**Model convergence and eigenvalue analysis of covariance matrix**

**Content**

[**Background on model convergence**](#_s65hskvrxdhf) **1**

[**Alternative initial parameter values (i.e., jitter run)**](#_gjdgxs) **6**

[**Convergence Criteria: First Order Conditions**](#_9p37klzb8q9) **7**

[**Convergence Criteria: Second Order Conditions**](#_im7tzi1mvty) **8**

Seems like some text on robustness and reparameterization could be added

# Background on model convergence

Convergence is a key property of a stock assessment model. A lack of model convergence indicates that the optimization criteria used for statistical estimation are not being satisfied and therefore that the model does not actually solve the estimation problem. Here one can think of the assessment model as a combination of model structure and input data that produces a particular mode of analysis for the population and fishery system dynamics. The construction of the assessment model requires assumptions, assertions and choices regarding the statistical and mathematical relationships between the model structure and input data. One of the key aspects of this construction is the choice of the statistical framework for estimating the free parameters in the model. In this context, modern stock assessment models use one of three estimation frameworks: frequentist, Bayesian, or random (a.k.a., mixed random and fixed parameters) effects. Convergence is the verification that the estimates of model parameters solve the optimization problem associated with each statistical framework. In what follows, we will describe some best practices for verifying that an individual assessment model converges to an optimal solution under each of the statistical frameworks.

Frequentist assessment models are built around the joint likelihood function of the observed input data. The frequentist interpretation of estimation asserts that model parameters are true unknowns that can be evaluated through the minimization of the negative joint loglikelihood of observed data. This is the first primary approach used for optimizing parameters of stock assessment models in the 1980s based on maximum likelihood estimation criteria. In this case, the model solution consists of a point estimate of the parameter vector that maximizes the joint likelihood. The uncertainty of the MLE is evaluated based on the inverse of the Hessian matrix evaluated at the MLE. The MLE solution exists under general conditions that the joint loglikelihood is a smooth differentiable function of the parameters and that the solution is in the interior of the parameter space. In this case, the MLE solution is asymptotically approximated by a multivariate normal distribution with mean equal to the calculated MLE and the covariance matrix equal to the inverse of the calculated Hessian evaluated at the MLE.

The MLE solution is calculated by applying an optimization algorithm to maximize the joint likelihood function, or equivalently to minimize the negative joint likelihood. There are various iterative approaches that can be applied to calculate the MLE based on numerical differentiation or other search algorithms. Here the convergence of a sequence of estimates to a solution is evaluated by measuring the distance between successive estimates of the MLE parameter vector or by measuring the distance of the model gradient relative to zero. This convergence check based on vector distances is essentially equivalent to measuring the absolute magnitude of elements of the gradient of the negative loglikelihood and ensuring that the gradient at the calculated solution is effectively equal to the zero vector. Thus, checking that the gradient of a potential MLE is effectively zero is the first order condition for minimizing the negative loglikelihood. The second necessary conditions for a local minimum to exist is that the curvature of the negative loglikelihood measured at the MLE is positive and convex. This second order condition is equivalent to checking that the inverse of the Hessian matrix, or the approximate covariance matrix of the parameter estimates, evaluated at the MLE is positive definite. If the Hessian is positive definite (e.g., all eigenvalues are positive) at the point 𝐱 , then 𝐱 is a local minimum of 𝑓 .This condition should be checked by applying an eigenvalue analysis to the estimated covariance matrix to verify that all of the eigenvalues are positive.

Other good practices include calculating the condition number of the covariance matrix, which is the ratio of the largest to the smallest eigenvalue. In this case, a very large condition number indicates that there is ill-conditioning, or collinearity in the calculated covariance matrix, which in turn, indicates that the parameter estimates are relatively imprecise. It is also very important to check whether the calculated MLE solution is consistent with a global minimum of the negative loglikelihood function. This can be accomplished in part by randomizing the initial parameter vector used to start the optimization process and rerunning the optimization algorithm, a technique sometimes called “jittering”. Note that the jittering process should be applied enough times to check if a better solution can be found (i.e., the first solution is not at a local minima). In complex assessments, one should check that if a new better solution is achieved, that the components of the likelihood that contributed to the finding are understood and plausible. Finally, such an exercise can avoid solutions in local minima, but there is no guarantee that the new solution is the global minimum.Here it is also important to note that the jitter analysis is a confirmatory analysis for model convergence.

Frequentist example: ISC Billfish Working Group. 2018. Stock Assessment for Swordfish *(Xiphias gladius*) in the Western and Central North Pacific Ocean through 2016. ISC18/Annex 16/ SAR. Available at: <http://isc.fra.go.jp/pdf/ISC18/ISC_18_ANNEX_16_Stock_Assessment_of_WCNPO_Swordfish_through_2016_FINAL.pdf>

Bayesian assessment models are built around the joint likelihood of the observed input data and the prior (or hyperprior) distributions of model parameters in a subjectivist interpretation of probability. The model parameters include observation error and also process error to describe system dynamics. This is the second primary approach used for optimizing parameters of stock assessment models in the 1990s based on calculating the joint posterior distribution of the vector of model parameters. Here the advent of Markov chain Monte Carlo methods to generate samples from the joint posterior distribution was a major advance in computational techniques to solve high dimensional Bayesian models. In this context, the Bayesian paradigm is to use numerical integration to solve the statistical optimization problem, instead of using numerical differentiation as in a frequentist framework, with the goal of computing a sequence of samples from the joint posterior distribution. That said, there may be limited practical difference between estimates calculated from a Bayesian model with uninformative priors or from a frequentist model because in both cases it is the joint loglikelihood of the observed data that drives the estimation process. The distinction between a Bayesian and frequentist approach is further lessened when the frequentist approach is expanded to include penalized loglikelihood models, which are effectively equivalent to parameter priors.

Posterior distributions for Bayesian assessment models can be calculated by numerically sampling the posterior distribution. For conjugate Bayesian models, exact solutions are available in some simple cases but this does not apply to nonlinear assessment models. The two primary approaches to sampling the posterior distribution of Bayesian models are by Markov chain Monte Carlo (MCMC) algorithms, of which there are a variety of types, or by approximate numerical integration of the posterior through the integrated nested Laplace approximation, or INLA. In both cases, it is useful to calculate the gradient of the model parameters to evaluate whether a stationary point has been achieved at the highest posterior density (HPD) parameter vector. This can be accomplished via MCMC by first maximizing the objective function comprised of the joint loglikelihood and prior distributions to calculate the HPD estimate and then using the HPD estimate as the initial vector in the iterative MCMC sampling process. In this case, one can apply the same approach as in a frequentist model for first and second order conditions using the HPD estimate and approximate covariance matrix. Here it also is important to use thinning to remove autocorrelation in successive parameter estimates sampled using MCMC.

Checking for convergence using MCMC can be accomplished using a number of approaches, including visual inspection of trace plots and density plots of MCMC iterates of parameters. Here the trace plots provide a quick initial check of whether successive iterates appear to be independent and identically distributed random samples while the density plots provide a quick check as to whether there is evidence of unimodality in the marginal densities of parameters. More formal convergence checks should also be applied and these typically require running more than one chain, or MCMC sequence of parameter estimates generated from different initial parameter vectors, in parallel. Here the R package “coda” has typically provided a simple means to produce some standard convergence tests with computational efficiency. The coda package can be used for application of standard diagnostics for the Gelman Rubin, Geweke, and Heidelberger and Welch,tests for convergence of an MCMC algorithm. It is also recommended to check the autocorrelations of MCMC sequences and apply posterior predictive checks to one or more key model parameters using the Bayesian P-value approach. This provides another means to assess whether the estimated marginal distribution of the parameter is consistent with a resampling of the observed data. One key issue with MCMC algorithms is that they can be slow to converge in stock assessment applications, a situation that can arise because there is a lack of curvature near to HPD estimate on the likelihood surface. In this context, it is notable that the “No U-Turn Sampler”, or NUTS can provide faster convergence for sampling the joint posterior and that this sampler is available in more recent versions of the ADMB and STAN modeling platforms. In contrast, the application of the integrated nested Laplace approximation provides a direct calculation of an HPD estimate and its approximate covariance for a Bayesian model with Gaussian error structures. This approach provides a faster calculation of an approximate posterior distribution but convergence in this case should also be checked by verifying first and second order conditions. It is also useful to simultaneously check whether convergence is supported by MCMC diagnostics applied to the sample the posterior distribution noting that INLA provides an approximate, and not exact HPD solution, and that the approximation is conditioned on Gaussian errors.

Bayesian Example: Langseth, B., J. Syslo, A. Yau, M. Kapur, and J. Brodziak. 2018. Stock Assessment for the Main Hawaiian Islands Deep 7 Bottomfish Complex in 2018, with Catch Projections Through 2022. NOAA Tech. Memo. NMFS-PIFSC-69, 218 p. Available at: <https://repository.library.noaa.gov/view/noaa/17252>

Random effects assessment models are built around the joint likelihood of the observed input data and the random effects distributions of model parameters in a hybrid interpretation of probability. Random effects models can be formulated for estimation using either a frequentist or a Bayesian framework, although we note that the former has been used in assessment models to date. This is the third primary approach used for optimizing parameters of stock assessment models developed in the 2010s based on calculating either the MLE or joint posterior distribution of the vector of model parameters. In general, assessment models that incorporate random effects can be expected to include random effects parameters for unobserved system state variables and fixed effects parameters for catch scaling, catchability, correlation and variance parameters in a state space modeling framework. Under the frequentist estimation approach, the estimation of the fixed effects parameters is based on optimizing the marginal likelihood. Here the marginal likelihood is calculated by integrating out the random effects from the joint likelihood for observation and process errors of the system dynamics. This integral can be calculated using an INLA approximation or using an MCMC sampling scheme. The random effects parameters are then predicted by maximizing the joint likelihood given the MLE of the fixed effects. Under the Bayesian estimation approach, the random effects model is set up to include prior distributions for model parameters with the goal of calculating the joint posterior distribution, or complete data likelihood, of both fixed and random effects. In this context, the approximation of the posterior distribution would be accomplished using an MCMC sampling approach given that both the fixed and random effects are random variables under the subjectivist interpretation. In general we note that parameter optimization under the random effects framework is based on a hierarchical statistical model wherein parameters are, in effect, estimated based on the observed data. This provides a more flexible treatment of hyperparameters for priors, which are estimated from data, in contrast to being assumed as in the Bayesian framework.

Stock assessment models that use random effects have been fitted primarily using the frequentist estimation approach to date. In this case, the integrated nested Laplace approximation has been applied to calculate and maximize the marginal likelihood. In terms of convergence, the maximization of the joint likelihood using the plug-in estimate of the MLE for the fixed effects is a calculation that should satisfy the first and second order conditions required for optimizing an objective function under the frequentist framework. Alternatively, generating representative samples from the joint posterior distribution should satisfy the convergence requirements of MCMC-based sampling under the Bayesian framework. In either case, convergence of the parameters of a random effects model can be confirmed using jittering or randomization of the initial parameter estimate combined with rerunning the optimization model and checking whether a more optimal solution is found.

Random Effects Example: Nielsen, A., Berg, C. 2014. Estimation of time-varying selectivity in stock assessments using state-space models. Fisheries Research. 158: 96-101. Available at: <https://reader.elsevier.com/reader/sd/pii/S0165783614000228?token=494E80C76B5F2D042822DF8AD02F1B14B45E501365135DEF55B5366C0321A482CAAC121D4EBFC381AB5CE689352F0255>

# Alternative initial parameter values (i.e., jitter run)

**Goal**: Ensure that the model converges upon a global solution regardless of the starting values used. The objective function for a stock assessment model may have more than one local minima. It is important to evaluate whether the calculated solution that minimizes the nonlinear objective function is sensitive to the initial parameter values.

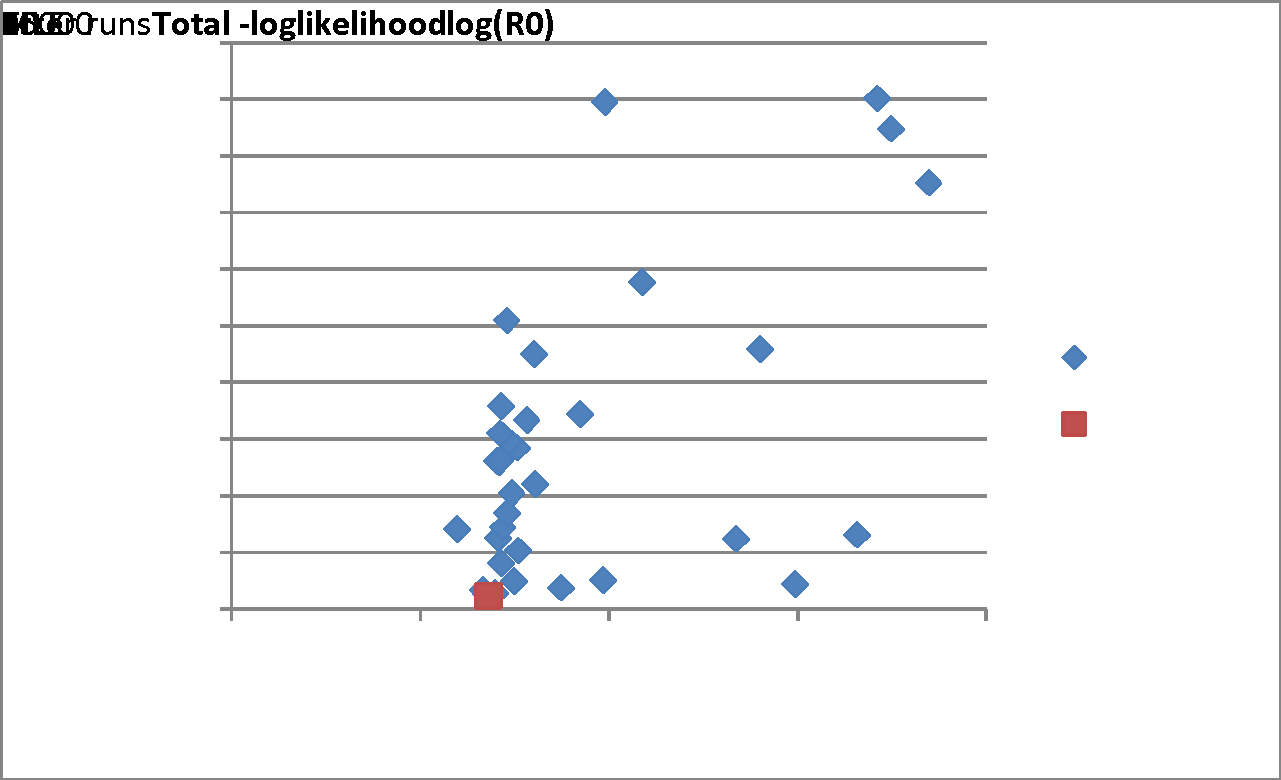
**Description:** Non-linear models need starting values and sometimes it is not trivial to generate these values. Poor choices can lead to non-convergence or convergence upon a local rather than global minimum. Starting the model using alternative but wise starting values can provide evidence that the model estimates are based on the global minimum if all values lead to the same likelihood.

**How to:** Change the initial values used for all estimated parameters and re-fit the model. Typically, this is done 50-100 times where the new initial values are chosen randomly given a range based on the original initial value and plus and minus a given fraction. Stock Synthesis offers internal functionality to do this and r4ss offers functionality to perform the analysis and summarize the results.

**What to do:**

1. Summarize the total likelihood values from each run.
2. Look at which likelihood components are changing if any are.
3. Evaluate the phases used for estimation.
4. If all runs converge upon a similar solution, consider a wider range of investigated starting values.

**Example of the randomized initial condition convergence, or jitter test from the stock assessment of Western and Central North Pacific Swordfish in 2018:**



Results of a randomized initial parameter value diagnostic for the base case model where 100 randomized initial conditions were used with a CV of 20% assigned to each parameter. Results are shown for the base case model (MLE, solid red square) and for the base case model with randomized initial parameter values (Jitter runs, solid blue diamond) that had a fitted total negative log-likelihood value of less than 9,000.

**Recommendations:** Move to Bayesian estimation where essentially every draw is a jitter run that is used in creating credible intervals rather than being runs that are essentially ignored. Jitter analyses are relevant and can be applied in any estimation framework used for a nonlinear stock assessment model. Sensitive dependence of the calculated optimal solution to initial parameter values should be evaluated for frequentist, Bayesian or random effects assessment models to ensure that the numerical results are robust.

**Key Literature:** Seber, G., Wild, C. 1989. Nonlinear regression. Wiley and Sons, New York, 768 p.

# Convergence Criteria: First Order Conditions

**Goal**: Optimization of a nonlinear objective function for a stock assessment model generally requires the iterative calculation of numerical estimates of the parameter vector. The convergence of the iterative sequence of estimates needs to be evaluated to ensure that a possible solution has been obtained.

**Description:** Describe the first order condition required to solve the numerical optimization of a nonlinear objective function. This condition requires that the first derivative, or gradient, of the nonlinear objective function be a stationary point at the calculated solution.

**How to:** Calculate the gradient of the objective function for each iterative estimate of the solution. Verify that the gradient is effectively equal to the zero vector at the final solution.

**What to do:**

1. Select a convergence criterion as a small positive constant epsilon, where epsilon is on the order of 10-4 or smaller.
2. For the current iterative estimate of the parameter vector indexed by i, calculate the gradient vector G(i).
3. Calculate the maximum absolute value of the components of the gradient G(i), call this max|G(i)|.
4. Check if max|G(i)| < epsilon. If this condition is true, then the first order condition is satisfied.

Note that other convergence criteria can be used for a given epsilon, such as: (i) the absolute difference between objective function values evaluated at successive parameter estimates, (ii) the Euclidean distance of the current gradient vector from the zero vector, or the Euclidean distance between successive parameter estimates. However, the use of the maximum absolute value of the components of the gradient G(i) is the natural mathematical condition to test, also known as the supremum norm.

**Example (include figures here):**

**Recommendations:** Do this calculation for frequentist models or for random effects models estimated in a frequentist framework when evaluating convergence to the MLE. Do this calculation for determining the HPD estimate to set the initial parameter vector in a Bayesian estimation framework.

**Key Literature:** Kennedy, W., Gentle, J. 1980. Statistical computing. Marcel Dekker, New York, 591 p.

# Convergence Criteria: Second Order Conditions

**Goal**: Optimization of a nonlinear objective function for a stock assessment model requires the calculation of a stationary point where the gradient of the objective function is zero. The second order condition for the existence of a local minimum at the stationary point is that the estimated covariance matrix is positive definite at the proposed solution.

**Description:** Describe the second order condition required for a stationary point to be a local minimum of a nonlinear objective function. This condition requires that the covariance matrix of the nonlinear objective function be positive definite at the calculated solution.

**How to:** Calculate the inverse of the Hessian matrix of the objective function, or estimated covariance matrix, at the proposed solution. Verify that the covariance is positive definite and well-conditioned.

**What to do:**

1. Calculate the solution that minimizes the objective function.
2. Calculate the Hessian matrix at the solution.
3. Calculate the covariance matrix A as the inverse of the Hessian matrix.
4. Apply eigenvalue analysis to the covariance matrix using the Cholesky or singular value decomposition.
5. If all of the eigenvalues of the symmetric matrix A are positive values, then the solution is a strict local minimum of the objective function, else it is not a strict local minimum.
6. Calculate the condition number k of the matrix A as the ratio of the largest to the smallest eigenvalue. Values of k that are large or on the order of 109 indicate that there is substantial collinearity in the estimated covariance matrix and that the calculated solution may be biased or imprecisely determined.

**Example (include figures here):**

**Recommendations:** Do this calculation for frequentist models or for random effects models estimated in a frequentist framework to verify that a possible solution is at least a strict local minimum of the objective function. Do this calculation for determining the covariance matrix at the HPD estimate positive definite for MCMC application in a Bayesian estimation framework.

**Key Literature:** Kennedy, W., Gentle, J. 1980. Statistical computing. Marcel Dekker, New York, 591 p.