications to the .71 value in the appended A have almost no effect on the quartimax and varimax rotations of the appended A . They all look poor. Indeed they are very nearly equal to the quartimax and varimax rotations given in Table 1 with a copy of the last three rows appended.

The performance of the minimum entropy method is not effected by small changes in the .71 value in A in Table 1. Changing this value over the range from .60 to .80 gives in each case a minimum entropy rotation of A that leaves A unchanged to the precision displayed.

9 References

- Browne, M. W. (1972). Orthogonal rotation to a partially specified target. *British Journal of Mathematical and Statistical Psychology*, 25, 115-120.
- Browne, M. W. (2001). An overview of analytic rotation in exploratory factor analysis. *Multivariate Behavioral Research*, 36, 111-150.
- Eber, H. W. (1966). Toward Oblique Simple Structure: Maxplane. *Multi-*

- variate Behavioral Research*, 1, 112-125.
- Ferguson, T. S. (1967). *Mathematical Statistics*. New York: Academic Press.
- Jennrich, R. I. (2001). A simple general procedure for orthogonal rotation. *Psychometrika*, 66, 289-306.
- Jennrich, R.I. (2002). A simple general method for oblique rotation. *Psychometrika*, 67, 7-19.
- 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.
- McCammon, R.B. (1966). Principal components analysis and its application in large scale correlation studies. *Journal of Geology*, 74, 721-733.
- Neuhaus, J. O. & Wrigley, C. (1954). The quartimax method: An analytical approach to orthogonal simple structure. *British Journal of Statistical Psychology*, 7, 81-91.
- Rozeboom, W. W. (1991). Theory and practice of analytic hyperplane optimization. *Psychometrika*, 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.