The Gauss-Newton Algorithm for the Weighted Least Squares Factor Analysis

Author(s): Sik Yum Lee

Source: *Journal of the Royal Statistical Society. Series D (The Statistician)*, Jun., 1978, Vol. 27, No. 2 (Jun., 1978), pp. 103-114

Published by: Wiley for the Royal Statistical Society

Stable URL: https://www.jstor.org/stable/2987906

The Gauss–Newton Algorithm for the Weighted Least Squares Factor Analysis

SIK YUM LEE
Department of Mathematics, University Science Centre, The Chinese University of Hong Kong, Hong Kong

A computational method, the Gauss–Newton algorithm, in obtaining the Weighted Least Squares estimates of the unknown parameters in the factor analytic model is presented. It is demonstrated that for both the exploratory and confirmatory factor analysis, this algorithm can successfully produce the estimates and the covariance matrix estimate of the estimators. The partial derivatives required for implementing the algorithm are obtained by matrix calculus.

1. Introduction

The factor analysis model such as given by Lawley and Maxwell (1963) states that a p by 1 vector $\boldsymbol{x}$ has a multivariate normal distribution with mean vector $\boldsymbol{u}$ and covariance matrix $\boldsymbol{C}=\boldsymbol{F}\boldsymbol{F}'+\boldsymbol{D}$. The matrix $\boldsymbol{F}$ is a p by k with $k<p$ and the matrix $\boldsymbol{D}$ is diagonal. One of the main problems in factor analysis is to estimate the vector $\boldsymbol{\theta}$, which consists of the unknown parameters in $\boldsymbol{F}$ and $\boldsymbol{D}$. For exploratory factor analysis, the Weighted Least Squares and Maximum Likelihood estimators have been obtained by Lawley and Maxwell (1963), Jöreskog (1967), Jennrich and Robinson (1969), Clarke (1970) and Jöreskog and Goldberger (1972). They either used the Fletcher–Powell algorithm or the Newton–Raphson algorithm to get the estimates. In confirmatory factor analysis, the investigator has knowledge about some of the parameters in the model and tries to estimate the remaining unknown parameters. Jöreskog (1969) has successfully applied the Fletcher–Powell algorithm to produce the Maximum Likelihood estimates. In this paper, we discuss the application of the Gauss–Newton algorithm to obtain the Weighted Least Squares estimates. The method discussed can be used for both the exploratory and confirmatory factor analysis. In subsequent sections, we shall use the following notations and abbreviations:

$\boldsymbol{I}_p$	the p by p identity matrix,
$\boldsymbol{X}'$ or $\boldsymbol{X}^{\mathrm{T}}$	the transpose of $\boldsymbol{X}$,
vec $\boldsymbol{X}$ or $\underline{\boldsymbol{X}}$	an mn by 1 vector which takes all the elements of the m by n matrix $\boldsymbol{X}$ row by row and places them in a vector,
$\boldsymbol{X}_{\mathrm{sym}}$	an $n(n+1)/2$ by 1 vector which takes the lower triangle

	elements (including the diagonal elements) of the n by n matrix $\boldsymbol{X}$ row by row and places them into a vector,
$\boldsymbol{X}_{\mathrm{dg}}$	an n by 1 vector whose components are the diagonal elements of the n by n matrix $\boldsymbol{X}$,
$\boldsymbol{X}^*\boldsymbol{Y}$	the Hadamard product of matrices $\boldsymbol{X}$ and $\boldsymbol{Y}$ of the same order, that is, $\boldsymbol{X}^*\boldsymbol{Y}=(x_{ij}y_{ij})$,
$\boldsymbol{X}@\boldsymbol{Y}$	the right Kronecker product of matrices $\boldsymbol{X}$ and $\boldsymbol{Y}$ of the same order, that is, $\boldsymbol{X}@\boldsymbol{Y}=(x_{ij}\boldsymbol{Y})$,
WLS	Weighted Least Squares,
GN	Gauss–Newton,
RMS	Root mean squares.

2. Discussion of WLS Problem and Preliminary Mathematical Results

Suppose $\boldsymbol{C}_0=\boldsymbol{C}(\theta_0)$ is a p by p covariance matrix, which is a function of an unknown q by 1 parameter vector $\boldsymbol{\theta}_0$. In general, we will regard the elements in $\boldsymbol{\theta}$ as mathematical variables. Let x_k, $k=1, \ldots, n+1$ be independently and multivariate normally distributed with mean vector $\boldsymbol{u}$ and covariance matrix $\boldsymbol{C}$; and $\boldsymbol{S}$ be the sample covariance matrix. Then $\boldsymbol{S}$ has the wishart distribution with expectation $\boldsymbol{C}$ and covariance matrix $2(\boldsymbol{C}@\boldsymbol{C})/n$. Consider the residual quadratic form

$$(\boldsymbol{S}_{\mathrm{sym}}-\boldsymbol{C}_{\mathrm{sym}})' \operatorname{cov}(\boldsymbol{S}_{\mathrm{sym}})^{-1}(\boldsymbol{S}_{\mathrm{sym}}-\boldsymbol{C}_{\mathrm{sym}})$$

By Browne (1974), this quadratic form equals to $(n/2)\operatorname{tr}\{(\boldsymbol{S}-\boldsymbol{C})\boldsymbol{C}_0^{-1}\}/2$. Since $\boldsymbol{C}_0^{-1}$ is unknown, we replace it by $\boldsymbol{S}^{-1}$ and consider the WLS function

$$\boldsymbol{Q}(\theta)=\operatorname{tr}(\boldsymbol{I}_p-\boldsymbol{C}\boldsymbol{S}^{-1})^2/2 \tag{2.1}$$

By definition, the WLS estimator $\hat{\boldsymbol{\theta}}_{\mathrm{WLS}}$ is the vector $\boldsymbol{\theta}$ such that $\boldsymbol{Q}(\theta)$ is minimized provided such a $\boldsymbol{\theta}$ exist. Under mild regularity conditions, Brown (1974) and Lee (1977) showed that the WLS estimator process the following asymptotic properties:

(a) The WLS estimator $\hat{\boldsymbol{\theta}}_{\mathrm{WLS}}$ is consistent.

(b) The WLS estimator $\hat{\boldsymbol{\theta}}_{\mathrm{WLS}}$ is asymptotically equivalent to the Maximum Likelihood estimator.

(c) The asymptotic distribution of $\hat{\boldsymbol{\theta}}_{\mathrm{WLS}}$ is multivariate normal with mean vector θ_0 and covariance matrix $2n^{-1}\boldsymbol{V}(\boldsymbol{C}_0^{-1})^{-1}$, where

$$\boldsymbol{V}(\boldsymbol{C}_0^{-1})=\frac{\partial \boldsymbol{C}}{\partial \boldsymbol{\theta}}(\boldsymbol{C}_0^{-1}@\boldsymbol{C}_0^{-1})\frac{\partial \boldsymbol{C}^{\mathrm{T}}}{\partial \theta}\bigg|_{\theta=\theta_0}$$

The matrix $2\boldsymbol{V}(\boldsymbol{C}_0^{-1})$ is the information matrix evaluated at $\boldsymbol{\theta}_0$.

(d) The asymptotic distribution of $n\boldsymbol{Q}(\hat{\boldsymbol{\theta}}_{\mathrm{WLS}})$ is χ^2 with degree of freedom $2^{-1}p(p+1)-q$.

In order to implement the GN algorithm, the first partial order derivatives of $\boldsymbol{Q}(\theta)$ with respect to $\boldsymbol{\theta}$ are required. The mathematical preliminary for deriving these derivatives are presented in the following definitions and theorems. The proof of the theorems can be found in Lee (1977).

Definition Given a p by q matrix $\boldsymbol{Y}$ whose elements are differentiable functions of elements of an m by n matrix $\boldsymbol{X}$, the matrix derivative $\partial \boldsymbol{Y}/\partial \boldsymbol{X}$ is equal to the mn by pq matrix whose (rs, uv) element is given by $\partial y_{rs}/\partial x_{uv}$, for $1 \leqslant r \leqslant p$, $1 \leqslant s \leqslant q$, $1 \leqslant u \leqslant m$ and $1 \leqslant v \leqslant n$.

In this paper, unless otherwise stated, we assume that all matrix derivatives exist. Obviously we have $\partial \text{ vec } \boldsymbol{Y}/\partial \boldsymbol{X} = \partial \boldsymbol{Y}/\partial \boldsymbol{X}$ by definition.

Suppose $\boldsymbol{X}$ is an m by n matrix, we denote $\boldsymbol{E}_{mn} = \partial \boldsymbol{X}'/\partial \boldsymbol{X}$. $\boldsymbol{E}_{mn}$ is a mn by mn matrix, whose general element is

$$(\boldsymbol{E}_{mn})_{n(j-1)+k,\ m(j'-1)+k'} = \begin{cases} 1 \text{ if } j = k' \text{ and } k = j' \\ 0 \text{ otherwise} \end{cases}$$

where $0 < j, k' \leqslant m$, and $0 < j', k \leqslant n$.

Also, we denote $\boldsymbol{K}_n = \partial \boldsymbol{X}/\partial \boldsymbol{X}_{\text{dg}}$, $\boldsymbol{K}_n$ is an n by n^2 matrix, whose general element is

$$(\boldsymbol{K}_n)_{ij} = \begin{cases} 1 \text{ if } j = n(i-1)+i \\ 0 \text{ otherwise} \end{cases}$$

where $0 < i \leqslant n$.

Theorem Suppose $\boldsymbol{A}$ is an m by n matrix, $\boldsymbol{B}$ is a p by q matrix, $\boldsymbol{C}$ is an n by r matrix, $\boldsymbol{D}$ is a q by s matrix, $\boldsymbol{F}$ is an n by q matrix, and $\boldsymbol{H}_1$ and $\boldsymbol{H}_2$ are non-singular matrices, then

$$(\boldsymbol{A} @ \boldsymbol{B})(\boldsymbol{C} @ \boldsymbol{D}) = \boldsymbol{AC} @ \boldsymbol{BD} \tag{2.2}$$

$$(\boldsymbol{A} @ \boldsymbol{B})' = \boldsymbol{A}' @ \boldsymbol{B}' \tag{2.3}$$

$$(\boldsymbol{A} @ \boldsymbol{B}) \text{ vec } \boldsymbol{F} = \text{vec } (\boldsymbol{AFB}') \tag{2.4}$$

$$(\boldsymbol{H}_1 @ \boldsymbol{H}_2)^{-1} = \boldsymbol{H}_1^{-1} @ \boldsymbol{H}_2^{-1} \tag{2.5}$$

Theorem (Chain Rule) If the elements of a p by q matrix $\boldsymbol{Z}$ are differentiable functions of an r by s matrix $\boldsymbol{Y}$, and the elements of $\boldsymbol{Y}$ are differentiable functions of the elements of an m by n matrix $\boldsymbol{X}$, then

$$\frac{\partial \boldsymbol{Z}}{\partial \boldsymbol{X}} = \frac{\partial \boldsymbol{Y}}{\partial \boldsymbol{X}} \frac{\partial \boldsymbol{Z}}{\partial \boldsymbol{Y}} \tag{2.6}$$

Theorem (Product Rule): Suppose each element of a p by r matrix $\boldsymbol{Y}$ and an r by q matrix $\boldsymbol{Z}$ is a differentiable function of the elements of an m by n matrix $\boldsymbol{X}$. Then

$$\frac{\partial (\boldsymbol{YZ})}{\partial \boldsymbol{X}} = \frac{\partial \boldsymbol{Y}}{\partial \boldsymbol{X}} (\boldsymbol{I}_p @ \boldsymbol{Z}) + \frac{\partial \boldsymbol{Z}}{\partial \boldsymbol{X}} (\boldsymbol{Y}' @ \boldsymbol{I}_q) \tag{2.7}$$

Theorem Let X be a p by p matrix, A and B are constant matrices of dimensions q by p and p by r respectively, then

$$\frac{\partial AXB}{\partial X} = A' @ B \tag{2.8}$$

$$\frac{\partial X^2}{\partial X} = (I_p @ X) + (X' @ I_p) \tag{2.9}$$

and

$$\frac{\partial \operatorname{tr} X}{\partial X} = \operatorname{vec} I_p \tag{2.10}$$

3. GN Algorithm on General WLS Function

The WLS estimator is determined as the solution of $\partial Q/\partial \theta = 0$. These equations cannot be solved algebraically, so that some iterative procedure has to be used instead. We first consider the first partial derivatives of $Q(\theta)$ with respect to θ. By (2.6), (2.9), (2.10), (2.2) and (2.4),

$$\begin{aligned}
\frac{\partial Q}{\partial \theta} &= \frac{1}{2} \frac{\partial \operatorname{tr} (I_p - CS^{-1})^2}{\partial \theta} \\
&= \frac{1}{2} \frac{\partial (I_p - CS^{-1})}{\partial \theta} \frac{\partial (I_p - CS^{-1})^2}{\partial (I_p - CS^{-1})} \frac{\partial \operatorname{tr} (I_p - CS^{-1})^2}{\partial (I_p - CS^{-1})^2} \\
&= \frac{-1}{2} \frac{\partial C}{\partial \theta} (I_p @ S^{-1}) \{ I_p @ (I_p - CS^{-1}) + (I_p - S^{-1}C) @ I_p \} \operatorname{vec} I_p \\
&= \frac{-1}{2} \frac{\partial C}{\partial \theta} [\{ I_p @ S^{-1}(S - C)S^{-1} \} + \{ S^{-1}(S - C)S^{-1} @ I_p \}] \operatorname{vec} I_p \\
&= \frac{1}{2} \frac{\partial C}{\partial \theta} 2 \operatorname{vec} \{ S^{-1}(S - C)S^{-1} \} \\
&= -\frac{\partial C}{\partial \theta} \operatorname{vec} \{ S^{-1}(S - C)S^{-1} \}
\end{aligned} \tag{3.1}$$

or

$$= -\frac{\partial C}{\partial \theta} (S^{-1} @ S^{-1})(\underline{S} - \underline{C}) \tag{3.2}$$

Differentiating $\partial Q/\partial \theta$ once more with respect to θ, by (2.7) and (2.8),

$$\begin{aligned}
\frac{\partial^2 Q}{\partial \theta \, \partial \theta} &= \frac{\partial}{\partial \theta} \left\{ -\frac{\partial C}{\partial \theta} (S^{-1} @ S^{-1})(\underline{S} - \underline{C}) \right\} \\
&= \frac{\partial^2 C}{\partial \theta \, \partial \theta} \{ I_q @ (S^{-1} @ S^{-1})(\underline{S} - \underline{C}) \} \\
&\quad - \frac{\partial (S^{-1} @ S^{-1})(\underline{S} - \underline{C})}{\partial \theta} \left(\frac{\partial C^{\mathrm{T}}}{\partial \theta} @ 1 \right) \\
&= \frac{\partial C}{\partial \theta} (S^{-1} @ S^{-1}) \frac{\partial C^{\mathrm{T}}}{\partial \theta} - \frac{\partial^2 C}{\partial \theta \, \partial \theta} \{ I_q @ (S^{-1} @ S^{-1})(\underline{S} - \underline{C}) \}
\end{aligned} \tag{3.3}$$

If θ_i is the ith approximation of $\boldsymbol{\theta}$ and the matrix $\partial^2 Q/\partial\theta\,\partial\theta$ is positive definite, then the Newton–Raphson algorithm is defined by

$$\theta_{i+1}=\theta_i-\left(\frac{\partial^2 Q}{\partial\theta\,\partial\theta}\right)^{-1}\frac{\partial Q}{\partial\theta}\bigg|_{\theta=\theta_i}$$

The second term of (3.3) consists $S-C$ as a factor. Since $S-C$ will be small, particularly near the minimum, it is dropped in the GN algorithm. Thus, from (3.2) and (3.3), the ith iteration of the GN algorithm is equal to

$$\theta_{i+1}=\theta_i+\Delta\theta_i=\theta_i-G_i^{-1}\left(\frac{\partial Q}{\partial\boldsymbol{\theta}}\right)\bigg|_{\theta=\theta_i} \tag{3.4}$$

where

$$G_i=\frac{\partial C}{\partial\boldsymbol{\theta}}(S^{-1}@S^{-1})\frac{\partial C^{\mathrm{T}}}{\partial\boldsymbol{\theta}}\bigg|_{\theta=\theta_i} \tag{3.5}$$

This algorithm may fail when it is not modified. A useful modification is the "step-halving", which chooses $\Delta\theta_i$, $\Delta\theta_i/2$, . . . , until a step is found which reduces $Q(\theta)$. There are many advantages in the GN algorithm. It is usually robust to bad starting values; it requires only the first order partial derivatives; it works well in a broad spectrum of problems and in comparative studies (e.g. Bard, 1970); and it can handle the problem of singularity very well (e.g. Greenstadt, 1967; Jennrich and Sampson, 1968; Marquardt, 1963).

After $\hat{\boldsymbol{\theta}}_{\mathrm{WLS}}$ has been obtained, its covariance matrix, $\widehat{\mathrm{cov}}\,(\hat{\boldsymbol{\theta}}_{\mathrm{WLS}})$ is obtained by taking $2/n$ times the matrix $G(\hat{\boldsymbol{\theta}}_{\mathrm{WLS}})^{-1}$. It has been shown by Lee (1977) that under mild regularity conditions,

$$n\,\widehat{\mathrm{cov}}\,(\hat{\boldsymbol{\theta}}_{\mathrm{WLS}})\xrightarrow{(p)}2V(C_0^{-1})^{-1} \tag{3.6}$$

Hence the GN algorithm automatically yields $\widehat{\mathrm{cov}}\,(\hat{\boldsymbol{\theta}}_{\mathrm{WLS}})$, which is a consistent estimator of the inverse of the information matrix. Thus, the estimator obtained is asymptotically efficient. The standard error estimates are obtained by taking the square roots of the diagonal elements of $\widehat{\mathrm{cos}}\,(\hat{\boldsymbol{\theta}}_{\mathrm{WLS}})$. These standard error estimates are important for testing hypothesis and setting confidence interval of $\boldsymbol{\theta}$.

4. GN Algorithm on WLS Factor Analysis

First we derive expressions for $\partial Q/\partial\boldsymbol{\theta}$ and G in (3.4). In factor analysis, $C=FF'+D$ and $\boldsymbol{\theta}=(\mathrm{vec}\,F,\ D_{\mathrm{dg}})$. By (2.7)

$$\begin{aligned}\frac{\partial C}{\partial\boldsymbol{\theta}}&=\begin{bmatrix}\partial C/\partial F\\ \partial C/\partial D_{\mathrm{dg}}\end{bmatrix}\\ &=\begin{bmatrix}(I_p@F')+E_{pk}(F'@I_p)\\ K_p\end{bmatrix}\end{aligned} \tag{4.1}$$

Thus, from (3.2) and (4.1),

$$\frac{\partial Q}{\partial \theta} = -\begin{bmatrix} 2 \text{ vec } \{S^{-1}(S-C)S^{-1}F'\} \\ K_p \text{ vec } \{S^{-1}(S-C)S^{-1}\} \end{bmatrix} \tag{4.2}$$

Furthermore, from (3.5) and (4.1),

$$G = \begin{bmatrix} 2\{(S^{-1} @ F'S^{-1}F) + E_{pk}(F'S^{-1} @ S^{-1}F)\} & 2(S^{-1} @ F'S^{-1})K_p{}^{\mathrm{T}} \\ 2K_p(S^{-1} @ S^{-1}F) & S^{-1} * S^{-1} \end{bmatrix} \tag{4.3}$$

It is well known that there is indeterminacy in the model, because if we replace F by FT, where T is some orthogonal transformation, the consequent model is unchanged. In order to remove the indeterminacy, we assign fixed values to some parameters. In confirmatory factor analysis, we usually have some ideas about the assigned values and positions of the fixed parameters in the model. In exploratory factor analysis, as suggested by Anderson and Rubin (1956), we can fix the upper triangular portion of F, above the element F_{ii}, to be some assigned values. Since the fixed parameters are no longer unknown variables, the corresponding elements in the gradient vector, and corresponding rows and columns in matrix G are deleted.

We generate the starting values from the sample covariance or correlation matrix $S=(s_{ij})$, by means of the spectral decomposition. Basically let

$$S = \sum_{i=1}^{p} e_i v_i v_i'$$

be the spectral decomposition of S, where $e_1 \geqslant \ldots \geqslant e_p$ are the eigenvalues of S and $v_1, \ldots, v_p$ are the corresponding eigenvectors. For each i, let

$$F_i^* = \sqrt{e_i}\, v_i.$$

We take

$$F_1 = (F_1^*, \ldots, F_k^*)$$

and

$$(D_1)_{ii} = s_{ii} - \sum_{j=1}^{k} F_{ij}^{*2}$$

to be our initial starting values.

Given these starting values, the procedure of the GN algorithm is: start with F_1 and D_1, set $Q(\theta_0)=10^6$, $\epsilon = 10^{-4}$, and $k=1$.

Step 1: Compute C_k, and $Q(\theta_k)$.

Step 2: If $Q(\theta_k) \leqslant Q(\theta_{k-1})$ go to step 3.
Otherwise, compute $\Delta\theta_{k-1} = \Delta\theta_{k-1}/2$. If RMS $\Delta\theta_k \leqslant \epsilon$, go to step 6; otherwise, set $\theta_k = \theta_{k-1} + \Delta\theta_{k-1}$, and go to step 1.

Step 3: Compute $\partial Q/\partial \theta \mid_{\theta=\theta_k}$ by (4.2).

Step 4: If RMS $\partial Q/\partial \theta \mid_{\theta=\theta_k} \leqslant \epsilon$, go to step 6.
Otherwise compute $G(\theta_k)$ by (4.3).

Step 5: Compute $\Delta\theta_k$ and $\theta_{k+1}=\theta_k+\Delta\theta_k$; update k and return to step 1.

Step 6: Print output.

5. Examples and Discussions

The GN algorithm has been implemented and run on several examples in the literature. For every example, it works nicely. One of the examples, the Holzinger and Swineford's (1939) data is discussed here. These data consist of nine mental ability tests; they are visual perception, cubes, lozenges, paragraph comprehension, sentence completion, word meaning, addition, counting dots, and straight-curved capitals. They can be regarded as measures of essentially visualization, verbal intelligence and mathematical ability. Data were obtained on seventh- and eighth-grade children from the Grant-White school. The correlation matrix based on 145 observations is taken directly from Jöreskog (1969). This correlation matrix and the starting values are presented in Table 1. The algorithm can

Table 1

(a) *The correlation matrix of Holzinger and Swineford's (1939) data*

1·000								
0·318	1·000							
0·436	0·419	1·000						
0·335	0·243	0·323	1·000					
0·304	0·157	0·283	0·722	1·000				
0·326	0·195	0·350	0·714	0·685	1·000			
0·116	0·057	0·056	0·203	0·246	0·170	1·000		
0·314	0·145	0·220	0·095	0·181	0·113	0·585	1·000	
0·489	0·239	0·361	0·309	0·345	0·280	0·408	0·512	1·000

(b) *The starting values for the Holzinger and Swineford data*

F			*D*
0·628	−0·101	0·000*	0·426
0·449	0·008	0·611	0·424
0·608	0·044	0·525	0·353
0·748	0·000*	0·000*	0·179
0·744	0·366	−0·341	0·196
0·730	0·459	−0·223	0·207
0·455	−0·580	−0·469	0·238
0·510	−0·712	−0·128	0·217
0·691	−0·413	0·045	0·351

* Fixed parameter.

produce the solution in 13 iterations; the convergence summary is displaced in Table 2. The convergence is quite rapid and the solution can be determined to about three decimal places. The final solution and the standard error estimates are presented in Tables 3 and 4. Jöreskog (1970) analysed

Table 2

Convergence of the GN algorithm

Iteration	*Function value*	RMS $\partial Q/\partial\theta$	RMS $\Delta\theta$
1	0·97128	0·4951	0·0584
2	0·87718	0·4886	0·0788
3	0·61982	0·4596	0·1553
4	0·13070	0·1446	0·1133
5	0·06823	0·0244	0·0454
6	0·06344	0·0073	0·0121
7	0·06276	0·0032	0·0058
8	0·06259	0·0016	0·0027
9	0·06254	0·0008	0·0014
10	0·06253	0·0004	0·0008
11	0·06253	0·0002	0·0004
12	0·06253	0·0001	0·0002
13	0·06253	0·0000	0·0001

Table 3

The WLS solution of the Holzinger and Swineford data

$\hat{F}$			$\hat{D}$
0·377	0·593	0·000*	0·486
0·240	0·445	−0·105	0·693
0·377	0·569	−0·121	0·497
0·874	0·000*	0·000*	0·232
0·830	−0·013	0·111	0·293
0·823	0·029	−0·016	0·316
0·217	0·102	0·768	0·342
0·124	0·445	0·668	0·337
0·354	0·524	0·368	0·447

these data by the Maximum Likelihood method, his solution is given in Table 5. Jöreskog and Goldberger (1972) showed that in exploratory factor analysis, the WLS and the Maximum Likelihood estimates are very close. Here, we see that for the confirmatory factor analysis, the WLS and Maximum Likelihood estimates are still very close to each other.

Table 4

The standard error estimates of $\hat{\theta}_{WLS}$ for the Holzinger and Swineford data

$\widehat{\text{Std}}(\hat{F})$			$\widehat{\text{Std}}(\hat{D})$
0·088	0·089	*	0·086
0·089	0·094	0·102	0·100
0·088	0·091	0·105	0·093
0·069	*	*	0·051
0·071	0·078	0·070	0·053
0·071	0·075	0·070	0·053
0·092	0·125	0·101	0·121
0·093	0·117	0·102	0·094
0·089	0·098	0·089	0·071

Table 5

Jöreskog's Maximum Likelihood solution for the Holzinger and Swineford's data

$\hat{F}$			$\hat{D}$
0·39	0·59	0·00*	0·50
0·24	0·44	−0·07	0·74
0·38	0·56	−0·08	0·54
0·87	0·00*	0·00*	0·24
0·83	−0·01	0·12	0·30
0·82	0·01	−0·01	0·32
0·22	0·08	0·75	0·39
0·14	0·42	0·70	0·32
0·36	0·51	0·39	0·46

Moreover, the algorithm has been applied to a larger artificial example, which consists of a known 20×20 diagonal matrix D and a known 20×5 matrix F. Their values are given in Table 6. From these matrices, we obtained the covariance matrix $C = FF' + D$. A sample covariance matrix is generated with sample size equal 451.† This sample covariance matrix is then analysed with five factors as before. In this example, excluding the fixed elements, which are the upper triangular elements of F, we are estimating 110 unknown parameters in the model. The algorithm converges in 11 iterations, the convergence summary is presented in Table 7. The final solutions, which are presented in Table 8, are pretty

† The sample covariance matrix was obtained from the "WSHRT" subroutine in HSCF library in UCLA.

Table 6

*The values of **F** and **D** for the artificial example*

F					**D**
1	0	0	0	0	1
0	1	0	0	0	1
0	0	1	0	0	1
0	0	0	1	0	1
0	0	0	0	1	1
1	0	0	0	0	1
0	1	0	0	0	1
0	0	1	0	0	1
0	0	0	1	0	1
0	0	0	0	1	1
1	0	0	0	0	1
0	1	0	0	0	1
0	0	1	0	0	1
0	0	0	1	0	1
0	0	0	0	1	1
1	0	0	0	0	1
0	1	0	0	0	1
0	0	1	0	0	1
0	0	0	0	0	1
0	0	0	0	0	1

Table 7

Convergence of the GN algorithm on the artificial example

Iteration	*Function value*	RMS $\partial Q/\partial\theta$	RMS $\Delta\theta$
1	1·89919	0·2523	0·1208
2	1·74748	0·2675	0·0967
3	1·52040	0·2811	0·1350
4	1·29562	0·2061	0·2611
5	0·32278	0·0425	0·1231
6	0·25115	0·0083	0·0417
7	0·24672	0·0027	0·0093
8	0·24624	0·0011	0·0027
9	0·24617	0·0004	0·0010
10	0·24616	0·0002	0·0004
11	0·24616	0·0000	0·0002

close to the actual values. For the 110 parameters, Table 9 displays the difference between the solution and the actual values in terms of the standard error estimates obtained from the GN algorithm.

We have shown that the GN algorithm is able to produce the WLS estimates and the standard error estimates. During each iteration, the

Table 8

The WLS solution of the artificial example

$\hat{F}$					$\hat{D}$
1·058	0*	0*	0*	0*	0·959
0·068	1·030	0*	0*	0*	0·989
0·110	−0·070	0·992	0*	0*	1·044
0·125	0·025	0·043	1·235	0*	0·940
−0·028	−0·033	0·060	−0·049	1·101	1·026
0·893	0·010	0·016	−0·143	−0·073	1·144
−0·021	0·966	−0·051	0·068	0·066	1·003
0·031	−0·033	0·963	−0·082	−0·092	0·918
0·199	−0·028	0·057	0·943	0·099	1·030
0·000	−0·006	−0·095	−0·015	0·994	0·999
1·041	−0·006	−0·034	−0·092	−0·014	0·833
0·043	1·000	−0·065	0·008	−0·008	0·829
0·008	0·031	1·090	0·043	−0·064	0·986
0·072	0·004	0·081	0·987	0·042	1·088
0·076	0·023	−0·026	−0·040	1·061	1·007
0·996	−0·005	−0·043	−0·100	−0·029	0·893
0·084	1·038	−0·063	0·014	0·088	0·867
0·041	0·048	1·091	−0·069	−0·009	0·800
−0·053	0·015	−0·006	0·066	−0·050	0·928
0·007	0·044	0·065	0·060	0·013	0·873

Table 9

Comparison of the GN solution with the actual values of the artificial example

	$\lvert\theta_i-\hat{\theta}_i\rvert<\hat{\sigma}_i$	$\hat{\sigma}_i<\lvert\theta_i-\hat{\theta}_i\rvert<2\hat{\sigma}_i$	$2\hat{\sigma}_i<\lvert\theta_i-\hat{\theta}_i\rvert$
Number	83	22	5
Percentage	75·4	20	4·6

algorithm must compute the gradient vector, the matrix $\boldsymbol{G}_i$ and its inverse. Most of the computer time in the algorithm is used to invert the matrix $\boldsymbol{G}_i$; so a good matrix inversion subroutine† is important in this algorithm. Using an IBM 360/91 model, in the Holzinger and Swineford example, it takes 1·07 s to converge, while in the larger artificial example, it takes 7·92 seconds to converge.

† Using an IBM 360/91 model, our subroutine inverts a 100×100 matrix within 0·5 s.

REFERENCES

BARD, Y. (1970). Comparison of gradient methods for the solution of non-linear parameter estimation problem. *SIAM Journal on Numerical Analysis*, **7**, 157–86.

CLARKE, M. R. B. (1970). A rapidly convergent method for maximum-likelihood factor analysis. *British Journal of Mathematical and Statistical Psychology*, **23**, 43–52.

GREENSTADT, J. (1967). On the relative efficiencies of gradient methods. *Mathematical Computing*, **21**, 360–7.

HOLZINGER, K. J. and SWINEFORD, F. A. (1939). A study in factor analysis: the stability of a bi-factor solution. *Supplementary Education Monographs, No. 48*. University of Chicago.

JENNRICH, R. I. and ROBINSON, S. M. (1969). A Newton–Raphson algorithm for Maximum Likelihood factor analysis. *Psychometrika*, **34**, 111–23.

JENNRICH, R. I. and SAMPSON, P. F. (1968). Application of stepwise regression to non-linear estimation. *Technometrics*, **10**, 63–72.

JÖRESKOG, K. G. (1967). Some contributions to maximum likelihood factor analysis. *Psychometrika*, **32**, 443–82.

JÖRESKOG, K. G. (1969). A general approach to confirmatory Maximum Likelihood factor analysis. *Psychometrika*, **34**, 138–202.

JÖRESKOG, K. G. and GOLDBERGER, A. S. (1972). Factor analysis by generalized least squares. *Psychometrika*, **37**, 243–60.

LAWLEY, D. W. and MAXWELL, A. E. (1963). *Factor Analysis as a Statistical Method*, 2nd edn. London: Butterworth.

LEE, S. Y. (1977). Some algorithms for the covariance structure analysis. Ph.D. Dissertation, Department of Biostatistics, UCLA.

MARQUARDT, D. W. (1963). An algorithm for least squares estimation of non-linear parameters. *SIAM Journal*, **11**, 431–41.