PROBLEMS WITH EM ALGORITHMS FOR ML FACTOR ANALYSIS

P. M. BENTLER AND JEFFREY S. TANAKA
UNIVERSITY OF CALIFORNIA, LOS ANGELES

Rubin and Thayer recently presented equations to implement maximum likelihood (ML) estimation in factor analysis via the EM algorithm. They present an example to demonstrate the efficacy of the algorithm, and propose that their recovery of multiple local maxima of the ML function "certainly should cast doubt on the general utility of second derivatives of the log likelihood as measures of precision of estimation." It is shown here, in contrast, that these second derivatives verify that Rubin and Thayer did not find multiple local maxima as claimed. The only known maximum remains the one found by Jöreskog over a decade earlier. The standard errors obtained from the second derivatives and the Fisher information matrix thus remain appropriate where ML assumptions are met. The advantages of the EM algorithm over other algorithms for ML factor analysis remain to be demonstrated.

Key words: factor analysis, EM algorithms, maximum likelihood.

In their classic paper, Dempster, Laird, and Rubin [1977] introduced the EM algorithm, an attractive general algorithm for maximum likelihood (ML) estimation, and described its application to numerous statistical models, including factor analysis. Several discussants of that paper (e.g., J. A. Nelder) mentioned that it would be desirable to have a detailed algebraic development of this application. Rubin and Thayer [1982] recently provided this valuable development. In addition, they illustrated the method on a four factor example previously discussed by Jöreskog [1969].

Some advantages of the EM algorithm mentioned by Rubin and Thayer are that 1) if the algorithm converges, it will converge to a local maximum; 2) it is simple to program; 3) it is computationally inexpensive; and 4) it yields a monotonically increasing continuous path, enabling it to find multiple maxima if they exist. Their example appeared to verify these claims, in that the method apparently rather easily and quickly found not only one, but two other local maxima of the ML function. As stated by Rubin and Thayer:

'The fact that three different starting values lead to three essentially different (i.e., different \( r^2 \) [unique variance]) solutions is quite interesting, and complicates the interpretation of any solution ... We have no reason to believe that these are the only maximums for this problem. In fact, one referee points out that a two factor solution (i.e., all \( \beta \)'s [factor loadings] in factors 3 and 4 set to zero) also appears to result in local maximum of the likelihood! Such behavior of the likelihood function should make any user of maximum likelihood factor analysis very uneasy. Minimally, the use of standard errors based on the second derivative matrix evaluated at a mode to measure precision of estimation should be viewed as being entirely questionable' [p. 75].

In general, Rubin and Thayer are pessimistic about the value of ML estimation in the factor model.

The purpose of this note is to point out that Rubin and Thayer's pessimism about ML factor analysis is not well founded. Rather than highlight the limitations of ML factor analysis, their example demonstrates weaknesses in the EM algorithm that were already evident to several discussants of the Dempster et al. paper.

Supported in part by grants DA00017 and DA01070 from the U.S. Public Health Service.

Requests for reprints should be sent to P. M. Bentler, Department of Psychology, University of California, Los Angeles, California 90024.
Rubin and Thayer analyzed a nine variable problem hypothesized to have a four factor solution. Based on $N = 710$, two factors were considered to have loadings free to be estimated, and two further factors were considered to have loadings in a simple cluster structure with no variable loading on both factors. They estimated these parameters and the associated unique variances using three sets of start values: ad hoc (Solution 1); based on an initial principal components solution (Solution 2); and based on Jöreskog's [1969] prior solution (Solution 3). (Rubin and Thayer incorrectly call Solution 2 the Jöreskog solution.) They terminated the EM algorithm after 50 iterations, and proposed that they had found three separate, different ML solutions. Further, as noted above, they imply that a fourth solution which zeros out factors 3 and 4, also exists.

In contrast to the conclusion reached by Rubin and Thayer, it will be shown that only one of the four solutions mentioned by them is in fact an ML solution. This is the solution previously described by Jöreskog.

It is well-known that two conditions must be met for a solution to be a candidate for a strict local maximum of the likelihood function: the gradient vector of first-order partial derivatives must vanish, and the Hessian matrix of second partial derivatives must be negative definite if the parameters are locally identified. The Hessian may also be negative semidefinite, as in this example, when identification conditions are not imposed. Since the first two factors in Rubin and Thayer's example (and in Jöreskog's solution) can be subjected to an arbitrary orthonormal rotation, one additional constraint can be imposed on any proposed solution. In order to achieve identification, the first two factors of all proposed solutions were orthonormally rotated so that the $(1, 2)$ element of the loading matrix could be fixed at zero, thus eliminating this parameter as one to be estimated. The gradient and Hessian for this reduced problem were evaluated in all three cases at Rubin and Thayer's proposed solution estimates, using Bentler and Lee's [1979] formulas and algorithm implemented in a double precision FORTRAN program on an IBM 3033. The gradient elements were in all cases close to zero, but probably not sufficiently close to meet strict criteria for zero (e.g., at least four absolute values were greater than .02 in each solution). Consequently, each of the proposed solutions was modified slightly by taking a few iterative steps until the likelihood functions equaled or exceeded those reported by Rubin and Thayer for the corresponding solutions. This occurred after 4 and 2 steepest ascent iterations in Solutions 1 and 2 respectively; at these points only 0 and 2 absolute gradient elements exceeded .01, but over two-thirds or more of these elements were greater than .001. (A referee had urged us to reach a point with an effectively zero gradient in each case, but this recommendation was not followed here since it would have required moving away from the Rubin and Thayer solutions which were to be evaluated. Of course, the absence of a zero gradient would, by itself, argue that the solutions cannot be local maxima of the likelihood function.) In the case of Solution 3, four Newton-Raphson steps were taken from the two-decimal solution so as to yield sufficient accuracy in the estimates to have the maximum absolute gradient element be less than .00001. The eigenvalues of the Hessians reported below were independently verified using the IMSL double precision routine EIGRS (IMSL, 1982).

Rubin and Thayer's Solution 1 had an indefinite Hessian matrix. Three of the 35 eigenvalues had a positive sign, and 32 were negative as required. The smallest among the

* Certain technical regularity conditions must also be met [e.g., Mäkeläinen, Schmidt, & Styan, 1981]. Shapiro and Browne [1982] provide a nonregular example with identified parameters which would yield a negative semidefinite Hessian. Solutions with parameters on the boundary of a feasible region (e.g., zero variances) require slightly modified conditions, since constrained rather than unconstrained optimization theory is involved [e.g., Lee & Bentler, 1980; Avriel, 1976].