Rotation Methods, Algorithms, and Standard Errors

Robert I. Jennrich

Outline

- The rotation problem

- Rotation methods
  - Graphical and analytic
  - Direct and indirect
  - Criteria for analytic methods

- Algorithms for rotation
  - Pairwise algorithms for quadratic criteria
  - Algorithms for general criteria

    Pairwise
    Gradient projection

- Standard errors for rotated loadings
  - Using the distribution of the initial loadings
  - Constrained MLE
  - Pseudo value methods
  - Jackknife, Bootstrap, MCMC
- A future direction for exploratory analysis

Introduction

Factor Analysis

- Model: $x = \Lambda f + u$

- Covariance structure:

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

The rotation problem

There are many Λ and Φ that satisify

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

Solution: Choose a Λ that looks nice

Rotation Methods: Oblique only

A parameterization for Λ and Φ

- Choose A so $\Lambda\Phi\Lambda' = AA'$

  A is called an initial loadings matrix

- Th: $\Lambda = AT^{-1}$, $\Phi = TT'$

  for some matrix T with rows of length one.

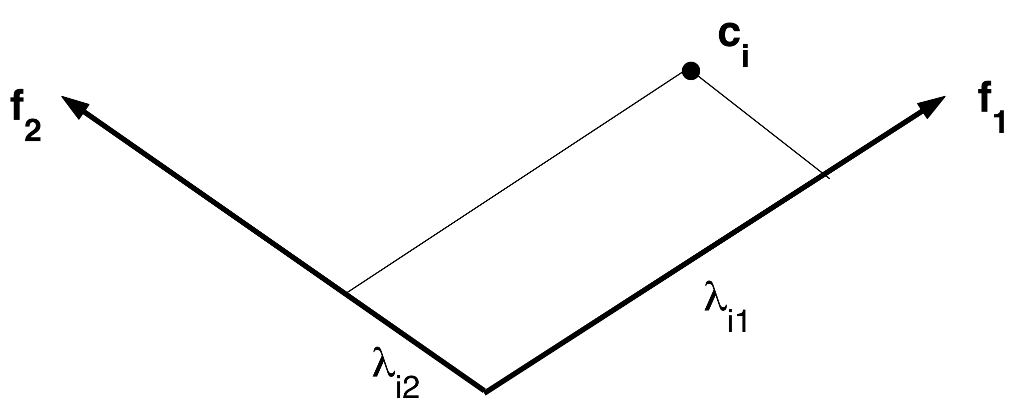
- The rows of T correspond to factors.

Graphical methods

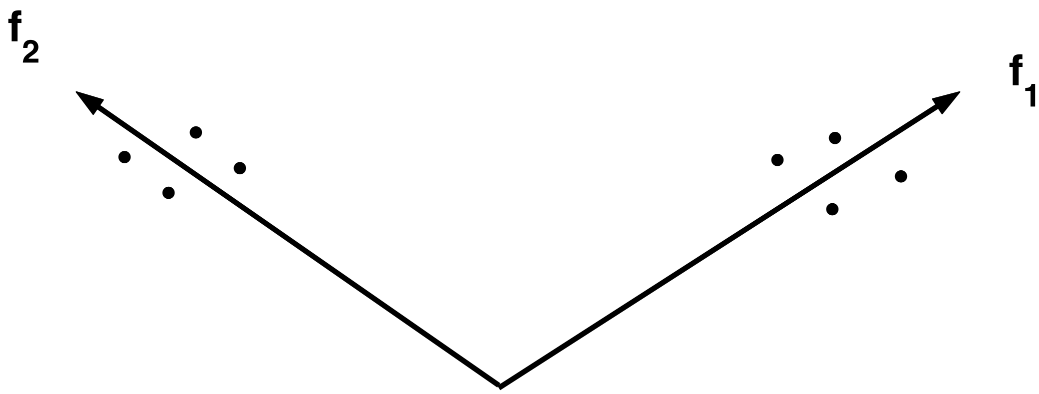
$c = \Lambda f$ = common part of x

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

In the case of two factors:



Plotting all c_i:



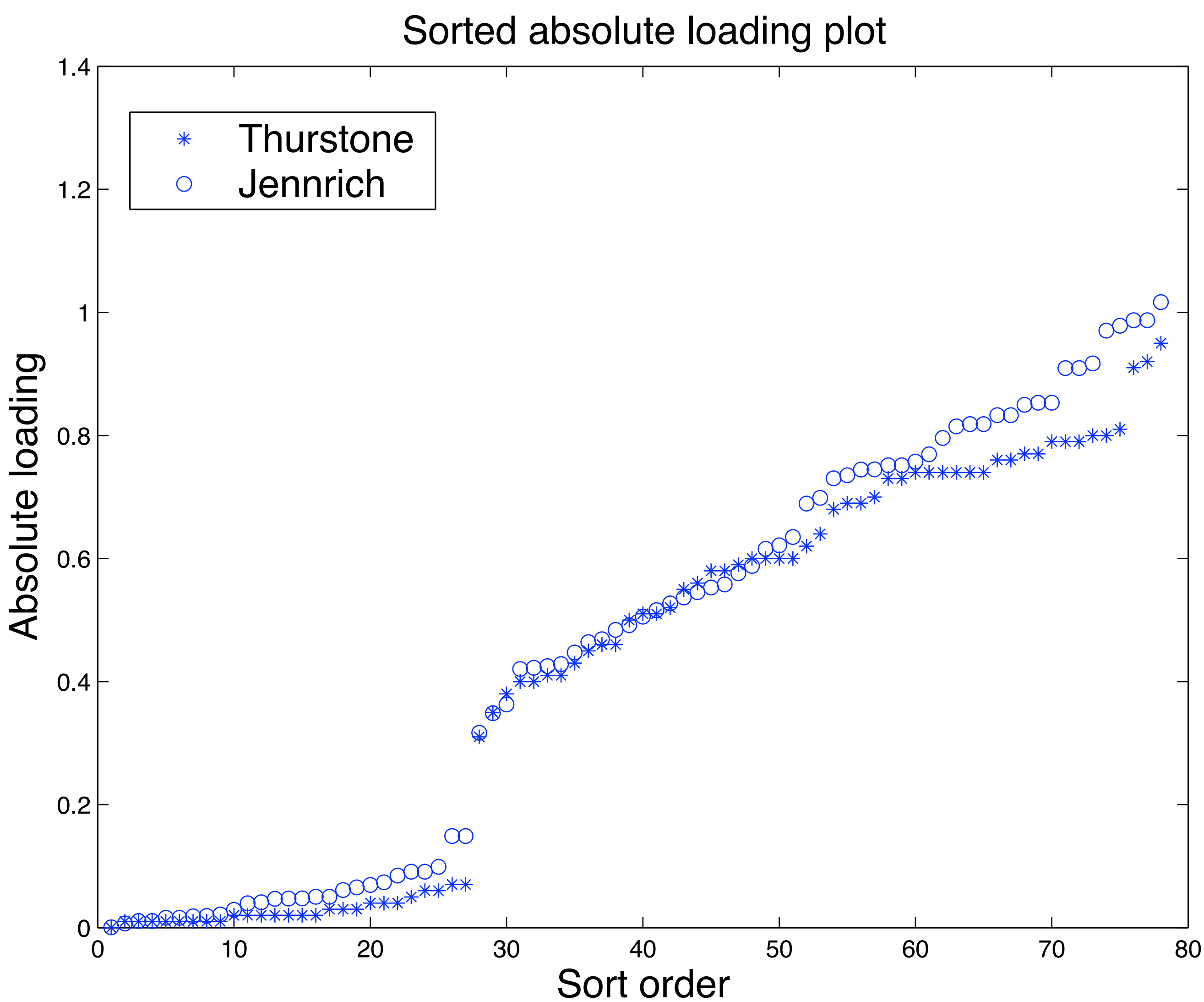
Choose f_1 and f_2 through the clusters.

For more than two factors one cycles through pairs of factors making similar plots.

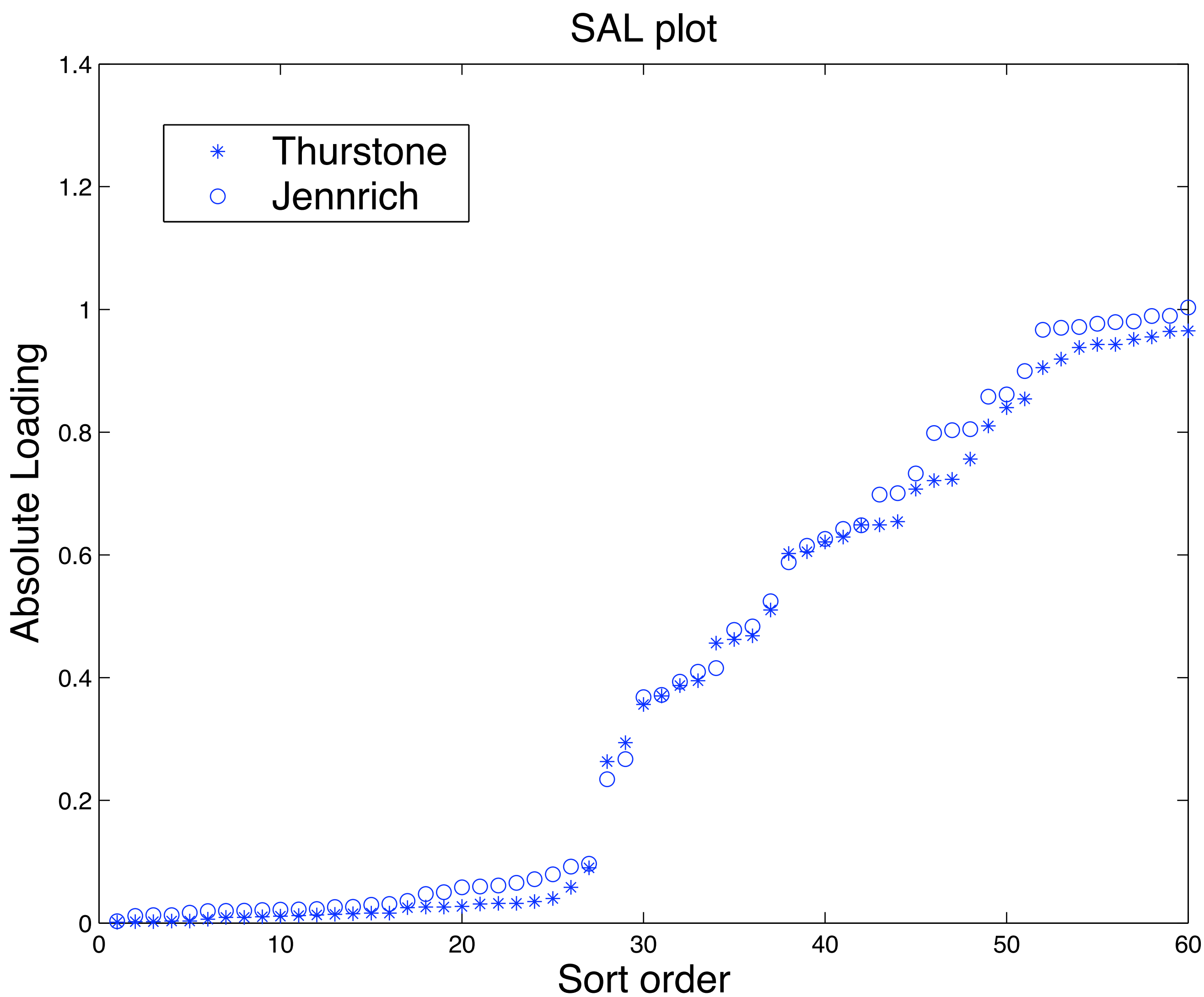
One can have the computer automate this.

15 lines of code.

For Thurstone's 26 variable box problem Thurstone and I got:



For Thurstone's 20 variable box problem Thurstone and I got:



Analytic methods: indirect

The motivation is to avoid T^{-1} in

$$Q(\Lambda) = Q(AT^{-1})$$

Following Thurstone (1947) let the rows of U be bi-orthogonal to the rows of T and use the reference structure

$$R = AU'$$

Th: $R = \Lambda\Delta \quad , \quad \Delta$ diagonal

R is simple $\Leftrightarrow \Lambda$ is simple

The analytic problem is to minimize

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

Harman calls making R simple an indirect method and making Λ simple a direct method.

Analytic methods: direct

Minimize

$$Q(\Lambda) = Q(AT^{-1})$$

Algorithms later.

Rotation criteria

- Quartic criteria

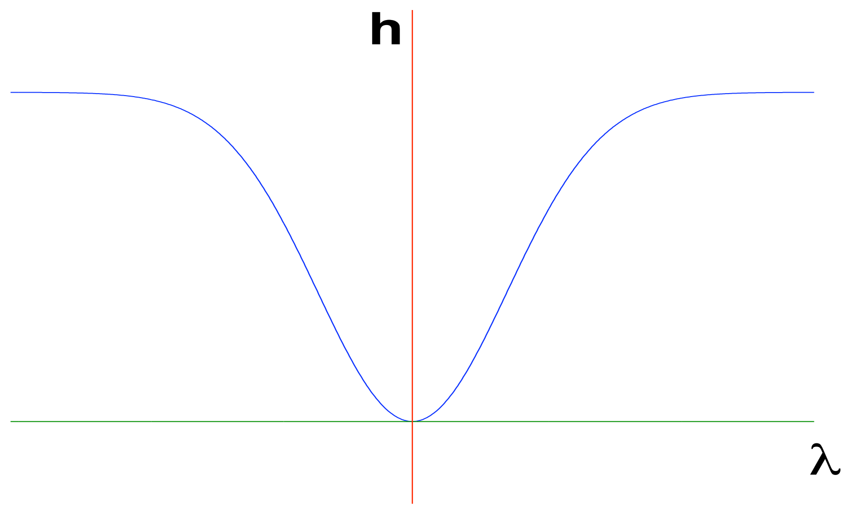
  For example the Crawford-Ferguson (1970) criteria:

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

- Component loss criteria

(Generalized hyperplane count criteria)

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



$$h(\lambda) = \text{component loss function (CLF)}$$

Long history, but sparse.

This is too simple and simply motivated a criterion to have been ignored for so long.

Choosing $h(\lambda)$:

Th: If

h is symmetric and is concave and nondecreasing on $[0, \infty)$

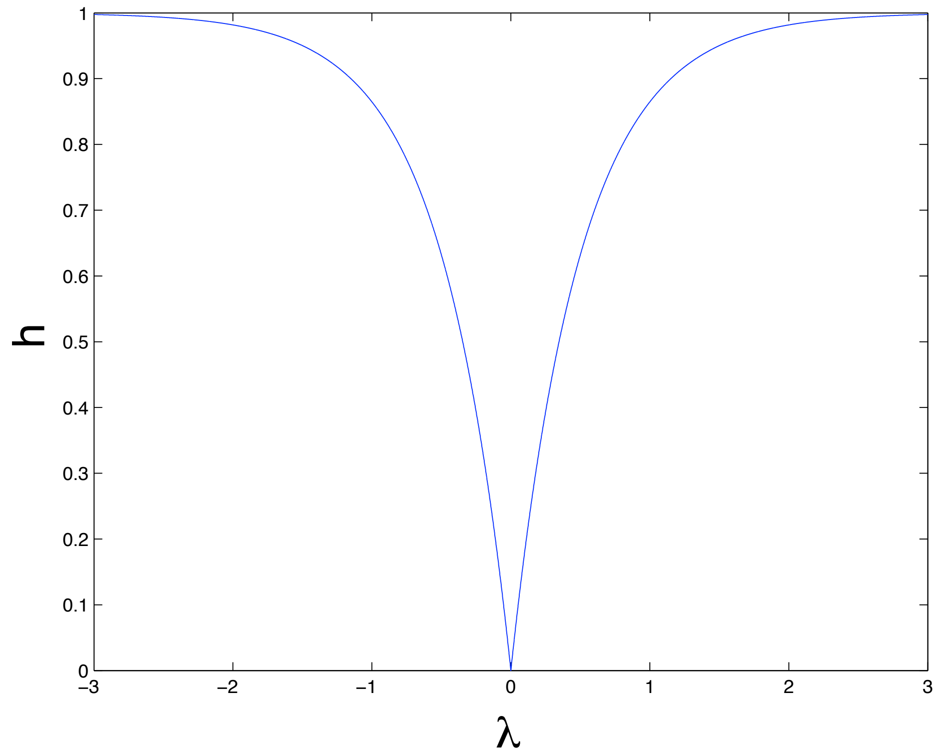
Λ is a rotation of A and has perfect simple structure

then

Λ mimimizes Q

Moreover, if h is strictly concave on $[0, \infty)$, any minimizer of Q must have perfect simple structure.

This means minimizing Q will produce perfect simple structure whenever it exists.



- Criteria related to CLF criteria:

  Browne's partially specified target criterion may be viewed as a weighted CLF criterion.

  Kiers' Simplimax criterion may be viewed as iteratively re-weighted CLF criterion.

  There is a subclass of CLF criteria that have the same local minima as the Simplimax criteria.

- Many other rotation criteria

  Quartimin
  Orthomin
  Geomin
  Simplimax
  Minimum entropy
  Oblimax
  Promax
  Invariant pattern
  Infomax

See Browne's Bible (2001)

Algorithms for Analytic Rotation

Direct oblique methods only

Before the appearance of the following theorem all analytic oblique rotation used indirect methods.

Th: If $Q(\Lambda)$ is invariant under sign changes in the columns of Λ, then when rotating one factor in the plane of two there is a scalar parameter δ such that

$$Q(\Lambda) = Q(\delta)$$

Pf: Jennrich and Sampson (1966) for quartimin.

This led to pairwise algorithms for quartimin and oblimin.

- General pairwise line search algorithms

  Browne & Cudeck in CEFA

  The previous result holds for arbitrary $Q(\Lambda)$

  A line search algorithm to minimize $Q(\delta)$ gives a pairwise algorithm for an arbitrary $Q(\Lambda)$.

  This is a remarkable algorithm:

  (a) It works for any rotation criterion $Q(\Lambda)$

  (b) All that is required is a formula for $Q(\Lambda)$

  (c) It is remarkably simple

  (d) It has been used successfully for many different criteria

  Browne and Cudeck should publish it.

- Gradient projection algorithms

  Jennrich (2002)

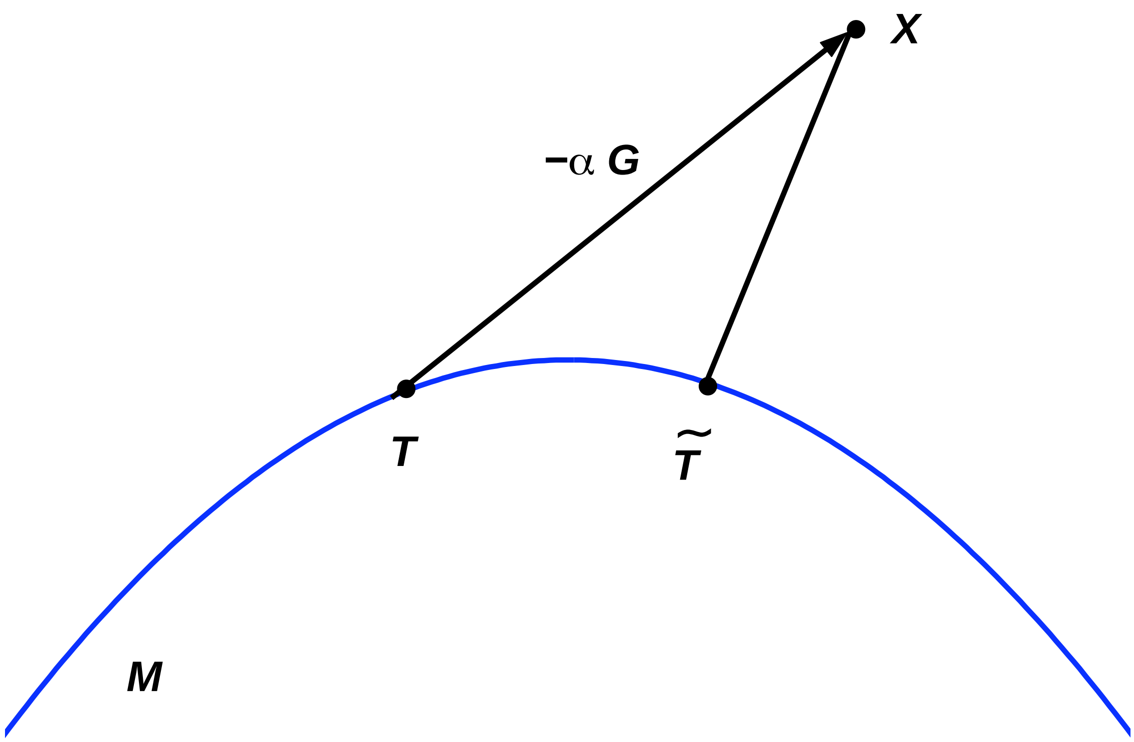
  The oblique rotation problem is to minimize

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

  over all T in the manifold $\mathcal{M}$ of nonsingular T with rows of length one.

  Algorithm: Choose a T in $\mathcal{M}$, a scalar $\alpha > 0$, and compute:

$$\begin{aligned} G &= \text{gradient of } f \text{ at } T \\ X &= T - \alpha G \\ \tilde{T} &= \text{projection of } X \text{ onto } \mathcal{M} \end{aligned}$$



The projection $\tilde{T}$ is simply X with its rows scaled to have length one.

Th: If T is not a stationary point of f restricted to $\mathcal{M}$,

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

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

If necessary halve α until $f(\tilde{T}) < f(T)$
Replace T by $\tilde{T}$ and repeat

This gives a strictly decreasing algorithm.

Obtaining the gradient G is not always easy.

Using numerical gradients leads to almost identical results.

With numerical gradients this algorithm has the same nice properties as Browne and Cudeck without requiring cycling through pairs of columns.

One can find free SAS, SPSS, R/S, and Matlab code at

`http://www.stat.ucla.edu/research/gpa`

Standard Errors for Rotated Loadings

Why standard errors?

Cudeck and O'Dell (1994) give an excellent discussion.

Unlike most statistical analyses, most EFA programs produce no standard errors.

Because they are not a by-product of the fitting process.

CEFA and SAS Factor do

Historical note:

An analytic rotation of $\hat{A}$ has the form

$$\hat{\Lambda} = \hat{A}(\hat{T})^{-1}$$

Because both $\hat{A}$ and $\hat{T}$ are random Lawley and Maxwell (1971) believed that:

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

Wexler (1968) provided some evidence that one cannot always ignore sampling errors in T.

Archer and Jennrich (1973) and Jennrich (1973) showed the Lawley and Maxwell approximation is not needed.

Using the asymptotic distribution of $\hat{A}$

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

Results of this form have been given by:

Anderson and Rubin (1956) for principal component loadings and normal sampling.

Lawley (1967) for canonical loadings and normal sampling.

Joreskog (1969) for confirmatory factor analysis and normal sampling.

Browne (1984) for confirmatory factor analysis and non-normal sampling.

Gershick (1939) for principal component analysis and normal sampling.

The confirmatory approach is particularly attractive. It assumes the upper diagonal part of A is zero. Then an estimate of acov($\hat{a}$) is a byproduct of the confirmatory analysis.

Let "alg" be a rotation algorithm and let

$$\Lambda = \text{alg}(A)$$

In vector form

$$\lambda = h(a)$$

and

$$d\lambda = \frac{dh}{da} da$$

By the delta method

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

- Finding dh/da by implicit differentiation

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

$$\text{dg}(TT') = I$$

$$\text{ndg}(\Lambda' \tfrac{dQ}{d\Lambda}(TT')^{-1}) = 0$$

The idea is to implicitly differentiate these equations. More specifically:

Let

$$f(a,t) = \text{vec}(AT^{-1})$$

$$\varphi(\lambda,t) = \text{vec}(\text{ndg}(\Lambda'\tfrac{dQ}{d\Lambda}(TT')^{-1}) + dg(TT') - I)$$

Then

$$f(a,t) = \lambda$$

$$\varphi(\lambda,t) = 0$$

Implicit differentiation gives

$$\dot{f}_1 da + \dot{f}_2 dt = d\lambda$$

$$\dot{\varphi}_1 d\lambda + \dot{\varphi}_2 dt = 0$$

Solving for $d\lambda$ in terms of da gives

$$\tfrac{dh}{da} = \dot{f}_1 - \dot{f}_2(\dot{\varphi}_1\dot{f}_2 + \dot{\varphi}_2)^{-1}\dot{\varphi}_1\dot{f}_1$$

Specific formulas

$$\dot{f}_1 = I \otimes (T')^{-1}$$

$$\dot{f}_2 = -\Lambda \otimes (T')^{-1}$$

$$\dot{\varphi}_1 = \text{ numerical derivative of } \varphi \text{ w.r.t. } \lambda$$

$$\dot{\varphi}_2 = \text{ numerical derivative of } \varphi \text{ w.r.t. } t$$

This is a fairly nice method:

(a) It is simple

(b) It can be used for any rotation method

- Comments on $dh/d\lambda$ formulas

Jennrich (1973)

Exact derivatives for the generalized
Crawford-Ferguson family

Cudeck & O'Dell FAS (1994)

General criteria
Numerical derivatives
Row normalization

Hayashi & Yung (1999)

Othhomax with row normalization
Exact derivatives
Matrix formulation

Constrained optimization methods

- Maximum likelihood (Jennrich, 1974)

A simple way to get standard errors.

Let

$$S = \text{the sample covariance matrix}$$

$$\lambda = \text{vec}(\Lambda)$$

$$\phi = \text{vecu}(\Phi)$$

$$\psi = \text{vecd}(\Psi)$$

$$\theta = \begin{pmatrix} \lambda \\ \phi \\ \psi \end{pmatrix}$$

$$\ell(\theta) = \text{ likelihood of } \theta \text{ given } S$$

This likelihood is over parameterized

For oblique rotation

$$\text{ndg}(\Lambda' \tfrac{dQ}{d\Lambda} \Phi^{-1}) = 0$$

Write this as

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

Let

$$\dot{\varphi} = \tfrac{d\varphi}{d\theta}$$

$$\mathcal{I} = \text{ information matrix at } \theta$$

Using Silvey's (1971) result on constrained MLE

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

If $\dot{\varphi}$ is computed numerically, this is a very simple approach when using maximum likelihood factor analysis.

25 years later a massive generalization of this approach appeared in CEFA.

- CEFA methods

CEFA is a free very general EFA program by

Browne, Cudeck, Tateneni, and Mels (1999)

`http://quantrm2.psy.ohio-state.edu/browne/`

I think it uses constrained minimum deviance methods.

Until Browne et al publish a paper, however, we have no way of knowing.

Their methods solve many problems including producing standard errors when factoring correlation matrices and when using normalized loadings.

Pseudo value methods

Jennrich & Clarckson (1980)

Let g be any algorithm that estimates Λ from S. More specifically let

$$\hat{\lambda} = g(S)$$

Let

$$\widehat{dg} = \text{ differential of } g \text{ at } S$$

For each $t = 1, \cdots, n$ define a pseudo value

$$\tilde{\lambda}_t = \widehat{dg}((x_t - \bar{x})(x_t - \bar{x})')$$

Let

$$S_{\tilde{\lambda}} = \text{ sample covariance matrix for the } \tilde{\lambda}_t$$

Th: $S_{\tilde{\lambda}} \to \text{acov}(\hat{\lambda})$

(a) This is a non-parametric result.

One can sample from any distribution, use any method of extraction and any method of rotation.

(b) 4-th sample moments of the x_t are not required.

(c) Except for finding $\widehat{dg}$ this is a very simple method.

(d) Jennrich and Clarkson give $\widehat{dg}$ for maximum likelihood extraction and othomax rotation.

(e) This needs to be extended to other extraction and rotation methods.

(f) This is my favorite method.

Less linear standard error methods

Jackkinfe

Bootstrap

MCMC

These require a mapping from $X \rightarrow \widehat{\Lambda}$

But rotations are determined only up to column sign and permutation.

Alignment failures have a disastrous effect.

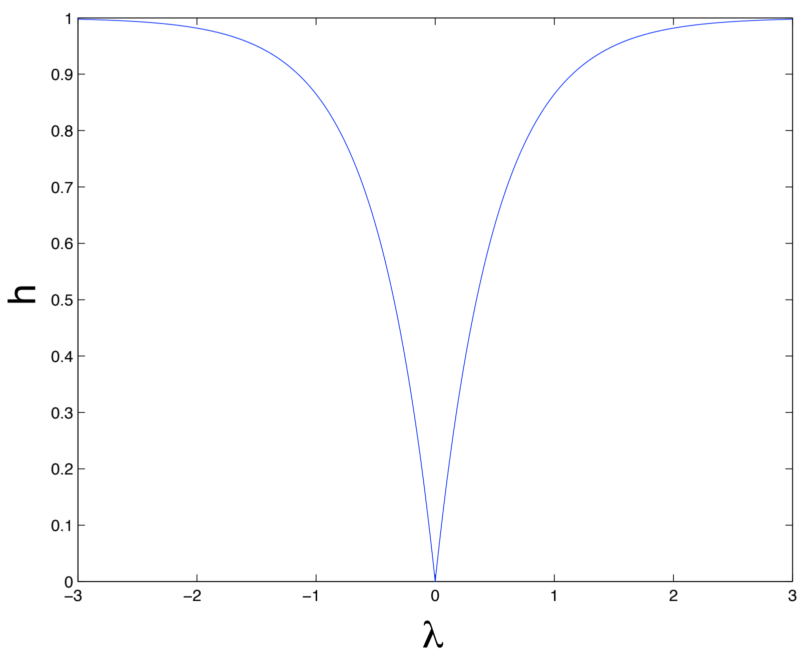
A future direction

A different approach to exploratory factor analysis.

Use penalized deviance for setting zero loadings:

$$F(S, \Sigma) + \gamma Q(\Lambda)$$

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



By increasing γ values of λ will be sucked into zero.

Consider least squares first:

$$\|y - X\beta\|^2 + \gamma \sum_j^p h(\beta_j)$$

This may be thought of as an alternative to step down regression.

Th: # zero $\widehat{\beta}_j \to p$ as $\gamma \to \infty$

A six predictor example:

γ	$\widehat{\beta}_1$	$\widehat{\beta}_2$	$\widehat{\beta}_3$	$\widehat{\beta}_4$	$\widehat{\beta}_5$	$\widehat{\beta}_6$
0	.63	2.01	3.37	3.97	5.13	5.96
50	0	1.28	3.00	3.94	4.93	5.66
100	0	0	2.18	4.12	4.61	4.85
200	0	0	0	3.54	4.45	3.52
400	0	0	0	0	4.55	0
1600	0	0	0	0	0	0

Now pretend the $\widehat{\beta}_j$ are $\widehat{\lambda}_{ir}$ and use this method to set zero loadings.