Saddle Points

Tihomir Asparouhov and Bengt Muthén

March 29, 2012

1 Introduction

To compute the maximum likelihood estimates the log-likelihood function L is maximized with respect to all model parameters. To check that the maximization has been achieved two things have to be satisfied:

1. The vector of first derivatives with respect to the model parameter L' should be equal to 0.

2. The negative of the matrix of the second derivatives $-L''$ should be a positive definite matrix.

Most maximization algorithms are based on following the direction of the derivative until maximization and thus continue iterating through the parameter spaces until condition 1 is met. At that point the maximization algorithm has no direction to move on and stops hoping that condition 2 is also satisfied. After the iterations are completed condition 2 is checked and if it is satisfied then Mplus will conclude that indeed the maximization is complete. If the ML estimator is used then the information matrix $(-L'')^{-1}$ is used as the estimator of the asymptotic variance covariance for the parameter estimates. If the MLF estimator is used the estimator for the asymptotic variance covariance is given by $\overline{L'(L')^T}$ where the bar symbol here means averaging over the independent units in the data (observations in single level models or clusters in multilevel models). If the MLR estimator is used the estimator for the asymptotic variance covariance is given by $(-L'')^{-1}\overline{L'(L')^T}(-L'')^{-1}$. All three estimators are asymptotically equivalent when the model is correctly specified, however for small or medium sample size the MLF estimator may overestimate the standard errors. This is particularly the case if the ratio between the number of independent units (observations in single level models or clusters in two-level models) and the number of parameters is less than 10.

In some cases condition 2 fails. Theoretically speaking all such points where condition 1 is satisfied but condition 2 is not are called saddle points. In many cases however the optimization has not reached a saddle point but there is a different reason for the failure of condition 2. Below we list some of the main reasons for the failure of condition 2.

If condition 2 is not satisfied Mplus will report that the optimization algorithm has reached a saddle point and at that point it will either report the MLF standard errors or it will use the following ad-hoc estimator for the standard errors (if the ML or MLR estimators are used). Instead of using the information matrix $(-L'')^{-1}$, Mplus uses a symmetric matrix M that is obtained from the eigendecomposition of $(-L'')^{-1}$ where we replace all negative eigenvalues with 10^{-10} (or equivalently 0). This way the estimated variance covariance matrix will be positive definite. This matrix is also nearly singular, however, since this matrix is not inverted, the near singularity is not a problem. It will be clear from the warning message which approach has been taken. In general, it is not clear in a particular case which of the two estimators MLF or the approximate information matrix M will provide most accurate approximation for the standard errors. If the sample size is large, the MLF standard errors may be better. If the sample size is moderate or small and we have a good approximation for the information matrix then the matrix M will provide better estimates. If the sample size is large and we have a good approximation for the information matrix the standard errors for the two approaches will be the same. If we don't have a good approximation for the information matrix then the MLF standard errors might be better.

Mplus will essentially attempt to provide the best estimates it can and generally no further action is needed. However, in many cases it may be possible to determine the exact reason for the failure of condition 2 and through manipulation of some technical options in Mplus the failure of condition 2 may be resolved. Below is a list of the most common reasons for condition 2 failure in order of most likely to least likely and the appropriate action that can be taken to resolve the failure in each case.

1. Incomplete convergence

The main convergence criteria in Mplus is to monitor condition 1, i.e., to monitor the first derivatives. When the derivatives become smaller than a predetermined small number (the convergence criterion value which depends on the model and can vary from 10^{-3} to 10^{-6}) they are considered zero but in many cases they may not be sufficiently small. This problem can be resolved by decreasing the convergence criterion by decreasing the CONVERGENCE and the MCONVERGENCE options in the ANALYSIS command.

2. Imprecise estimate of the information matrix

If numerical integration is used in the evaluation of the likelihood the information matrix is estimated with error. This is quite common especially when Montecarlo integration is used. The error can be minimized by increasing the number of integration points using the INTEGRATION option of the ANALYSIS command. In certain cases imprecision in the estimate can occur due to nearly singular variance covariance matrices for the latent variables in the model. In such a case the problem can be resolved by reformulating the model to avoid such near singularities.

3. A true saddle point

A true saddle point can be reached by the optimization algorithm. This is especially common in mixture models. The problem can be resolved by using different starting values or by using the randomized starts option STARTS in the ANALYSIS command.

4. Expected information matrix

In certain models the expected information matrix (or a part of it) can be used as a simplified version of the actual observed information matrix. This is never the default in Mplus so it should not be a common occurrence. The problem can be resolved by using the observed information matrix using the INFORMATION option of the ANALYSIS command. The observed information matrix is preferable in general as it gives unbiased estimates of the standard errors in the presence of MAR missing data while the expected information may be biased.

2 Simulation Study

In this section we illustrate the saddle point situation with a simulation study. We use a simple two-level model with a random intercept and a random slope

$$Y_{ij} = \alpha_j + \beta_j X_{ij} + \varepsilon_{ij}.$$

Table 1: Ratio between average standard error and standard deviation of parameter estimates.

Method	μ_α	v_α
ML(M)	0.96	1.00
MLR(M)	1.06	1.12
ML(MLF)	1.05	1.16
MLR(MLF)	1.14	1.28
MLF	1.18	1.43

where the variance of ε_{ij} is set to 1, and the random effects α_j and β_j have means 0 and variances 1 and 0.01 respectively and a covariance of 0.09, i.e., highly correlated. The covariate X_{ij} is generated from a standard normal distribution. We generate 100 samples using the above model with 20 clusters each of size 5. Table 1 reports the ratios between the average standard errors and the standard deviation of the parameter estimates. If the standard errors are correct, this ratio should be close to 1. We report the results for the mean μ_α and the variance v_α parameters for the random effect α_j. In 35 of the samples the information matrix is not positive definite and thus condition 2 fails. We analyze the results with 5 different methods

- ML (M) uses the ML estimator with the M matrix approach when saddle points occur
- MLR (M) uses the MLR estimator with the M matrix approach when saddle points occur
- ML (MLF) uses the ML estimator with the MLF matrix approach when saddle points occur
- MLR (MLF) uses the MLR estimator with the MLF matrix approach when saddle points occur
- MLF uses the MLF estimator

The above results suggest that as expected the MLF estimator overestimates the standard errors. This is due to the small number of clusters in the samples. The results also suggest that the M approach of dealing with saddle points may be superior to the MLF approach. In Mplus Version 7 the ML(M) and MLR(M) methods are implemented and these replace the ML(MLF) and MLR(MLF) methods used in the previous versions. If needed the MLF estimates can be obtained by requesting the MLF estimator.