PSYCHOMETRIKA—VOL. 34, NO. 1
MARCH, 1969

A NEWTON-RAPHSON ALGORITHM FOR MAXIMUM LIKELIHOOD FACTOR ANALYSIS*

ROBERT I. JENNRICH AND STEPHEN M. ROBINSON

UNIVERSITY OF CALIFORNIA, LOS ANGELES, AND UNITED STATES ARMY

This paper demonstrates the feasibility of using a Newton–Raphson algorithm to solve the likelihood equations which arise in maximum likelihood factor analysis. The algorithm leads to clean easily identifiable convergence and provides a means of verifying that the solution obtained is at least a local maximum of the likelihood function. It is shown that a popular iteration algorithm is numerically unstable under conditions which are encountered in practice and that, as a result, inaccurate solutions have been presented in the literature. The key result is a computationally feasible formula for the second differential of a partially maximized form of the likelihood function. In addition to implementing the Newton–Raphson algorithm, this formula provides a means for estimating the asymptotic variances and covariances of the maximum likelihood estimators.

1. *Introduction*

The maximum likelihood method of estimating factor loadings has a number of desirable statistical properties which include asymptotic efficiency, invariance under change of scale, and the existence of a χ^2-test for additional factors. The method has been extensively discussed by Lawley [1940], Rao [1955], and Anderson and Rubin [1956]. These authors suggest simple iteration algorithms for estimation which, with minor variations, consist in finding factor loadings Λ which maximize the likelihood function for a specified set Δ^2 of unique variances. The unique variances are then modified so that the sum of each communality and the corresponding unique variance is equal to the corresponding sample variance. This process, which is summarized in (9) below, is repeated until Δ^2 and Λ converge. While it can be shown that the maximum likelihood estimates of Δ^2 and Λ constitute a stationary point of this process, a number of difficulties arises. The convergence of the algorithm is at best linear. This means that from a computational point of view it may be, and in fact frequently is, difficult to recognize when convergence has occurred. A still greater difficulty is that in some cases the algorithm will not converge at all. In addition to showing why this can happen, it will be shown that real data analyzed by Rao [1955] and Harman [1960] are of this unfavorable type and that apparently as a consequence these authors were led to erroneous solutions. To eliminate these difficulties the authors propose

*This research was supported by the Air Force Office of Scientific Research, Grant No. AF-AFOSR-4.59-66 and by National Institutes of Health, Grant No. FR-3.

solving the appropriate likelihood equations by a method based on the Newton–Raphson algorithm [Henrici, 1964, p. 105]. With sufficiently good starting values the Newton–Raphson algorithm converges and the convergence is quadratic. That sufficiently good starting values can be found is illustrated by the examples in Section 5. The value of quadratic convergence is indicated by the clean, easily identifiable convergence they display.

2. Discussion of the Problem

The factor analysis model as given by Anderson and Rubin [1956] states that a vector x of n scores has a normal distribution with mean vector μ and covariance matrix $\Sigma = \Lambda\Lambda' + \Delta^2$. The matrix Λ is n by m with $m < n$ and the matrix Δ is diagonal. Our problem is to find maximum likelihood estimates of μ, Λ, and Δ based on a random sample $x_1, \cdots, x_N$ of score vectors. The likelihood of μ, Λ, and Δ corresponding to the sample is

$$L = (2\pi)^{-N/2} |\Sigma|^{-N/2} \exp\left[-\frac{1}{2}\sum_{i=1}^{N}(x_i - \mu)'\Sigma^{-1}(x_i - \mu)\right]. \tag{1}$$

Letting

$$A = \frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})(x_i - \bar{x})',$$

equation (1) can be put in the form

$$\log(L) = -\tfrac{1}{2}N[\log(2\pi) + \log|\Sigma| + \operatorname{tr}(A\Sigma^{-1}) + (\bar{x} - \mu)'\Sigma^{-1}(\bar{x} - \mu)]. \tag{2}$$

The maximum likelihood estimate of μ is given by $\bar{x}$. Setting $\mu = \bar{x}$, (2) becomes

$$\phi = -\tfrac{1}{2}[\log(2\pi) + \log|\Sigma| + \operatorname{tr}(A\Sigma^{-1})] \tag{3}$$

when $\phi = \log(L)/N$. Viewing ϕ and Σ as functions of Λ and Δ, let $d\Sigma$ and $d\phi$ denote the differential elements corresponding to $d\Lambda$ and $d\Delta$. Then

$$\begin{aligned} d\phi &= \tfrac{1}{2}\operatorname{tr}[\Sigma^{-1}(A - \Sigma)\Sigma^{-1}\, d\Sigma] \\ d\Sigma &= d\Lambda\,\Lambda' + \Lambda\, d\Lambda' + 2\Delta\, d\Delta. \end{aligned} \tag{4}$$

The element $d\phi$ equals 0 for all $d\Lambda$ and $d\Delta$ if and only if

$$\begin{aligned} \Sigma^{-1}(A - \Sigma)\Sigma^{-1}\Lambda &= 0 \\ \operatorname{diag}[\Sigma^{-1}(A - \Sigma)\Sigma^{-1}\Delta] &= 0. \end{aligned} \tag{5}$$

These are the likelihood equations for Λ and Δ corresponding to the sample $x_1, \cdots, x_N$. The first likelihood equation can be put into the equivalent form

$$(A - \Delta^2)A^{-1}\Lambda = \Lambda\Lambda'A^{-1}\Lambda. \tag{6}$$

Let $\gamma_1, \cdots, \gamma_m$ denote the m largest eigenvalues of $(A - \Delta^2)A^{-1}$ and let $\ell_1, \cdots, \ell_m$ denote the corresponding eigenvectors normalized so that $\ell_i' A^{-1} \ell_i = \gamma_i$. Let

$$\Lambda = (\ell_1, \cdots, \ell_m). \tag{7}$$

Then Λ satisfies (6), and the chain of implications

$$(A - \Delta^2)A^{-1}\Lambda = \Lambda\Lambda' A^{-1}\Lambda \Rightarrow (A - \Delta^2 - \Lambda\Lambda')A^{-1}\Lambda = 0$$
$$\Rightarrow (A - \Sigma)A^{-1}\Lambda = 0 \Rightarrow \Lambda = \Sigma A^{-1}\Lambda \Rightarrow A\Sigma^{-1}\Lambda = \Lambda$$
$$\Rightarrow (A - \Sigma)\Sigma^{-1}\Lambda = 0 \Rightarrow \Sigma^{-1}(A - \Sigma)\Sigma^{-1}\Lambda = 0$$

shows that Λ satisfies the first likelihood equation in (5). Using the first, the second likelihood equation can be put in the form

$$\text{diag}\,(A - \Sigma) = 0. \tag{8}$$

View Λ as a function of Δ given by (7). A root Δ of (8) together with the corresponding Λ gives a solution to the likelihood equations (5). Equation (8) may be put in the form

$$\Delta^2 = \text{diag}\,(A - \Lambda\Lambda') \tag{9}$$

which suggests a simple iteration algorithm for obtaining roots of (8). This algorithm with or without minor variations has been proposed by Lawley [1940], Rao [1955], and others and has become the standard algorithm for maximum likelihood estimation. Two problems arise. First, it will be shown that when Δ is singular the differential of the transformation which takes Δ^2 into diag $(A - \Lambda\Lambda')$ has an eigenvalue of 1. This means that the iteration suggested by (9) is numerically unstable at a singular Δ and that it will converge slowly, at best, to a nearly singular Δ. The examples of Section 5 will show that solutions involving singular Δ's can and do arise in practice. Second, even when the simple iteration does converge to a solution it does so at best linearly. As pointed out earlier this makes it difficult, in practice, to identify convergence.

Rather than using simple iteration the authors propose solving the second likelihood equation by means of the Newton–Raphson algorithm. The basic problem is to find a practical formula for the differential of the transformation which takes Δ into diag $[\Sigma^{-1}(A - \Sigma)\Sigma^{-1}\Delta]$.

3. Newton–Raphson Solution

As in the previous section, assume that Λ is a function of Δ given by (7). Let ϕ' denote the differential of the transformation which takes Δ into ϕ and let ϕ'' denote the second differential of this transformation. The Newton–Raphson algorithm for finding a solution to the equation $\phi' = 0$ is given by the repeated replacement of Δ by

$$\tilde{\Delta} = \Delta - (\phi'')^{-1}\phi'. \tag{10}$$

The algorithm is not applied to the entire set of $(m + 1)n$ parameters in Λ and Δ but only to the n parameters in Δ. In effect the Newton–Raphson algorithm is being used to solve the second likelihood equation in (5) under the assumption that the first is satisfied.

In order to find a formula for ϕ'' it is convenient to let $\tilde{\Lambda} = A^{-1/2}\Lambda$, $\tilde{\ell}_i = A^{-1/2}\ell_i$, $\tilde{\Delta} = A^{-1/2}\Delta^2 A^{-1/2}$, and $\tilde{\Sigma} = A^{-1/2}\Sigma A^{-1/2}$. Let f denote the function which takes $\tilde{\Delta}$ into ϕ. It follows from equations (4) and (5) that the differential of this function is given by

$$df = \tfrac{1}{2}\,\mathrm{tr}\,[\tilde{\Sigma}^{-1}(I - \tilde{\Sigma})\tilde{\Sigma}^{-1}d\tilde{\Delta}]. \tag{11}$$

It follows from (6) that the m largest eigenvalues of $I - \tilde{\Delta}$ are $\gamma_1, \cdots, \gamma_m$. Let $\gamma_1, \cdots, \gamma_n$ denote all the eigenvalues of $I - \tilde{\Delta}$ and let $v_1, \cdots, v_n$ denote the corresponding eigenvectors normalized so that $v_i'v_i = 1$. Then

$$(I - \tilde{\Delta})v_i = \gamma_i v_i, \qquad v_i'v_i = 1. \tag{12}$$

It follows that $\tilde{\ell}_i = \gamma_i^{1/2}v_i$ for $i \leq m$ and hence

$$\tilde{\Lambda}\tilde{\Lambda}'v_i = \gamma_i v_i, \qquad i \leq m. \tag{13}$$

Thus by (12) and (13), $\tilde{\Sigma}$ has the spectral representation

$$\tilde{\Sigma} = \sum_{i=1}^{m} v_i v_i' + \sum_{i=m+1}^{n} (1 - \gamma_i)v_i v_i' \tag{14}$$

and hence,

$$\tilde{\Sigma}^{-1}(I - \tilde{\Sigma})\tilde{\Sigma}^{-1} = \sum_{i=m+1}^{n} \gamma_i(1 - \gamma_i)^{-2}v_i v_i'. \tag{15}$$

Finally from (11),

$$df = \frac{1}{2}\sum_{i=m+1}^{n} \alpha_i v_i'\, d\tilde{\Delta}\, v_i \tag{16}$$

where $\alpha_i = \gamma_i(1 - \gamma_i)^{-2}$. Let ddf denote the second differential element of the function f corresponding to the differential elements $d\tilde{\Delta}_1$ and $d\tilde{\Delta}_2$. Then

$$ddf = \frac{1}{2}\sum_{i=m+1}^{n} (\alpha_i'\, d\gamma_i\, v_i'\, d\tilde{\Delta}_1\, v_i + 2\alpha_i v_i'\, d\tilde{\Delta}_1\, dv_i) \tag{17}$$

where $\alpha_i' = d\alpha_i/d\gamma_i$ and $d\gamma_i$ and dv_i are the differential values of γ_i and v_i corresponding to $d\tilde{\Delta}_2$. From (12),

$$-d\tilde{\Delta}_2 v_i + (I - \tilde{\Delta})\, dv_i = d\gamma_i v_i + \gamma_i\, dv_i \tag{18}$$

and $v'\, dv_i = 0$. Multiplying both members of (18) by v_i' gives

$$d\gamma_i = -v_i'\, d\tilde{\Delta}_2 v_i \tag{19}$$

and multiplying both members of (18) by v_j', where $j \neq i$, gives

$$v_j' \, dv_i = -(\gamma_i - \gamma_j)^{-1} v_j' \, d\tilde{\Delta}_2 v_i \, .$$

Multiplying both sides of this equation by v_j and summing over all $j \neq i$ gives

$$dv_i = -\sum_{j \neq i} (\gamma_i - \gamma_j)^{-1} v_j v_j' \, d\tilde{\Delta}_2 \, v_i \, . \tag{20}$$

Using (19) and (20), (17) becomes

$$ddf = -\sum_{i=m+1}^{n} \sum_{j=1}^{n} \beta_{ij} v_i' \, d\tilde{\Delta}_1 \, v_j v_i' \, d\tilde{\Delta}_2 \, v_j \tag{21}$$

where

$$\beta_{ij} = \begin{cases} \frac{1}{2}\alpha_i' = \frac{1}{2}(1 + \gamma_i)(1 - \gamma_i)^{-3}, & i = j \\ \alpha_i(\gamma_i - \gamma_j)^{-1} = \gamma_i(1 - \gamma_i)^{-2}(\gamma_i - \gamma_j)^{-1}, & i \neq j. \end{cases} \tag{22}$$

Let g denote the function which takes Δ^2 into ϕ and let $u_i = A^{-1/2} v_i$. It follows from (16) and (21) that the first and second differentials g' and g'' of g are given by

$$\begin{aligned} dg &= \frac{1}{2} \sum_{i=m+1}^{n} \alpha_i u_i' \, d\Delta^2 \, u_i \\ ddg &= -\sum_{i=m+1}^{n} \sum_{j=1}^{n} \beta_{ij} u_i' \, d\Delta_1^2 \, u_j u_i' \, d\Delta_2^2 \, u_j \, . \end{aligned} \tag{23}$$

The matrices (g_r') and (g_{rs}'') of these differentials are given by

$$\begin{aligned} g_r' &= \frac{1}{2} \sum_{i=m+1}^{n} \alpha_i u_{ri}^2 \\ g_{rs}'' &= \sum_{i=m+1}^{n} \sum_{j=1}^{n} \beta_{ij} u_{ri} u_{rj} u_{si} u_{sj} \, . \end{aligned} \tag{24}$$

Finally, the matrices (ϕ_r') and (ϕ_{rs}'') of ϕ' and ϕ'' are given by

$$\begin{aligned} \phi_r' &= 2 \, \Delta_r \, g_r' \\ \phi_{rs}'' &= 4 \, \Delta_r \, g_{rs}'' \, \Delta_s + 2 \, \delta_{rs} \, g_r' \end{aligned} \tag{25}$$

where δ_{rs} denotes the Kronecker delta. In summary, the computational form of the Newton–Raphson algorithm is given by (10) where ϕ' and ϕ'' are obtained from equations (22), (24), and (25) and the solution to the eigen problem

$$A^{-1/2}(A - \Delta^2)A^{-1/2} v_i = \gamma_i v_i \, , \qquad v_i' v_i = 1. \tag{26}$$

It is useful to have a computational formula for the likelihood L or, equivalently, for ϕ. Such a formula, which follows from (3) and (14), is given by

$$(27) \qquad \phi = -\frac{1}{2}\left[\log(2\pi) + \log|A| + m + \sum_{i=m+1}^{n} \{\log(1-\gamma_i) + (1-\gamma_i)^{-1}\}\right].$$

This equation clears up a difficult point. In effect we have presented an algorithm for solving the likelihood equations under the restriction that Λ is a function of Δ given by (7). It would be comforting to know that such a pair of values can produce an absolute maximum likelihood. We may assume without loss of generality that for an absolute maximum $\Lambda' A^{-1}\Lambda$ is diagonal. It follows from (6) that $\ell_1, \cdots, \ell_m$ must be eigenvectors of $(A - \Delta)A^{-1}$ but not that they must, as we demanded, correspond to the m largest eigenvalues. Since $\Lambda' A^{-1}\Lambda$ is non-negative definite, the eigenvalues corresponding to the ℓ_i must be non-negative. Since $\log(1-\gamma) + (1-\gamma)^{-1}$ is a monotonically increasing function of γ for $\gamma \geq 0$ it follows from (27) that for an absolute maximum $\gamma_1, \cdots, \gamma_m$ must be the largest eigenvalues of $(A - \Delta)A^{-1}$.

4. Why Simple Iteration Doesn't Work

It was stated earlier that when Δ is singular the differential of the transformation which takes Δ^2 into diag $(A - \Lambda\Lambda')$ has an eigenvalue of 1 and hence the simple iteration given by (9) is numerically unstable for singular Δ. It is sufficient to prove that the differential of the transformation h which takes Δ^2 into diag $(A - \Sigma)$ is singular when Δ is singular. We have

$$(28) \qquad A - \Sigma = (\Lambda\Lambda' + \Delta^2)\Sigma^{-1}(A - \Sigma)\Sigma^{-1}(\Lambda\Lambda' + \Delta^2)$$

and hence, using the first equation in (5),

$$(29) \qquad \text{diag}\,(A - \Sigma) = \Delta^2\,\text{diag}\,[\Sigma^{-1}(A - \Sigma)\Sigma^{-1}]\Delta^2.$$

Viewing diag $[\Sigma^{-1}(A - \Sigma)\Sigma^{-1}]$ as a function w of Δ^2, let dh and dw denote the differential elements which correspond to $d\Delta^2$. Then,

$$dh = 2\Delta^2 w\, d\Delta^2 + \Delta^4\, dw.$$

If Δ is singular, then as $d\Delta^2$ ranges over the space of n by n diagonal matrices, dh ranges over a proper subspace of such matrices. It follows that the differential of h is singular when Δ is singular.

5. Examples

In this section we shall look at the results of applying the Newton–Raphson algorithm to two examples in the literature, both of which use real data. The first was given by Rao [1955, p. 110] and the second by Harman [1960, p. 376]. Both authors used simple iteration algorithms similar to (9) to obtain the solutions they present. Table 1 shows the results of applying the Newton–Raphson algorithm to Rao's example using his solution for starting values. The algorithm converged cleanly and solidly in six steps to a singular

TABLE 1

Convergence of the Newton-Raphson algorithm applied to Rao's example using Rao's solution for starting values. After convergence $\max|\phi'_r| = 1.3 \cdot 10^{-6}$ and tol = .963.

Iteration	Δ_1	Δ_2	Δ_3	Δ_4	Δ_5	Δ_6	Δ_7	Δ_8	Δ_9	ϕ
0	.44000	.57000	.88000	.90000	.73000	.43000	.55000	.75000	.63000	-2.8733
1	.35047	.58388	.87870	.90010	.72858	.43078	.54384	.74687	.52029	-2.8601
2	.25366	.58521	.87909	.90224	.72929	.43155	.54455	.74778	.54483	-2.8577
3	.13189	.58761	.87930	.90283	.72993	.43214	.54444	.74786	.54662	-2.8574
4	.03105	.58762	.87934	.90297	.72990	.43210	.54443	.74791	.54663	-2.8574
5	.00117	.58717	.87931	.90289	.72981	.43208	.54445	.74790	.54664	-2.8573
6	.00000	.58713	.87931	.90289	.72980	.43207	.54445	.74790	.54664	-2.8573
7	.00000	.58713	.87931	.90289	.72980	.43207	.54445	.74790	.54664	-2.8573

Δ. That Rao's solution does not correspond to a maximum likelihood solution is indicated by the fact that ϕ increases from -2.8733 to -2.8573. That the solutions differ substantially can be seen by looking at the corresponding factor loadings presented in Table 2. The fact that $\max_r |\phi'_r| = 1.3 \cdot 10^{-6}$ indicates that the solution given in Table 2 is at least a stationary point of the likelihood function. To show that it is a local maximum one could compute the eigenvalues of the matrix (ϕ''_{rs}) and verify that they are all negative. Equivalently one could verify that the tolerance of the matrix $-(\phi''_{rs})$ is positive [Jennrich and Sampson, 1968]. Since this tolerance is a standard output of the matrix inversion routines used by the authors, the latter criterion was used. The tolerance of the solution given in Table 1 is .963 and hence the solution is at least a local maximum of the likelihood function.

TABLE 2

Rao's Loadings

Factor 1	.845	.817	.477	.401	.669	.891	.834	.651	.833
Factor 2	-.309	-.084	.012	.153	.161	.145	.081	.122	.080

Newton-Raphson Loadings

Factor 1	1.000	.720	.410	.280	.520	.710	.680	.510	.680
Factor 2	.000	.370	.242	.326	.444	.556	.491	.425	.489

The results of applying the Newton–Raphson algorithm to Harman's example are given in Table 3. Originally the authors used Harman's answers as starting values, but the likelihood function was not concave down at this starting point and the algorithm converged to a saddle point of the likelihood function. Investigation of the solution suggested that the starting value of Δ_8 was too small. After increasing this value to .700 the Newton–Raphson iteration converged to the local maximum presented in Table 3. The value of ϕ corresponding to Harman's solution was -1.5273. The Newton–Raphson value of -1.4863 indicates again that Harman's solution does not correspond to a maximum of the likelihood function. That the solutions differ substantially can be seen by comparing the factor loadings presented in Table 4.

TABLE 3

Convergence of the Newton-Raphson algorithm applied to Harman's example using Harman's solution for starting values with Δ_8 replaced by .700. After convergence $\max|\phi'_r| = 2.5 \cdot 10^{-6}$ and tol = .537.

Iteration	Δ_1	Δ_2	Δ_3	Δ_4	Δ_5	Δ_6	Δ_7	Δ_8	ϕ
0	.39900	.30500	.41100	.43800	.29200	.60400	.64700	.70000	-1.5144
1	.34229	.22020	.42361	.38893	.29645	.60041	.64175	.68895	-1.4928
2	.36919	.13994	.43790	.38478	.29690	.59888	.64271	.69673	-1.4883
3	.35652	.06816	.44072	.39715	.30059	.59972	.64049	.70049	-1.4867
4	.35719	.01369	.44132	.39516	.30102	.59938	.64066	.70123	-1.4864
5	.35693	.00067	.44053	.39555	.30070	.59941	.64071	.70073	-1.4864
6	.35692	.00000	.44048	.39557	.30068	.59941	.64072	.70071	-1.4863
7	.35692	.00000	.44048	.39557	.30068	.59941	.64072	.70071	-1.4863

It should be pointed out that Harman apparently performed his calculations by hand and carried out only five iterations. He also suggested that he may not have achieved convergence. Rao's calculations on the other hand were carried out on an electronic computer presumably to a point that appeared stationary.

While both examples produced singular Δ estimates, this is the exception rather than the rule. The examples demonstrate that this exception occurs in practice and with it the convergence problems discussed in Section 4. The fact that the estimate of Δ is singular does not imply that the population Δ is singular. Thus, the zero estimate for Δ_1 in the first example suggests that the corresponding population value is small, but it need not be zero.

An interesting consequence of singular Δ estimates, or singular Δ's in general, can be seen from Tables 2 and 4. In each table one variable has a loading of exactly one on one factor and exactly zero on the remaining.

TABLE 4

Harman's Loadings

Factor 1	.874	.874	.838	.849	.710	.596	.534	.612
Factor 2	-.258	-.370	-.358	-.281	.625	.529	.544	.423
Factor 3	-.102	.081	-.013	-.093	-.141	-.021	-.028	.493

Newton-Raphson Loadings

Factor 1	.846	1.000	.881	.826	.376	.326	.277	.415
Factor 2	.189	.000	.055	.160	.876	.725	.704	.543
Factor 3	-.348	.000	-.164	-.368	.031	.093	.130	.204

This phenomenon occurs as follows. Any variable with a zero unique variance must lie in the common factor space. If there is only one such variable it must be a canonical variable and, in the case of Rao's canonical factor analysis, colinear with a factor. The phenomenon would disappear if some other rotation criterion, such as varimax, were used.

The investigation presented here is similar in some respects to that of Karl Jöreskog [1967]. His basic iteration uses the method of Fletcher and Powell [1963] and resembles our Newton–Raphson iteration in a number of ways. His treatment of the singular Δ problem, however, differs substantially. No attempt will be made to compare the two algorithms systematically but it is perhaps worth pointing out that, in personal correspondence, Harry Harman has observed that, when applied to the second example discussed above, both algorithms converged rapidly to identical answers.

6. *Some Practical Considerations*

The existence of solutions involving singular Δ suggests that if the diagonal components of Δ^2 were permitted to be negative, then larger values of the likelihood function might be obtained. Such solutions would suggest that the factor analysis model may not represent the population being sampled. It was hoped that if Δ^2 were replaced by Δ in the original model, thus in effect allowing negative unique variances, the resulting solutions would indicate the degree, if any, to which the original model failed to represent the sampled population. It was also hoped that such a modification would eliminate saddle points associated with singular Δ's. In some cases, including Harman's example, finite solutions involving negative unique variances were found, but in others neither hope was realized. Instead, zero unique variances were replaced by negatively infinite unique variances and singular Δ saddle points by saddle points at infinity. Because of this and because negative variances make little sense, it seems that for a general purpose algorithm it is best to retain the Δ^2 formulation.

The major computational effort in our Newton–Raphson algorithm is devoted to the computation of the matrix (g''_{rs}). This requires on the order of $\frac{1}{2}(n - m)n^3$ multiplications. It is in theory possible to reduce this to about $\frac{1}{2}mn^3$ multiplications. Since in practice m is frequently less than $n/10$ this modification could result in as much as ten-fold increase in computation speed. The modification is based on the observation (stated here without proof) that g''_{rs} can be written in the form

$$g''_{rs} = (\tfrac{1}{2}\,\Delta_r^{-4} - \Delta_r^{-6}\,A_{rr})\,\delta_{rs} + \sum_{i=1}^{m}\sum_{j=1}^{n}\beta_{ij}u_{ri}u_{rj}u_{si}u_{sj}\,. \tag{30}$$

When Δ is nearly singular this formulation has obvious numerical problems which may or may not be serious. It was a fear that they might be serious that prompted the authors to use formula (24) rather than (30). Moreover, the actual computation time required by the original formulation seems quite modest. Using an IBM Model 75 computer, the 8 and 9 variable problems discussed in Section 5 required about 10 seconds of computing time each.

The algorithm presented here has been used on a total of about ten problems. While this is sufficient to prove its feasibility, it does not demonstrate its practicality. The latter, in the opinion of the authors, requires a user-oriented program and extensive use on a large variety of problems.

7. *Conclusion*

This paper demonstrates the feasibility of using a Newton–Raphson algorithm to solve the likelihood equations which arise in maximum likelihood factor analysis. The algorithm leads to clean and solid quadratic convergence which in applied problems contrasts significantly with the slow linear convergence of popular simple iteration algorithms. In addition, the Newton–Raphson algorithm provides a means of verifying that the solution obtained is at least a local maximum of the likelihood function.

It has been shown that for a popular simple iteration algorithm, convergence to a solution is impossible under conditions which are encountered in practice. Because of this and because of the characteristically slow convergence of simple iteration algorithms, false solutions have been presented in the literature.

The maximum likelihood method of factor analysis, as compared to other methods, is relatively unpopular. Considering the difficulties encountered in the use of formula (9), this is perhaps a tribute to the insight of factor analysis practitioners. However, the simple iteration algorithm given by (9) is quite similar to the simple iteration algorithm used in least squares factor analysis [Harman, 1960, p. 89]. The latter algorithm, which is usually referred to as "iteration for communalities," is a feature of most standard factor analysis programs and its use is quite popular. We know from experience that the algorithm converges slowly and one wonders to what extent it is plagued by the same difficulties that plague algorithm (9). Using the techniques of Section 3 the authors have derived Newton–Raphson formulas for the least squares factor analysis problem, but to date the algorithm has not been programmed.

The key to deriving a Newton–Raphson algorithm for maximum likelihood factor analysis was the development of a formula for the matrix (ϕ_{rs}''). This matrix is also the key to formulas for the asymptotic variances of parameter estimates. For example, $(\phi_{rs}'')^{-1}$ is a consistent estimate of the asymptotic covariance matrix of the maximum likelihood estimates of the unique standard deviations. The authors hope to discuss related results in a subsequent paper.

REFERENCES

Anderson, T. W., & Rubin, H. Statistical inference in factor analysis. *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, Univ. of California Press, 1956, **5**, 111–149.

Fletcher, R., & Powell, M. J. D. A rapidly convergent descent method for minimization. *Computer Journal*, 1963, **6**, 163–168.

Harman, H. H. *Modern factor analysis*. Univ. of Chicago Press, 1960.

Henrici, P. *Elements of numerical analysis*. Wiley, New York, 1964.

Jennrich, R. I., & Sampson, P. F. Application of stepwise regression to non-linear estimation. *Technometrics*, 1968, **10**, 63–72.

Jöreskog, K. G. Some contributions to maximum likelihood factor analysis. *Psychometrika*, 1967, **32**, 443–482.

Lawley, D. N. The estimation of factor loadings by the method of maximum likelihood. *Proceedings of the Royal Society of Edinburgh*, 1940, **60**, 64–82.

Rao, C. R. Estimation and tests of significance in factor analysis. *Psychometrika*, 1955, **20**, 93–111.

Manuscript received 8/9/67
Revised manuscript received 1/19/68