Factor Analysis as Matrix Decomposition and Approximation: Theory

Jan de Leeuw

Version 96, May 05, 2017

Abstract

A general form of linear factor analysis is defined, and presented as a method to factor a data matrix, similar in many respects to principal component analysis. We discuss necessary and sufficient conditions for solvability of the factor analysis equations and give a constructive method to compute all solutions. A follow up paper will present the corresponding algorithm.

Contents

Note: This is a working paper which will be expanded/updated frequently. All suggestions for improvement are welcome. The directory gifi.stat.ucla.edu/factor has a pdf version, the complete Rmd file with all code chunks, and the bib file.

1 Introduction

Problem: Suppose X is an $n \times m$ real matrix. Suppose $\mathcal{H}$ is a subset of $\mathbb{R}^{p \times p}$ and $\mathcal{S}$ is a subset of $\mathbb{R}^{m \times p}$. We want to find the solutions of the equation $X = FS'$, with the $n \times p$ matrix F of *factor scores* such that $F'F$ is in $\mathcal{H}$ and the $m \times p$ matrix S of *factor loadings* in $\mathcal{S}$.

For reference purposes we repeat our general *factor analysis model*.

$$X = FS',\tag{1}$$
$$F'F = H,\tag{2}$$
$$H \in \mathcal{H} \subseteq \mathbb{R}^{p \times p},\tag{3}$$
$$S \in \mathcal{S} \subseteq \mathbb{R}^{m \times p}.\tag{4}$$

Note that we do not require that $p \leq \min(m, n)$ or even $p = \mathbf{rank}(X)$. The interesting case is $p > m$, more factors than variables. There is also no requirement that $n \geq m$, by the way.

In classical *exploratory orthogonal common factor analysis* we require $H = I$ and $S = \begin{bmatrix} S_1 & | & S_2 \end{bmatrix}$ with the $m \times m$ matrix S_2 diagonal. The $m \times (p - m)$ matrix S_1 has the *common factor loadings*. In *confirmatory common factor analysis* we keep the basic structure of $\mathcal{S}$ but we may fix some loadings at fixed and known values, mostly zero. In non-orthogonal versions of common factor analysis we allow for non-zero elements in H, usually to model correlations between common factors.

In this paper we rederive some of the basic algebraic results for factor analysis. These results, in various forms and disguises, can be found in many places, starting with Wilson (1928) and culminating – at least for me – in Guttman (1955). I desperately needed some clarity after reading much of the factor indeterminacy literature, which – at least for me – has a light-to-heat ratio close to zero. I admire Steiger and Schönemann (1978) and Steiger (1979) for going through half a century of clunky notation, polemics, scientific denial, and doubtful results.

Our proofs are based on powerful matrix decomposition tools, the *singular value decomposition* and the *eigenvalue decomposition*. This makes our proofs quite different from what one normally finds, especially in the older literature. Matrix algebra has come a long way.

2 Two Steps

Instead of directly tackling the problem of solving equations $(1) - (4)$ we proceed in two steps. We first give a trivial necessary condition for solvability, which has been important throughout the history of factor analysis.

Theorem 1: [**Necessary**] If S and F solve $(1) - (2)$ then $X'X = SHS'$.

Proof: Duh. **QED**

As we shall see in a moment, the condition $X'X = SHS'$ is also sufficient for solvability of $(1) - (2)$, and consequently $(1) - (4)$ is solvable if and only if

$$X'X = SHS', \tag{5}$$
$$S \in \mathcal{S}, \tag{6}$$
$$H \in \mathcal{H}. \tag{7}$$

Of course it does not follow that conditions $(5) - (7)$ determine S and H uniquely. Because of the generality of our constraints on S and H it is quite useless to look for identification conditions.

The usual approach in factor analysis theory is to first solve $(5) - (7)$ for S and H, and then find all F for which $X = FS'$ and $F'F = H$. The fact that such an F always exists if $(5) - (7)$ are solvable is sometimes called the *Fundamental Theorem of Factor Analysis* (Kestelman (1952)). The fact that there are usually multiple solutions for F to $X = FS'$ and $F'F = H$ for given S and H that solve $(5) - (7)$ is called the *Factor Score Indeterminacy Problem.*

In factor analysis computation a similar two step process is followed. First an approximate solution to $(5) - (7)$ is computed, using multinormal maximum likelihood or least squares. This gives an $S \in \mathcal{S}$ and an $H \in \mathcal{H}$. Then, in the second step, we compute an approximate solution to $X = FS'$ and $F'F = H$, using S and H from the first step. This is sometimes called *Factor Score Estimation.*

3   Least Squares

In this paper we study only the second step only. We assume we have some S and H that may or may not satisfy $(5) - (7)$. Then find an expression for all F with $F'F = H$ that minimize

$$\sigma(F) = \mathbf{tr} \ (X - FS')'(X - FS'). \tag{8}$$

This is "factor score estimation". We then look into conditions for which the minimum of the least squares loss function σ is zero. Those conditions give us the "fundamental theorem of factor analysis". And finally if the minimum of σ is zero the set of all solutions defines the "factor score indeterminacy problem".

Lemma 1: **[Crossproduct]** If the $p \times p$ positive semi-definite matrix H of rank s has eigenvalue decomposition

$$H = \begin{bmatrix} N & N_\perp \end{bmatrix} \begin{bmatrix} \Phi^2 & \emptyset \\ \emptyset & \emptyset \end{bmatrix} \begin{bmatrix} N' \\ N'_\perp \end{bmatrix}, \tag{9}$$

with the $s \times s$ diagonal matrix Φ^2 positive definite, then the $n \times p$ matrix F satisfies $F'F = H$ if and only if F has singular value decomposition $F = T\Phi N'$, with some $n \times s$ matrix T with $T'T = I$.

Proof: Write F in the form

$$F = \begin{bmatrix} A & B \end{bmatrix} \begin{bmatrix} N' \\ N'_\perp \end{bmatrix}. \tag{10}$$

Then $F'F = H$ if and only if

$$\begin{bmatrix} A'A & A'B \\ B'A & B'B \end{bmatrix} = \begin{bmatrix} \Phi^2 & 0 \\ 0 & 0 \end{bmatrix}.$$

Thus $F'F = H$ if and only if $B = 0$ and $A'A = \Phi^2$, which can be written as $A = T\Phi$ with $T'T = I$. **QED**

Theorem 2: [Least_Squares] Suppose X is an $n \times m$ matrix, S is an $m \times p$ matrix , and H is a positive semi-definite $p \times p$ matrix of rank s with eigenvalue decomposition (9). Define the $n \times s$ matrix $R = XSN\Phi$ and suppose

$$R = \begin{bmatrix} P & P_\perp \end{bmatrix} \begin{bmatrix} \Psi & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} Q' \\ Q'_\perp \end{bmatrix} \tag{11}$$

is a singular value decomposition of R and rank R is r. The minimum of $\mathbf{tr}\ (X - FS')'(X - FS')$ over the $n \times p$ matrices F that satisfy $F'F = H$ is

$$\min_{F'F=H} \mathbf{tr}\ (X - FS')'(X - FS') = \mathbf{tr}\ X'X + \mathbf{tr}\ SHS' - 2\ \mathbf{tr}\ \Psi \tag{12}$$

and the minimum is attained at any F with singular value decomposition

$$F = (PQ' + P_\perp DQ'_\perp)\Phi N', \tag{13}$$

where the $(n - r) \times (s - r)$ matrix D satisfies $D'D = I$ but is otherwise arbitrary.

Proof: It follows from lemma 1 that minimizing the sum of squares is equivalent to maximizing $\mathbf{tr}\ T'R$ over $T'T = I$. Let

$$T = \begin{bmatrix} P & P_\perp \end{bmatrix} \begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{bmatrix} Q' \\ Q'_\perp \end{bmatrix}$$

Then $\mathbf{tr}\ T'R = \mathbf{tr}\ A'\Psi$ while $T'T = I$ when

$$\begin{bmatrix} A'A + C'C & A'B + C'D \\ B'A + D'C & B'B + D'D \end{bmatrix} = \begin{bmatrix} I & 0 \\ 0 & I \end{bmatrix}.$$

The maximum $\mathbf{tr}\ A'\Psi$ over $A'A \lesssim I$ is equal to $\mathbf{tr}\ \Psi$ and is attained (uniquely) for $A = I$, which means that $C = 0$ and $B = 0$. Thus

$$T = PQ' + P_\perp DQ'_\perp,$$

and F has singular value decomposition

$$F = (PQ' + P_\perp DQ'_\perp)\Phi N'$$

where $D'D = I$. **QED**

4

4   Fundamental Theorem

Theorem 3: [**Fundamental**]

1. If $X \in \mathbb{R}^{n \times m}$, $S \in \mathbb{R}^{m \times p}$ and $F \in \mathbb{R}^{n \times p}$ and $H \in \mathbb{R}^{p \times p}$ satisfy $X = FS'$ and $F'F = H$ then $X'X = SHS'$.
2. If $X \in \mathbb{R}^{n \times m}$, $S \in \mathbb{R}^{m \times p}$ and $H \in \mathbb{R}^{p \times p}$ satisfy $X'X = SHS'$ then there is an $F \in \mathbb{R}^{n \times p}$ such that $X = FS'$ and $F'F = H$.

Proof: Part 1 is just a restatement of theorem 1. To prove part 2, we show that the minimum least squares loss is zero if and only if $X'X = SHS'$. If $X'X = SHS'$ then $RR' = XSHS'X' = (XX')^2$. Thus the singular values of R are the same as the singular values of $X'X$ and those of SHS', and from (12) we find that minimum loss is zero. Conversely, if minimum loss is zero there is an F with $X = FS'$ and $F'F = H$ and thus $X'X = SHS'$.
QED

5   Quantifying Indeterminacy

For two different choices D_1 and D_2 of D in (13) we have

$$\Phi^{-1}N'F_1'F_2N\Phi^{-1} = PP' + P_\perp D_1'D_2P_\perp'.$$

Thus the canonical correlations of F_1 and F_2 are the $s-r$ singular values of $D_1'D_2$, in addition to r canonical correlations equal to one. Because $D_1'D_2 \gtrsim -I$ we see that

$$\Phi^{-1}N'F_1'F_2N\Phi^{-1} \gtrsim PP' - P_\perp P_\perp' = I - 2P_\perp P_\perp' = 2PP' - I,$$

and thus $\mathbf{tr}\ \Phi^{-1}N'F_1'F_2N\Phi^{-1} \geq 2r - s$, a result due to Schönemann (1971) in classical exploratory factor analysis. In that case we have $p = s = m + c$, with c the number of common factors, and $r = m$. Thus $2r - s = m - c$ and the average correlation between corresponding factors in two solutions is $(1 - c/m)/(1 + c/m)$.

6   Example

Our example is completely fictional, but it does illustrate some of the computations, even in the case where there are singularities. In factor analytic terminology, it has some correlated common factors and some correlated unique factors. The matris S is

```
##      [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9] [,10]
## [1,]    1    0    1    0    1    0    0    0    0     0
## [2,]    2    0    0    0    0    2    0    0    0     0
## [3,]    3    0    0    0    0    0    3    0    0     0
## [4,]    0   -1    0    0    0    0    0    3    0     0
## [5,]    0   -2    0    0    0    0    0    0    2     0
## [6,]    0   -3    0    1    0    0    0    0    0     1
```

and the matrix H is

```
##       [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9] [,10]
##  [1,]    1   -1    0    0    0    0    0    0    0     0
##  [2,]   -1    1    0    0    0    0    0    0    0     0
##  [3,]    0    0    1    0    0    0    0    0    0     0
##  [4,]    0    0    0    1    0    0    0    0    0     0
##  [5,]    0    0    0    0    1    0    0    0    0     0
##  [6,]    0    0    0    0    0    1    0    0    0     0
##  [7,]    0    0    0    0    0    0    1    0    0     0
##  [8,]    0    0    0    0    0    0    0    1    0     0
##  [9,]    0    0    0    0    0    0    0    0    1    -1
## [10,]    0    0    0    0    0    0    0    0   -1     1
```

Matrix S has rank 6 and H has rank 8 (two eigenvalues equal to two, six eigenvalues equal to one, two eigenvalues equal to zero). Eigenvalue decomposition of H defines N and Φ.

Thus SHS' is

```
##      [,1] [,2] [,3] [,4] [,5] [,6]
## [1,]    3    2    3    1    2    3
## [2,]    2    8    6    2    4    6
## [3,]    3    6   18    3    6    9
## [4,]    1    2    3   10    2    3
## [5,]    2    4    6    2    8    4
## [6,]    3    6    9    3    4   11
```

with a trace equal to 58. We fill a 30×6 matrix X with random normal deviates and compute $R = XSN\Phi$.

The sum of squares of X is 220.2584815037, and the trace norm of R (sum of singular values) is 100.3767382715, which means minimum least squares loss is 77.5050049607.

We can now use (13) to construct F. Choose the 24×2 matrix D by

```
d <- matrix (0, 24, 2)
diag (d) <- 1
```

The corresponding F is

Using the sup-norm we find $\|H - F'F\|_\infty$ equal to 1.2212e-15 and $\mathbf{tr}(X - FS')'(X - FS')$ equal to 77.5050049607.

Now we cheat a bit and simply make an X such that $X'X = SHS'$. We use a 30×30 matrix of standard normals, orthonormalize it with `qr()`, and take the first 6 columns as K and the last 24 as $K_\perp$. Then Λ and L are taken from the eigenvalue decomposition of SHS'.

The sum of squares of X and the trace norm of R are both 58, which means minimum least squares loss is zero. With the same D as before we use (13) to construct F. Again $\|H - F'F\|_\infty$ is equal to 1.3323e-15 and now $\|X - FS'\|_\infty$ is equal to 1.4211e-14.

7 And Now ... For Our Next Act

In the second part of this paper we will discuss an algorithm in R that minimizes (8) over $S \in \mathcal{S}$, $H \in \mathcal{H}$ and F with $F'F = H$. This is a one-step algorithm, in the sense that we construct S, H and F simultaneously. Clearly choice of $\mathcal{S}$ and $\mathcal{H}$ is critical here. In fact, we will modify the problem somewhat so that we minimize

$$\sigma(F, H, S) = \mathbf{tr}(X - FHS')'(X - FHS'),$$

where we require $F'F = I$ and $H \in \mathcal{H}$. Both $\mathcal{S}$ and $\mathcal{H}$ are defined by elementwise box constraints, which include the extreme cases of no constraint and constrained to be a fixed real number.

References

Guttman, L. 1955. "The Determinacy of Factor Score Matrices with Implications for Five Other Problems of Common Factor Theory." *The British Journal of Statistical Psychology* 8 (2): 65–81.

Kestelman, H. 1952. "The Fundamental Equation of Factor Analysis." *British Journal of Psychology, Statistical Section* 5: 1–6.

Schönemann, P.H. 1971. "The Minimum Average Correlation between Equivalent Sets of Uncorrelated Factors." *Psychometrika* 36: 21–30.

Steiger, J.H. 1979. "Factor Indeterminacy in the 1930's and the 1970's. Some Interesting Parallels." *Psychometrika* 44 (4): 157–67.

Steiger, J.H., and P.H. Schönemann. 1978. "A History of Factor Indeterminacy." In *Theory Construction and Data Analysis in the Behavioral Sciences*, edited by S. Shye. San Francisco: Jossey-Bass.

Wilson, E.B. 1928. "Reviewed Work(s): The Abilities of Man, Their Nature and Measurement by C. Spearman." *Science, New Series* 67 (1731): 244–48.