Sensitivity Equations Provide More Robust Gradients and Faster Computation of the FOCE Approximation to the Population Likelihood

Joachim Almquist1,2, Jacob Leander1,3,* and Mats Jirstrand1
1Fraunhofer-Chalmers Centre, Göteborg, 2Department of Chemical and Biological Engineering, Chalmers University of Technology, Göteborg
3Department of Mathematical Sciences, Chalmers University of Technology, Göteborg, *Current affiliation AstraZeneca R&D, Mölndal

Background
The first order conditional estimation (FOCE) method [1] is still one of the parameter estimation workhorses for nonlinear mixed effects (NLME) modeling used in population pharmacokinetics and pharmacodynamics [2]. We propose a novel implementation of the FOCE and FOCEI methods where instead of obtaining the gradients needed for the two levels of quasi-Newton optimizations from the standard finite difference approximation, gradients are computed using so called sensitivity equations [3].

The Approximate Population Likelihood
The state-space model for a single individual is described by a system of ordinary differential equations and a corresponding set of measurement equations

\[ \frac{dx(t)}{dt} = f(x(t), \theta, \eta_j) \]

where \( x(t) \) and \( \eta_j \) are the state and individual-specific parameters, respectively. The log-likelihood function for a single individual \( i \) is given by

\[ \log L_i(\theta) = \frac{1}{2} \sum_{j=1}^{N} \left( \frac{1}{2} \log \det(2\pi \Omega) \right) - \frac{1}{2} \eta_j^T \Omega^{-1} \eta_j - \frac{1}{2} \log \det(2\pi \Omega) \]

The Outer Optimization Problem
The outer optimization problem consists of finding the \( \theta \) that maximizes the log-likelihood. The \( m \)th component of the gradient of the log-likelihood wrt \( \theta \)

\[ \frac{d \log L}{d \theta_m} = \sum_{i=1}^{N} \frac{d \log L_i}{d \theta_m} \]

where the total derivatives of \( l_i \) and \( H_i \) wrt \( \theta \) can be expressed in terms of solutions to sensitivity differential equations, e.g.,

\[ \frac{d \eta_i}{d \theta_m} = \frac{d \eta_i}{d \theta_m} \]

The sensitivity differential equations wrt \( \theta_m \)

\[ \frac{d}{dt} \frac{d \eta_i}{d \theta_m} = \frac{d}{dt} \frac{d \eta_i}{d \theta_m} \]

How to find \( \frac{d \eta_i}{d \theta_m} \)

\[ \frac{d \eta_i}{d \theta_m} = 0 \text{ if } \frac{d}{dt} \frac{d \eta_i}{d \theta_m} = 0 \]

Second order sensitivities are also required: \( \frac{d^2 \eta_i}{d \theta_m} \) and \( \frac{d^2 \eta_i}{d \theta_m} \).

The Inner Optimization Problem
The inner optimization problem consists of finding the \( \eta_j \) that maximize the individual \( l_i \) (for a given \( \theta \)). Gradient based optimization methods need accurate gradients. The \( k \)th component of the gradient of the log-likelihood wrt \( \eta_j \)

\[ \frac{d l_i}{d \eta_k} = \frac{1}{2} \sum_{j=1}^{N} \left( c^T \Omega^{-1} c \right) \frac{d l_i}{d \eta_k} \]

where \( \frac{d l_i}{d \eta_k} = \frac{d l_i}{d \eta_k} \)

The sensitivity differential equations wrt \( \eta_k \)

\[ \frac{d}{dt} \frac{d \eta_k}{d \theta_m} = \frac{d}{dt} \frac{d \eta_k}{d \theta_m} \]

Starting Values for Random Parameters
Using that \( \eta_i = \eta_i(\theta) \) is a function of \( \theta \) and that we have \( \frac{d \eta_i}{d \theta} \) give improved starting values of the inner optimization problem

\[ \eta_{i+1} = \eta_i + \frac{d \eta_i}{d \theta} \theta_{i+1} \]

Acknowledgements
This project has in part been supported by the Swedish Foundation for Strategic Research, which is gratefully acknowledged.

References