J Pharmacokinet Pharmacodyn
DOI 10.1007/s10928-007-9060-6

Derivation of various NONMEM estimation methods

Yaning Wang

Received: 11 April 2006 / Accepted: 19 April 2007

Abstract Various estimation methods and the lack of a systematic derivation of the core objective function implemented in NONMEM for nonlinear mixed effect modeling has caused consistent confusion and inquiry among scientists who routinely use NONMEM for data analysis. This paper provides a detailed derivation of the objective functions for the most commonly used estimation methods in NONMEM, such as the Laplacian method, the first-order conditional estimation method (FOCE) with or without interaction, and the first-order method (FO). In addition, models with homogenous or heterogeneous residual error were used to demonstrate the relationship between the objective functions derived from two different types of approximation, namely Laplacian approximation of log-likelihood and linearized model approximation. The relationship between these estimation methods and those implemented in SAS and Splus is discussed.

Keywords Nonlinear mixed-effects, Likelihood approximation, Laplacian, First-order conditional method (FOCE), First-order method (FO)

Introduction

Nonlinear mixed effect models (NLMEM) have become a popular platform in the field of population pharmacokinetic/pharmacodynamic data analysis since the pioneering work by Sheiner et al. [1] and Beal and Sheiner [2]. As the first software for nonlinear mixed effect modeling, NONMEM has evolved from version I to its current version VI [3]. Despite the emergence of other general-purpose software, such as Splus and

Y. Wang (✉)
Food and Drug Administration, Office of Clinical Pharmacology, CDER, WO21 RM3662 HFD-880,
10903 New Hampshire Avenue, Silver Spring, MD 20993, USA
Tel.: +1-301-7961624, e-mail: yaning.wang@fda.hhs.gov

SAS, and new pharmaceutical-specific software, such as WinNonmix and Kinetica, that incorporate NLMEM, NONMEM still remains the most popular tool in the pharmaceutical industry for population pharmacokinetic/pharmacodynamic data analysis because of its superior flexibility in the pharmaceutical field [4]. Even though thousands of pharmaceutical scientists utilize NONMEM routinely for their data analysis, the various estimation methods implemented in NONMEM remain a mystery for most of these scientists due to the complex statistical and mathematical derivations underlying the algorithm used in NONMEM. Several extensive derivations for the various estimation methods for NLMEM have been published in the statistical literature [5–8]. None of them, however, can be directly applied to obtain the core likelihood function described in the NONMEM Users Guide Part VII [3], triggering confusion and inquiry among those scientists who are eager to know the underlying mathematics and statistics for further understanding of NONMEM estimation algorithms and better application of various estimation methods under different scenarios. This paper gives the detailed derivation for the most commonly used estimation methods in NONMEM, namely the Laplacian method, the first-order conditional method (FOCE) with or without interaction, and the first-order method (FO).

Estimation methods in NONMEM

To be consistent with the NONMEM terminology, the following variables and parameters are defined:

y_i: the response vector for ith subject
ψ: mean parameter (θ) and residual variance-covariance matrix (Σ)
Ω: inter-individual variance-covariance matrix
η_i: random vector with mean $\mathbf{0}$ and variance-covariance matrix Ω (inter-individual variability)
ε_i: random vector with mean $\mathbf{0}$ and variance-covariance matrix Σ, which could be as simple as $\sigma^2 I$ (intra-individual variability)

The difficulty of NLMEM lies in the fact that the marginal density of y_i, $P(y_i|\psi, \Omega)$, or the marginal likelihood of ψ and Ω for the data y_i, $L_i(\psi, \Omega|y_i)$, is an integral that is difficult to compute exactly.

$$\begin{aligned} P(y_i|\psi, \Omega) = L_i(\psi, \Omega|y_i) &= \int P(y_i, \eta_i|\psi, \Omega) d\eta_i \\ &= \int P(y_i|\eta_i, \psi, \Omega) \cdot P(\eta_i|\psi, \Omega) d\eta_i \end{aligned} \tag{1}$$

Since $y_i|\eta_i, \psi$ does not involve any parameter in Ω and $\eta_i|\Omega$ does not involve any parameter in ψ, Eq. 1 can be further simplified to be

$$P(y_i|\psi, \Omega) = L_i(\psi, \Omega|y_i) = \int P(y_i|\eta_i, \psi) \cdot P(\eta_i|\Omega) d\eta_i \tag{2}$$

where $P(y_i|\eta_i, \psi)$ is the conditional density of y_i given the random subject effect η_i and $P(\eta_i|\Omega)$ is the density of η_i. In NONMEM, $P(y_i|\eta_i, \psi)$ is expressed as $l_i(\eta; \psi)$ and $P(\eta_i|\Omega)$ is expressed as $h(\eta; \Omega)$. Therefore

$$L_i(\psi, \Omega|y_i) = \int l_i(\eta; \psi) \cdot h(\eta; \Omega) d\eta \tag{3}$$

In the later derivation, $l_i(\eta; \psi)$ and $h(\eta; \Omega)$ are sometimes simplified to be l_i and h, respectively.

The conditional density of η_i is then

$$P(\eta_i|\psi, \Omega, y_i) = \frac{P(\eta_i, y_i|\psi, \Omega)}{P(y_i|\psi, \Omega)} = \frac{l_i(\eta_i; \psi) \cdot h(\eta_i; \Omega)}{L_i(\psi, \Omega|y_i)} \propto l_i(\eta_i; \psi) \cdot h(\eta_i; \Omega) \tag{4}$$

When ψ and Ω are replaced with their maximum likelihood estimators (MLE), Eq. 4 is called the empirical Bayes posterior distribution of η_i.

In order to apply the maximum likelihood estimation approach, different approximation methods are applied for the marginal likelihood (Eq. 3). In this paper, we will focus on the Laplacian method, the first-order conditional method (FOCE) with or without interaction, and the first-order method (FO) implemented in NONMEM.

Laplacian approximation (First level of approximation)

Given a complex integral, $\int f(x)dx$ (where $f(x) > 0 \forall x \psi$ in the integration range and $f(\cdot)$ is twice differentiable), $f(x)$ can be re-expressed as $e^{\log f(x)} = e^{g(x)}$. If $g''(x_0)$ is negative (scalar case) or negative definite (matrix case), $g(x)$ can be approximated by a second-order Taylor expansion of $g(x)$ about a point x_0

$$g(x) \approx g(x_0) + (x - x_0)g'(x_0) + \frac{(x - x_0)^2}{2!} g''(x_0)$$

The approximate integration is called a first order Laplacian approximation to the true integration.

If x_0 is $\hat{x}$, the mode of $f(x)$, the second term will be zero since $g'(x_0) = 0$. Then $\int f(x)dx = \int e^{g(x)}dx \approx \int e^{g(x_0)+\frac{(x-x_0)^2}{2!}g''(x_0)}dx = f(x_0) \cdot \sqrt{\frac{2\pi}{-g''(x_0)}}$ (see Appendix 1 for detailed derivation).

But if x_0 is not the mode, the result will be

$$\int f(x)dx = \int e^{g(x)}dx \approx \int e^{g(x_0)+(x-x_0)g'(x_0)+\frac{(x-x_0)^2}{2!}g''(x_0)}dx$$
$$= f(x_0) \cdot \int e^{(x-x_0)g'(x_0)+\frac{(x-x_0)^2}{2!}g''(x_0)}dx = f(x_0) \cdot \sqrt{\frac{2\pi}{-g''(x_0)}} \cdot e^{\frac{-g'(x_0)^2}{2g''(x_0)}}$$

(see Appendix 2 for detailed derivation)

If x is a vector,

$$\int f(x)dx = f(x_0) \cdot \sqrt{\frac{(2\pi)^p}{|-g''(x_0)|}} e^{\frac{-g'(x_0)^T g''(x_0)^{-1} g'(x_0)}{2}} \tag{5}$$

where p is the dimension of x.

In NONMEM,

$$-2 \log l_i(\eta; \psi) = \Phi_i(\eta)$$

$$\Gamma_i(\eta) = [-2 \log l_i(\eta; \psi)]' = \frac{\partial[-2 \log l_i(\eta; \psi)]}{\partial \eta} = -2\frac{l_i'}{l_i} \tag{6}$$

$$\Delta_i(\eta) = [\Gamma_i(\eta)]' = \frac{\partial \Gamma_i(\eta)}{\partial \eta} = -2\left(\frac{l_i'}{l_i}\right)' \tag{7}$$

$\Gamma_i(\eta)$ and $\Delta_i(\eta)$ are the gradient vector and the Hessian matrix of Φ_i evaluated at η. Notice that these two entities are a little different from the gradient vector (6a) and the Hessian matrix (7a) of log-likelihood (conditional likelihood of η_i).

$$G_i(\eta) = [\log l_i(\eta; \psi)]' = -\frac{\Gamma_i(\eta)}{2} \tag{6a}$$

$$H_i(\eta) = [G_i(\eta)]' = -\frac{\Delta_i(\eta)}{2} \tag{7a}$$

From Eqs. 6 and 7, we can obtain

$$\frac{l_i'}{l_i} = \frac{\Gamma_i(\eta)}{-2} \tag{8}$$

$$\left(\frac{l_i'}{l_i}\right)' = \frac{\Delta_i(\eta)}{-2} \tag{9}$$

And we know

$$h = \frac{1}{(2\pi)^{\frac{p}{2}} |\Omega|^{\frac{1}{2}}} \cdot e^{-\frac{\eta^T \Omega^{-1} \eta}{2}} \quad \text{and}$$

$$h' = \frac{\partial h}{\partial \eta} = \frac{1}{(2\pi)^{\frac{p}{2}} |\Omega|^{\frac{1}{2}}} \cdot e^{-\frac{\eta^T \Omega^{-1} \eta}{2}} \cdot \left(-\frac{2\Omega^{-1}\eta}{2}\right)$$

Therefore

$$\frac{h'}{h} = -\Omega^{-1}\eta \tag{10}$$

$$\left(\frac{h'}{h}\right)' = -\Omega^{-1} \tag{11}$$

Since the integrand in Eq. 3 is $f(\eta) = l_i(\eta;\psi)\cdot h(\eta;\Omega)$, we know

$$g(\eta) = \log l_i(\eta;\psi) + \log h(\eta;\Omega)$$

and

$$g'(\eta) = \frac{l_i'}{l_i} + \frac{h'}{h} = -\frac{\Gamma_i(\eta)}{2} - \Omega^{-1}\eta \quad \text{(12, apply 8 and 10)}$$

$$g''(\eta) = \left(\frac{l_i'}{l_i}\right)' + \left(\frac{h'}{h}\right)' = -\frac{\Delta_i(\eta)}{2} - \Omega^{-1} \quad \text{(13, apply 9 and 11)}$$

The objective function defined in NONMEM is then

$$\begin{aligned}
&-2\log L_i(\psi,\Omega|y_i) = -2\log\int l_i(\eta;\psi)\cdot h(\eta;\Omega)d\eta\\
&\approx -2\log\left\{l_i(\eta_{i0};\psi)\cdot h(\eta_{i0};\Omega)\sqrt{\frac{(2\pi)^p}{|-g''(\eta_{i0})|}}\cdot e^{\frac{-g'(\eta_{i0})^T g''(\eta_{i0})^{-1}g'(\eta_{i0})}{2}}\right\} \quad \text{(apply 5)}\\
&= -2\log l_i(\eta_{i0};\psi) - 2\log h(\eta_{i0};\Omega) - 2\log\sqrt{\frac{(2\pi)^p}{|-g''(\eta_{i0})|}} - 2\log e^{\frac{-g'(\eta_{i0})^T g''(\eta_{i0})^{-1}g'(\eta_{i0})}{2}}\\
&= \Phi_i(\eta_{i0}) - 2\log\frac{1}{(2\pi)^{\frac{p}{2}}|\Omega|^{\frac{1}{2}}} - 2\log e^{-\frac{\eta_{i0}^T\Omega^{-1}\eta_{i0}}{2}}\\
&\quad -\log\frac{(2\pi)^p}{|-g''(\eta_{i0})|} + g'(\eta_{i0})^T g''(\eta_{i0})^{-1}g'(\eta_{i0})\\
&= \Phi_i(\eta_{i0}) + \log|\Omega| + \eta_{i0}^T\Omega^{-1}\eta_{i0} + \log\left|-g''(\eta_{i0})\right| + g'(\eta_{i0})^T g''(\eta_{i0})^{-1}g'(\eta_{i0})\\
&= \Phi_{i0} + \log|\Omega| + \eta_{i0}^T\Omega^{-1}\eta_{i0} + \log\left|\Omega^{-1} + \frac{\Delta_{i0}}{2}\right|\\
&\quad -\left(\frac{\Gamma_{i0}}{2} + \Omega^{-1}\eta_{i0}\right)^T\left(\Omega^{-1} + \frac{\Delta_{i0}}{2}\right)^{-1}\left(\frac{\Gamma_{i0}}{2} + \Omega^{-1}\eta_{i0}\right)
\end{aligned} \tag{14}$$

η_{i0} is some estimate of η_i. Φ_{i0}, Δ_{i0}, and Γ_{i0} are Φ_i, Δ_i, and Γ_i evaluated at η_{i0}. This is the core objective function in NONMEM user guide part VII. If η_{i0} is the mode of [4], $\hat{\eta}_i$, the last term in equation above is zero since$g'(\hat{\eta}_i)$ is 0. And

$-2 \log L_i(\psi, \Omega|y_i)$ is approximated by

$$\hat{\Phi}_i + \log|\Omega| + \hat{\eta}_i' \Omega^{-1} \hat{\eta}_i + \log \left| \Omega^{-1} + \frac{\hat{\Delta}_i}{2} \right| \tag{15}$$

The sum over i of Eq. 15 is the Laplacian objective function that is minimized in NONMEM. Notice that $\hat{\eta}_i$ is dependent on ψ and Ω since it is the mode of Eq. 4 assuming ψ and Ω are known. In other words, $\hat{\eta}_i$ is a function of ψ and Ω. Therefore, objective function 15 is a function of only ψ and Ω given y_i even though the dependence of $\hat{\eta}_i$ on ψ and Ω is not obvious.

FOCE (Second level of approximation)

Due to the difficulty for the direct computation of second-order derivatives, The Hessian matrix is approximated by a function of the gradient vector

$$\Delta_i(\eta) \approx \frac{1}{2} E \left(\Gamma_i(\eta) \Gamma_i(\eta)^T \right)$$

which is called the first order approximation in NONMEM. E represents the expectation over y_i under the intraindividual model. This approximation could be derived from the well known equality

$$-E\left(H_i(\eta)\right) = E\left(G_i(\eta) G_i(\eta)^T\right).$$

Substituting (6a) and (7a) into this equation and we get

$$E\left(\Delta_i(\eta)\right) = \frac{1}{2} E\left(\Gamma_i(\eta) \Gamma_i(\eta)^T\right).$$

Therefore $\Delta_i(\eta)$ is basically approximated by its expectation.

Then $-2 \log L_i(\psi, \Omega|y_i)$ is approximated by

$$\hat{\Phi}_i + \log|\Omega| + \hat{\eta}_i' \Omega^{-1} \hat{\eta}_i + \log \left| \Omega^{-1} + \frac{\hat{E}\left(\Gamma_i(\eta) \Gamma_i(\eta)^T\right)}{4} \right| \tag{16}$$

where $\hat{E}\left(\Gamma_i(\eta)\Gamma_i(\eta)^T\right)$ means that after expectation is taken relative to y_i, η is replaced by $\hat{\eta}_i$. The sum over i of Eq. 16 is the objective function of FOCE. Pinheiro and Bates [8] applied a similar approximation to the Hessian matrix defined in their method (different from the Hessian matrix defined in NONMEM) by ignoring a term involving the second derivative of f(·). It can be easily shown that their approximation is equivalent to taking the expectation of their Hessian matrix relative to y_i. Therefore, these two types of approximation give the same likelihood (Appendix 4). And the modified Laplacian approximation discussed in [8] is simply NONMEM FOCE.

FO (Third level of approximation)

If $\Delta_i(\eta)$ is approximated by $\frac{1}{2}E\left(\Gamma_{i(k)}(\eta)\Gamma_{i(k)}(\eta)^T\right)$ and $\eta_{i0} = 0$, then $-2\log L_i(\psi, \Omega|y_i)$ is approximated by

$$\Phi_{i0} + \log|\Omega| + \log\left|\Omega^{-1} + \frac{E\left(\Gamma_i(\eta)\Gamma_i(\eta)^T\right)\big|_{\eta=0}}{4}\right| - \left(\frac{\Gamma_{i0}}{2}\right)^T\left(\Omega^{-1} + \frac{E\left(\Gamma_i(\eta)\Gamma_i(\eta)^T\right)\big|_{\eta=0}}{4}\right)^{-1}\left(\frac{\Gamma_{i0}}{2}\right) \tag{17}$$

since $\hat{\eta}_i^T\Omega^{-1}\hat{\eta}_i$ in Eq. 14 is zero but the last term is not zero. The sum over i of Eq. 17 is the objective function of FO.

In NONMEM, the constant in $\Phi_{i0}, n_i\log 2\pi$, is not included in the objective function evaluation for any estimation method, which is why the objective function value from NONMEM is not equal to that reported from SAS for the same model and data set.

Further discussion on additive and proportional intraindividual models

For an additive intraindividual model, $y_i = f_i(\eta_i, \theta) + \varepsilon_i$, the objective functions 16 and 17 can also be derived from a linearized model by taking a first order Taylor expansion on the nonlinear structural model around $\eta = \hat{\eta}_i$ for FOCE or $\eta = 0$ for FO. This equivalence also holds for proportional intraindividual model, $y_i = f_i(\eta_i, \theta) + f_i(\eta_i, \theta)\cdot\varepsilon_i$, if FOCE without interaction or FO method is used for estimation. If FOCE with interaction is used for a proportional intraindividual model, however, the objective function 16 is different from that derived based on a linearized model around $\eta = \hat{\eta}_i$.

To demonstrate the relationship between the Laplacian approximation based derivation and the model linearization based derivation, univariate random variables are used to simplify the algebra. The general matrix version can be found in Appendix 3 for those who are interested.

Additive intraindividual model

For $y_i = f(\eta_i, \theta) + \varepsilon_i$,
we assume

$$\begin{aligned} &y_i|\eta_i, \theta, \sigma^2 \sim N(f(\eta_i, \theta), \sigma^2)\\ &\varepsilon_i|\sigma^2 \sim N(0, \sigma^2)\\ &\eta_i|\omega^2 \sim N(0, \omega^2) \end{aligned}$$

The model can be linearized around $\eta = \hat{\eta}$ based on a first-order Taylor expansion if the FOCE method is used:

$$y_i = f(\hat{\eta}_i, \theta) + f'(\hat{\eta}_i, \theta)(\eta_i - \hat{\eta}_i) + \varepsilon_i$$

The marginal distribution for y_i is

$$y_i \sim N(f(\hat{\eta}_i, \theta) - f'(\hat{\eta}_i, \theta)\hat{\eta},\ f'(\hat{\eta}_i, \theta)^2\omega^2 + \sigma^2)$$

and

$$\begin{aligned} -2\log L_i(\psi, \Omega|y_i) = &\ \text{constant} + \log[f'(\hat{\eta}_i, \theta)^2\omega^2 + \sigma^2] \\ &+ \frac{[y_i - f(\hat{\eta}_i, \theta) + f'(\hat{\eta}_i, \theta)\hat{\eta}_i]^2}{f'(\hat{\eta}_i, \theta)^2\omega^2 + \sigma^2} \end{aligned} \tag{18}$$

Equation 18 can be shown to be the same as objective function 16.

$$\begin{aligned} \text{Since } \Gamma_i(\eta) &= -2\frac{l_i'(\eta)}{l_i(\eta)} \quad \text{and} \quad l_i(\eta) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{[y_i - f(\eta_i, \theta)]^2}{2\sigma^2}} \\ l_i'(\eta) &= \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{[y_i - f(\eta_i, \theta)]^2}{2\sigma^2}} \cdot \left(-\frac{2[y_i - f(\eta_i, \theta)]}{2\sigma^2}\right) \cdot \left(-f'(\eta_i, \theta)\right) \\ &= l_i(\eta) \cdot \left(\frac{[y_i - f(\eta_i, \theta)]}{\sigma^2}\right) \cdot f'(\eta_i, \theta) \\ \Gamma_i(\eta) &= -2\frac{l_i'(\eta)}{l_i(\eta)} = -2 \cdot \left(\frac{[y_i - f(\eta_i, \theta)]}{\sigma^2}\right) \cdot f'(\eta_i, \theta) \\ \Gamma_i(\hat{\eta}) &= -2 \cdot \left(\frac{[y_i - f(\hat{\eta}_i, \theta)]}{\sigma^2}\right) \cdot f'(\hat{\eta}_i, \theta) \end{aligned} \tag{19}$$

$$\begin{aligned} \Delta_i(\eta) &\approx \frac{1}{2} E\left(\Gamma_i(\eta)\Gamma_i(\eta)^T\right) = 2E\left(\frac{[y_i - f(\eta_i, \theta)]}{\sigma^2}\right)^2 f'(\eta_i, \theta)^2 \\ &= 2\frac{f'(\eta_i, \theta)^2}{\sigma^4} E[y_i - f(\eta_i, \theta)]^2 = 2\frac{f'(\eta_i, \theta)^2}{\sigma^4}\sigma^2 = 2\frac{f'(\eta_i, \theta)^2}{\sigma^2}. \end{aligned}$$

Hence,

$$\Delta_i(\hat{\eta}) \approx 2\frac{f'(\hat{\eta}_i, \theta)^2}{\sigma^2} \tag{20}$$

Objective function 16 can be simplified to be $\log 2\pi + \log \sigma^2 + \frac{[y_i - f(\hat{\eta}_i, \theta)]^2}{\sigma^2} + \log \omega^2 + \frac{\hat{\eta}_i^2}{\omega^2} + \log\left(\frac{1}{\omega^2} + \frac{f'(\hat{\eta}_i, \theta)^2}{\sigma^2}\right)$, which can be further simplified to Eq. 21 by dropping the constant $\log 2\pi$ without affecting the optimization:

$$\log[f'(\hat{\eta}_i,\theta)^2\omega^2+\sigma^2]+\frac{[y_i-f(\hat{\eta}_i,\theta)]^2}{\sigma^2}+\frac{\hat{\eta}_i^2}{\omega^2} \tag{21}$$

In order to show Eqs. 18 and 21 are equivalent, we only need to show the second term in Eq. 18 is equal to the sum of the 2nd and 3rd terms in Eq. 21. Since $g'(\hat{\eta}_i)=-\frac{\hat{\Gamma}_i}{2}-\Omega^{-1}\hat{\eta}_i$ and $g'(\hat{\eta}_i)=0$

$$\frac{\hat{\Gamma}_i}{2}+\Omega^{-1}\hat{\eta}=0 \tag{22}$$

Substituting Eq. 19 into Eq. 22, we get

$$-\left(\frac{[y_i-f(\hat{\eta}_i,\theta)]}{\sigma^2}\right)\cdot f'(\hat{\eta}_i,\theta)+\frac{\hat{\eta}_i}{\omega^2}=0 \text{ and}$$
$$\hat{\eta}_i=\left(\frac{[y_i-f(\hat{\eta}_i,\theta)]}{\sigma^2}\right)\cdot f'(\hat{\eta}_i,\theta)\cdot\omega^2 \tag{23}$$

Substituting Eq. 23 into the last term in Eq. 18, we have

$$\begin{aligned}
&\frac{[y_i-f(\hat{\eta}_i,\theta)+f'(\hat{\eta}_i,\theta)\hat{\eta}_i]^2}{f'(\hat{\eta}_i,\theta)^2\omega^2+\sigma^2}=\frac{[y_i-f(\hat{\eta}_i,\theta)+f'(\hat{\eta}_i,\theta)\cdot\frac{y_i-f(\hat{\eta}_i,\theta)}{\sigma^2}\cdot f'(\hat{\eta}_i,\theta)\cdot\omega^2]^2}{f'(\hat{\eta}_i,\theta)^2\omega^2+\sigma^2}\\
&=\frac{\{[y_i-f(\hat{\eta}_i,\theta)]\cdot(1+\frac{f'(\hat{\eta}_i,\theta)^2\cdot\omega^2}{\sigma^2})\}^2}{f'(\hat{\eta}_i,\theta)^2\omega^2+\sigma^2}=\frac{\{[y_i-f(\hat{\eta}_i,\theta)]\cdot\frac{\sigma^2+f'(\hat{\eta}_i,\theta)^2\cdot\omega^2}{\sigma^2}\}^2}{f'(\hat{\eta}_i,\theta)^2\omega^2+\sigma^2}\\
&=\frac{[y_i-f(\hat{\eta}_i,\theta)]^2}{\sigma^2}\cdot\frac{\sigma^2+f'(\hat{\eta}_i,\theta)^2\cdot\omega^2}{\sigma^2}\\
&=\frac{[y_i-f(\hat{\eta}_i,\theta)]^2}{\sigma^2}\left(1+\frac{f'(\hat{\eta}_i,\theta)^2\cdot\omega^2}{\sigma^2}\right)\\
&=\frac{[y_i-f(\hat{\eta}_i,\theta)]^2}{\sigma^2}+\frac{[y_i-f(\hat{\eta}_i,\theta)]^2}{\sigma^4}\cdot f'(\hat{\eta}_i,\theta)^2\cdot\omega^2 \quad \text{(substituting Eq. 23 again)}\\
&=\frac{[y_i-f(\hat{\eta}_i,\theta)]^2}{\sigma^2}+\frac{\hat{\eta}_i^2}{\omega^2}
\end{aligned}$$

Therefore, Eqs. 18 and 21 are the same objective function.

If the FO method is used, the model can be linearized around $\eta=0$ based on a first-order Taylor expansion:

$$y_i=f(0,\theta)+f'(0,\theta)\eta_i+\varepsilon_i$$

The marginal distribution for y_i is

$$y_i\sim N(f(0,\theta),\,f'(0,\theta)^2\omega^2+\sigma^2)$$

and

$$-2\log L_i(\psi,\Omega|y_i) = \text{constant} + \log[f'(0,\theta)^2\omega^2+\sigma^2] + \frac{[y_i - f(0,\theta)]^2}{f'(0,\theta)^2\omega^2+\sigma^2} \quad (24)$$

Equation 24 can be shown to be the same as objective function 17.
From Eqs. 19 and 20, we know

$$\Gamma_i(0) = -2\cdot\left(\frac{[y_i - f(0,\theta)]}{\sigma^2}\right)\cdot f'(0,\theta) \text{ and}$$
$$\Delta_i(0) \approx 2\frac{f'(0,\theta)^2}{\sigma^2}$$

Objective function 17 can be simplified to be

$$\begin{aligned}\Phi_i(0) &+ \log\omega^2 + \log\left(\frac{1}{\omega^2}+\frac{\Delta_i(0)}{2}\right) - \left(\frac{\Gamma_i(0)}{2}\right)^2\left(\frac{1}{\omega^2}+\frac{\Delta_i(0)}{2}\right)^{-1}\\ &= \log 2\pi + \log\sigma^2 + \frac{[y_i - f(0,\theta)]^2}{\sigma^2} + \log\omega^2 + \log\left(\frac{1}{\omega^2}+\frac{f'(0,\theta)^2}{\sigma^2}\right)\\ &\quad - \left(\frac{\Gamma_i(0)}{2}\right)^2\left(\frac{1}{\omega^2}+\frac{f'(0,\theta)^2}{\sigma^2}\right)^{-1}\end{aligned}$$

which ignoring the constant becomes

$$\begin{aligned}\log&\left(\sigma^2 + f'(0,\theta)^2\omega^2\right) + \frac{[y_i - f(0,\theta)]^2}{\sigma^2} - \frac{\left(\frac{[y_i - f(0,\theta)]}{\sigma^2}\right)^2 f'(0,\theta)^2}{\frac{1}{\omega^2}+\frac{f'(0,\theta)^2}{\sigma^2}}\\ &= \log\left(\sigma^2 + f'(0,\theta)^2\omega^2\right) + \frac{[y - f(0,\theta)]^2}{\sigma^2}\left[1 - \frac{\frac{f'(0,\theta)^2}{\sigma^2}}{\frac{1}{\omega^2}+\frac{f'(0,\theta)^2}{\sigma^2}}\right]\\ &= \log\left(\sigma^2 + f'(0,\theta)^2\omega^2\right) + \frac{[y - f(0,\theta)]^2}{\sigma^2 + f'(0,\theta)^2\omega^2} \qquad (25)\end{aligned}$$

Then Eq. 25 is identical to Eq. 24.

Proportional intraindividual model

For $y_i = f(\eta_i,\theta) + f(\eta_i,\theta)\cdot\varepsilon_i$,
we assume

$$\begin{aligned}&y_i|\eta_i,\theta,\sigma^2 \sim N(f(\eta_i,\theta),\, f(\eta_i,\theta)^2\sigma^2)\\ &\varepsilon_i|\sigma^2 \sim N(0,\sigma^2)\\ &\eta_i|\omega^2 \sim N(0,\omega^2)\end{aligned}$$

If FOCE without interaction or FO method is used, the original model is transformed to $y_i = f(\eta_i, \theta) + f(0, \theta) \cdot \varepsilon_i$ and the conditional distribution of y_i is $y_i | \eta_i, \theta, \sigma^2 \sim N(f(\eta_i, \theta), f(0, \theta)^2 \sigma^2)$. Then the Laplacian approximation based derivation and model linearization based derivation still yield equivalent objective functions. All the previous derivations remain the same except that σ^2 is replaced by $f(0, \theta)^2 \sigma^2$. In NONMEM V, the Laplacian method also transforms the original model to $y_i = f(\eta_i, \theta) + f(0, \theta) \cdot \varepsilon$ before any further calculation is performed. In other words, no interaction is automatically assumed for the Laplacian method in NONMEM V.

When the FOCE with interaction method is used, the Laplacian approximation based derivation and model linearization based derivation yield different objective functions. NONMEM uses the Laplacian approximation based derivation. Linearization of the structural model around $\eta = \hat{\eta}$ and replacement of $f(\eta_i, \theta) \cdot \varepsilon_i$ with $f(\hat{\eta}_i, \theta) \cdot \varepsilon_i$ yields

$$y_i = f(\hat{\eta}_i, \theta) + f'(\hat{\eta}_i, \theta) \cdot (\eta_i - \hat{\eta}_i) + f(\hat{\eta}_i, \theta) \cdot \varepsilon_i$$

The marginal distribution for y_i is

$$y_i \sim N(f(\hat{\eta}_i, \theta) - f'(\hat{\eta}_i, \theta)\hat{\eta}_i, f'(\hat{\eta}_i, \theta)^2 \omega^2 + f(\hat{\eta}_i, \theta)\sigma^2)$$

and

$$\begin{aligned} -2\log L_i(\psi, \Omega | y_i) = \text{constant} &+ \log[f'(\hat{\eta}_i, \theta)^2 \omega^2 + f(\hat{\eta}_i, \theta)^2 \sigma^2] \\ &+ \frac{[y_i - f(\hat{\eta}_i, \theta) + f'(\hat{\eta}_i, \theta)\hat{\eta}_i]^2}{f'(\hat{\eta}_i, \theta)^2 \omega^2 + f(\hat{\eta}_i, \theta)^2 \sigma^2} \end{aligned} \tag{26}$$

Equation 26 is different from objective function 16 as demonstrated below.

$$\begin{aligned} \text{Since } l_i(\eta) &= \frac{1}{\sqrt{2\pi} \cdot f(\eta_i, \theta)\sigma} e^{-\frac{[y_i - f(\eta_i, \theta)]^2}{2 f(\eta_i, \theta)^2 \sigma^2}} \\ l_i'(\eta) &= -l_i(\eta) \frac{f'(\eta_i, \theta)}{f(\eta_i, \theta)} + l_i(\eta) \frac{[y_i - f(\eta_i, \theta)] \cdot y_i \cdot f'(\eta_i, \theta)}{f(\eta_i, \theta)^3 \sigma^2} \\ \Gamma_i(\eta) &= -2\frac{l_i'(\eta)}{l_i(\eta)} = -2 * \left(\frac{[y_i - f(\eta_i, \theta)] \cdot y_i}{f(\eta_i, \theta)^3 \sigma^2} - \frac{1}{f(\eta_i, \theta)} \right) \cdot f'(\eta_i, \theta) \end{aligned} \tag{27}$$

$$\begin{aligned} \Delta_i(\eta) &\approx \frac{1}{2} E\left(\Gamma_i(\eta)\Gamma_i(\eta)^T\right) = 2E\left(\frac{[y_i - f(\eta_i, \theta)] \cdot y_i}{f(\eta_i, \theta)^3 \cdot \sigma^2} - \frac{1}{f(\eta_i, \theta)} \right)^2 f'(\eta_i, \theta)^2 \\ &= 2f'(\eta_i, \theta)^2 E\left(\frac{[y_i - f(\eta_i, \theta)] \cdot y_i - f(\eta_i, \theta)^2 \cdot \sigma^2}{f(\eta_i, \theta)^3 \cdot \sigma^2} \right)^2 \\ &= \frac{2f'(\eta_i, \theta)^2}{f(\eta_i, \theta)^2 \cdot \sigma^2} + \frac{4f'(\eta_i, \theta)^2}{f(\eta_i, \theta)^2} \end{aligned}$$

where as objective function 16 can be simplified to be

$$
\begin{aligned}
&\Phi\ (\hat{\eta}_i) + \log \omega^2 + \frac{\hat{\eta}_i^2}{\omega^2} + \log\left(\frac{1}{\omega^2} + \frac{\Delta_i(\hat{\eta})}{2}\right) \\
&\approx \log 2\pi + \log f(\hat{\eta}_i, \theta)^2 \sigma^2 + \frac{[y_i - f(\hat{\eta}_i, \theta)]^2}{f(\hat{\eta}_i, \theta)^2 \sigma^2} + \log \omega^2 + \frac{\hat{\eta}_i^2}{\omega^2} \\
&\quad + \log\left(\frac{1}{\omega^2} + \frac{f'(\hat{\eta}_i, \theta)^2}{f(\hat{\eta}_i, \theta)^2 \cdot \sigma^2} + \frac{2 f'(\hat{\eta}_i, \theta)^2}{f(\hat{\eta}_i, \theta)^2}\right) \\
&= \text{constant} + \log[f(\hat{\eta}_i, \theta)^2 \sigma^2 + f'(\hat{\eta}_i, \theta)^2 \omega^2 + 2 f'(\hat{\eta}_i, \theta) \sigma^2 \omega^2] \\
&\quad + \frac{[y_i - f(\hat{\eta}_i, \theta)]^2}{f(\hat{\eta}_i, \theta)^2 \sigma^2} + \frac{\hat{\eta}_i^2}{\omega^2}
\end{aligned} \tag{28}
$$

Comparing Eqs. 26 and 28, we can see that they are not equal. Therefore, when there is an interaction between intra- and inter-individual variability as in a proportional error model, the Laplacian approximation based method is different from the linearization method unless the interaction is ignored. Even though it is believed that the Laplacian method should perform no worse than the FOCE method (the former avoids the first-order approximation), this belief may not be true when there is an intra- and inter-individual variability interaction because, for the current version of NONMEM (version V), the Laplacian method ignores this interaction while FOCE does not.

In NONMEM, when the residual error model is not linear in ε, the model is first linearized around $\varepsilon = 0$ based on a first-order Taylor expansion. Therefore, for an exponential residual error model, $y_i = f(\eta_i, \theta)\exp(\varepsilon_i)$, the linearization around $\varepsilon = 0$ leads to $y_i = f(\eta_i, \theta) + f(\eta_i, \theta) \cdot \varepsilon_i$, the proportional residual error model. Therefore, these two residual error models are treated equally in NONMEM. All the above derivations also apply to an exponential residual error model.

A numerical example

To illustrate the objective function calculation discussed in "Further discussion on additive and proportional intraindividual models" section, a simulated data set was used (Table 1). The NONMEM code for fitting the data and Splus code for objective function calculation are included in Appendix 5. Table 2 lists the NONMEM reported results and calculated results based on the alternative expressions discussed in "Further discussion on additive and proportional intraindividual models" section.

Summary

For the first time, the objective functions for the major estimation methods used in NONMEM were systematically derived and their relationships were clearly demonstrated through the derivation. This paper can serve as a reference for understanding the underlying mathematics and statistics of nonlinear mixed effect models, especially in the context of NONMEM. Also the original approximation method based on a

Table 1 Simulated data set

ID	Time	Y
1	0	10.68
1	1	3.6837
2	0	10.402
2	1	6.454
3	0	9.8814
3	1	5.8565
4	0	9.3408
4	1	5.6209
5	0	10.082
5	1	6.7583
6	0	9.8938
6	1	6.5049
7	0	9.8908
7	1	6.9557
8	0	10.234
8	1	6.4488
9	0	9.9882
9	1	6.7112
10	0	9.6736
10	1	6.6402

Table 2 Comparison of NONMEM and calculated objective functions

	NONMEM	Calculated
Additive residual error		
FO	0.026	0.02580146
FOCE	−2.059	−2.058504
Proportional residual error		
FO	39.213	39.21323
FOCE without interaction	39.207	39.20673
FOCE with interaction	39.458	39.20108
Exponential residual error		
FO	39.213	39.21323
FOCE without interaction	39.207	39.20673
FOCE with interaction	39.458	39.20108

linearized model was shown to be equivalent to the new approximation method based on the Laplacian approximation of the log-likelihood when the residual error model is additive and the FOCE or FO method is used or when the residual error model is proportional and FOCE without interaction or the FO method is used. This equivalence does not hold when FOCE with interaction is used and the residual error model is proportional, which was also indicated by Vonesh and Chinchilli [7]. Even though the general belief is that the Laplacian method should perform no worse than the FOCE method because the former avoids the first-order approximation [3], the FOCE with interaction may be superior to the Laplacian method (no interaction) when there is intra- and inter-individual variability interaction as in a proportional residual error model. The option of the Laplacian method with interaction is available in NONMEM VI. Following the rationale as outlined in NONMEM Users Guide Part VII [3], the Laplacian with interaction should be the most accurate approximation method. More detailed discussion about the trade-off between approximation accuracy and computation time and numerical stability under various scenarios can be found in NONMEM Users Guide Part VII [3].

Even though the Laplacian method is the most accurate approximation method for the integral of the likelihood implemented in NONMEM, a more accurate approximation method, such as adaptive Gaussian approximation, is available and implemented in the SAS nlmixed procedure. One point adaptive Gaussian approximation, a special case of adaptive Gaussian approximation with only one quadrature point, is simply the Laplacian method in NONMEM, but without the constraint of assuming no interaction. Since NONMEM FOCE is equivalent to the modified Laplacian approximation discussed in [8], the Lindstrom and Bates Algorithm, implemented in Splus (nlme) is inferior to NONMEM FOCE in terms of approximation accuracy [8] even though nlme has the advantage of providing restricted maximum likelihood.

Acknowledgements The author would like to thank two anonymous referees and Leon Aarons for valuable inputs which greatly improved the manuscript.

Appendix

1. $\int e^{g(x_0)+\frac{(x-x_0)^2}{2!}g''(x_0)}dx = f(x_0)\cdot\sqrt{\frac{2\pi}{-g''(x_0)}}$

The left side of the equation comes from the second order Taylor expansion around the mode. On the right side of the equation, $f(x_0)$ is from $\exp(g(x_0))=\exp(\log f(x_0))$ $=f(x_0)$, which is constant. The square root part is from the rest of the integral, which is the kernel of a normal density with mean $\mu=x_0$ and variance $\sigma^2=-1/g''(x_0)$. $\left(\int e^{-\frac{(x-\mu)^2}{2\sigma^2}}dx = \sqrt{2\pi\sigma^2}\right)$

2. $f(x_0)\cdot\int e^{(x-x_0)g'(x_0)+\frac{(x-x_0)^2}{2!}g''(x_0)}dx = f(x_0)\cdot\sqrt{\frac{2\pi}{-g''(x_0)}}\cdot e^{\frac{-g'(x_0)^2}{2g''(x_0)}}$

We know the moment generating function (MGF), $M_x(t)$, of a normal random variable $x\sim N(\mu,\sigma^2)$ is

$$M_x(t)=E(e^{xt})=\int\frac{1}{\sqrt{2\pi}\sigma}e^{xt}e^{-\frac{(x-\mu)^2}{2\sigma^2}}dx=e^{\mu t+\frac{\sigma^2t^2}{2}}$$

$$\int e^{(x-x_0)g'(x_0)+\frac{(x-x_0)^2}{2!}g''(x_0)}dx = \int e^{yg'(x_0)+\frac{y^2}{2}g''(x_0)}dy$$

This is just the kernel of MGF with $\mu = 0$, $t = g'(x_0)$ and $\sigma^2 = -\frac{1}{g''(x_0)}$.

$$\int e^{yg'(x_0)+\frac{y^2}{2}g''(x_0)}dy = \int e^{yt}e^{-\frac{y^2}{2\sigma^2}}dy = \sqrt{2\pi\sigma} \cdot \int \frac{1}{\sqrt{2\pi\sigma}}e^{yt}e^{-\frac{y^2}{2\sigma^2}}dy$$
$$= \sqrt{2\pi\sigma} \cdot e^{\frac{\sigma^2 t^2}{2}}$$
$$= \sqrt{\frac{2\pi}{-g''(x_0)}} \cdot e^{\frac{-g'(x_0)^2}{2g''(x_0)}}$$

3. Derivation based on multivariate random variables

For $y_i = f(\eta_i, \theta) + \varepsilon_i$, the model can be linearized around $\eta = \hat{\eta}$, $\varepsilon = 0$ based on a first-order Taylor expansion if the FOCE method is used:

$$y_i = f_i(\hat{\eta}_i, \theta) - f'(\hat{\eta}_i, \theta)\hat{\eta}_i + f'(\hat{\eta}_i, \theta)\eta_i + \varepsilon_i$$

The marginal distribution for y_i under this linearized model is

$$y_i \sim N(f(\hat{\eta}_i, \theta) - f'(\hat{\eta}_i, \theta)\hat{\eta}_i,\, f'(\hat{\eta}_i, \theta)\Omega f'(\hat{\eta}_i, \theta)^T + \Sigma) \tag{3.1}$$

where Σ is $\sigma^2 I$.

This marginal distribution can also be derived based on the Laplacian approximation of the log-likelihood.

We remember that the approximated marginal distribution (Laplacian method) of y_i under the original model is

$$P(y_i|\psi, \Omega) = \int l_i(\eta; \psi) \cdot h(\eta; \Omega)d\eta$$
$$\approx l_i(\hat{\eta}_i; \psi) \cdot h(\hat{\eta}_i; \Omega)\sqrt{\frac{(2\pi)^p}{\left|-g''(\hat{\eta}_i)\right|}}$$
$$= \frac{1}{(2\pi)^{\frac{n_i}{2}}\,|\Sigma|^{\frac{1}{2}}}e^{-\frac{(y_i-f(\hat{\eta}_i,\theta))^T\Sigma^{-1}(y_i-f(\hat{\eta}_i,\theta))}{2}}\frac{1}{(2\pi)^{\frac{p}{2}}\,|\Omega|^{\frac{1}{2}}}e^{-\frac{\hat{\eta}_i^T\Omega^{-1}\hat{\eta}_i}{2}}\sqrt{\frac{(2\pi)^p}{\left|-g''(\hat{\eta}_i)\right|}}$$
$$= \frac{1}{(2\pi)^{\frac{n_i}{2}}\,|\Sigma|^{\frac{1}{2}}\,|\Omega|^{\frac{1}{2}}\,\left|-g''(\hat{\eta}_i)\right|^{\frac{1}{2}}}e^{-\frac{(y_i-f(\hat{\eta}_i,\theta))^T\Sigma^{-1}(y_i-f(\hat{\eta}_i,\theta))+\hat{\eta}_i^T\Omega^{-1}\hat{\eta}_i}{2}} \tag{3.2}$$

This density function can be shown to be the same as the density function for 3.1.

We know

$$\Gamma_i(\eta) = -2\frac{l_i'}{l_i} = -2\frac{\frac{1}{(2\pi)^{\frac{n_i}{2}}|\Sigma|^{\frac{1}{2}}} e^{\frac{-(y_i - f(\eta_i,\theta))^T\Sigma^{-1}(y_i - f(\eta_i,\theta))}{2}}}{\frac{1}{(2\pi)^{\frac{n_i}{2}}|\Sigma|^{\frac{1}{2}}} e^{\frac{-(y_i - f(\eta_i,\theta))^T\Sigma^{-1}(y_i - f(\eta_i,\theta))}{2}}} f'(\eta_i,\theta)^T\Sigma^{-1}\left(y_i - f(\eta_i,\theta)\right)$$
$$= -2f'(\eta_i,\theta)^T\Sigma^{-1}\left(y_i - f(\eta_i,\theta)\right)$$

Since $g'(\hat{\eta}) = -\frac{\Gamma(\hat{\eta})}{2}\Omega^{-1}\hat{\eta}$ and $g'(\hat{\eta}) = 0$
$\frac{\Gamma(\hat{\eta})}{2} + \Omega^{-1}\hat{\eta} = 0$ and $\hat{\eta} = -\Omega\frac{\Gamma(\hat{\eta})}{2} = \Omega f'(\hat{\eta}_i,\theta)^T\Sigma^{-1}\left(y_i - f(\hat{\eta}_i,\theta)\right)$

The exponent in Eq. 3.2 is

$$\begin{aligned}
&(y_i - f(\hat{\eta}_i,\theta))^T\Sigma^{-1}(y_i - f(\hat{\eta}_i,\theta)) + \hat{\eta}_i^T\Omega^{-1}\hat{\eta}_i \\
&= \left(y_i - f(\hat{\eta}_i,\theta)\right)^T\Sigma^{-1}\left(y_i - f(\hat{\eta}_i,\theta)\right) \\
&\quad + \left(y_i - f(\hat{\eta}_i,\theta)\right)^T\Sigma^{-1}f'(\hat{\eta}_i,\theta)\Omega f'(\hat{\eta}_i,\theta)^T\Sigma^{-1}\left(y_i - f(\hat{\eta}_i,\theta)\right) \\
&= \left(y_i - f(\hat{\eta}_i,\theta)\right)^T\Sigma^{-1}\left(I + f'(\hat{\eta}_i,\theta)\Omega f'(\hat{\eta}_i,\theta)^T\Sigma^{-1}\right)\left(y_i - f(\hat{\eta}_i,\theta)\right) \\
&= \left(y_i - f(\hat{\eta}_i,\theta)\right)^T\left(I + \Sigma^{-1}f'(\hat{\eta}_i,\theta)\Omega f'(\hat{\eta}_i,\theta)^T\right) \\
&\quad \times\left(\Sigma + f'(\hat{\eta}_i,\theta)\Omega f'(\hat{\eta}_i,\theta)^T\right)^{-1}\left(I + f'(\hat{\eta}_i,\theta)\Omega f'(\hat{\eta}_i,\theta)^T\Sigma^{-1}\right)\left(y_i - f(\hat{\eta}_i,\theta)\right) \\
&= \left(\left(y_i - f(\hat{\eta}_i,\theta)\right)^T + \hat{\eta}_i^T f'(\hat{\eta}_i,\theta)^T\right)\left(\Sigma + f'(\hat{\eta}_i,\theta)\Omega f'(\hat{\eta}_i,\theta)^T\right)^{-1} \\
&\quad \times\left(\left(y_i - f(\hat{\eta}_i,\theta)\right) + f'(\hat{\eta}_i,\theta)\hat{\eta}_i\right) \\
&= \left(y_i - f(\hat{\eta}_i,\theta) + f'(\hat{\eta}_i,\theta)\hat{\eta}_i\right)^T\left(\Sigma + f'(\hat{\eta}_i,\theta)\Omega f'(\hat{\eta}_i,\theta)^T\right)^{-1} \\
&\quad \times\left(y_i - f(\hat{\eta}_i,\theta) + f'(\hat{\eta}_i,\theta)\hat{\eta}_i\right)
\end{aligned} \tag{3.3}$$

We also know the approximation of Hessian matrix by the function of gradient vectors:

$$\begin{aligned}
\Delta_i(\eta) &\approx \frac{1}{2}E\left(\Gamma_i(\eta)\Gamma_i(\eta)^T\right) \\
&= 2E\left(f'(\eta_i,\theta)^T\Sigma^{-1}\left(y_i - f(\eta_i,\theta)\right)\left(y_i - f(\eta_i,\theta)\right)^T\Sigma^{-1}f'(\eta_i,\theta)\right) \\
&= 2\left(f'(\eta_i,\theta)^T\Sigma^{-1}f'(\eta_i,\theta)\right)
\end{aligned} \tag{3.4}$$

Since $g''(\eta) = -\frac{\Delta_i(\eta)}{2} - \Omega^{-1}$

$$\begin{aligned}
\left|-g''(\hat{\eta}_i)\right| &= \left|\frac{\Delta_i(\hat{\eta})}{2} + \Omega^{-1}\right| \\
&= \left|f'(\hat{\eta}_i,\theta)^T\Sigma^{-1}f'(\hat{\eta}_i,\theta) + \Omega^{-1}\right|
\end{aligned}$$

$$
\begin{aligned}
&= \left| f'(\hat{\eta}_i, \theta)^T \Sigma^{-1} f'(\hat{\eta}_i, \theta) \Omega + I_{p \times p} \right| \left| \Omega^{-1} \right| \\
&= \left| f'(\hat{\eta}_i, \theta) \Omega f'(\hat{\eta}_i, \theta)^T \Sigma^{-1} + I_{n_i \times n_i} \right| \left| \Omega^{-1} \right| \\
&= \left| f'(\hat{\eta}_i, \theta) \Omega f'(\hat{\eta}_i, \theta)^T + \Sigma \right| \left| \Sigma^{-1} \right| \left| \Omega^{-1} \right|
\end{aligned}
$$

Therefore the constant part of Eq. 3.2 is

$$
\begin{aligned}
&\frac{1}{(2\pi)^{\frac{n_i}{2}} |\Sigma|^{\frac{1}{2}} |\Omega|^{\frac{1}{2}} \left| -g''(\hat{\eta}_i) \right|^{\frac{1}{2}}} \\
&= \frac{1}{(2\pi)^{\frac{n_i}{2}} |\Sigma|^{\frac{1}{2}} |\Omega|^{\frac{1}{2}} \left| f'(\hat{\eta}_i, \theta) \Omega f'(\hat{\eta}_i, \theta)^T + \Sigma \right|^{\frac{1}{2}} \left| \Sigma^{-1} \right|^{\frac{1}{2}} \left| \Omega^{-1} \right|^{\frac{1}{2}}} \\
&= \frac{1}{(2\pi)^{\frac{n_i}{2}} \left| f'(\hat{\eta}_i, \theta) \Omega f'(\hat{\eta}_i, \theta)^T + \Sigma \right|^{\frac{1}{2}}}
\end{aligned}
$$

In combination with Eq. 3.3, Eq. 3.2 can be expressed as
$\frac{1}{(2\pi)^{\frac{n_i}{2}} \left| f'(\hat{\eta}_i,\theta)\Omega f'(\hat{\eta}_i,\theta)^T + \Sigma \right|^{\frac{1}{2}}} e^{-\frac{\left(y_i - f(\hat{\eta}_i,\theta) + f'(\hat{\eta}_i,\theta)\hat{\eta}_i\right)^T \left(\Sigma + f'(\hat{\eta}_i,\theta)\Omega f'(\hat{\eta}_i,\theta)^T\right)^{-1} \left(y_i - f(\hat{\eta}_i,\theta) + f'(\hat{\eta}_i,\theta)\hat{\eta}_i\right)}{2}}$
which is exactly the same distribution as that for the marginal distribution of y_i for the linearized model (3.1).

Then the objective function without the constant is $\log\left| f'(\hat{\eta}_i, \theta)\Omega f'(\hat{\eta}_i, \theta)^T + \Sigma \right| + \left(y_i - f(\hat{\eta}_i, \theta) + f'(\hat{\eta}_i, \theta)\hat{\eta}_i\right)^T \left(f'(\hat{\eta}_i, \theta)\Omega f'(\hat{\eta}_i, \theta)^T + \Sigma\right)^{-1} \left(y_i - f(\hat{\eta}_i, \theta) + f'(\hat{\eta}_i, \theta)\hat{\eta}_i\right)$

4. Equivalence of NONMEM FOCE and modified Laplacian approximation in [8]

Equation 16 can be further expanded to

$$
\begin{aligned}
&\hat{\Phi}_i + \log|\Omega| + \hat{\eta}_i' \Omega^{-1} \hat{\eta}_i + \log\left| \Omega^{-1} + \frac{\hat{E}\left(\Gamma_i(\eta)\Gamma_i(\eta)^T\right)}{4} \right| \\
&\quad = -2\log\left(\frac{1}{(2\pi\sigma^2)^{\frac{n_i}{2}}} e^{-\frac{(y_i - f(\hat{\eta}_i,\theta))^2}{2\sigma^2}} \right) \\
&\qquad + \log|\Omega| + \hat{\eta}_i' \Omega^{-1} \hat{\eta}_i + \log\left| \Omega^{-1} + \frac{f'(\hat{\eta}_i, \theta)^T f'(\hat{\eta}_i, \theta)}{\sigma^2} \right|
\end{aligned}
$$

(using 3.4)

$$
\begin{aligned}
&= -2\left[-\frac{n_i}{2}\log(2\pi\sigma^2) - \frac{(y_i - f(\hat{\eta}_i, \theta))^2}{2\sigma^2} - \frac{1}{2}\log|\Omega| \right. \\
&\qquad \left. + \frac{\hat{\eta}_i' \Omega^{-1} \hat{\eta}_i}{2} - \frac{1}{2}\log\left| \Omega^{-1} + \frac{f'(\hat{\eta}_i, \theta)^T f'(\hat{\eta}_i, \theta)}{\sigma^2} \right| \right]
\end{aligned}
$$

$$
\begin{aligned}
&= -2\left[-\frac{n_i}{2}\log(2\pi\sigma^2) - \frac{(y_i - f(\hat{\eta}_i,\theta))^2}{2\sigma^2}\right.\\
&\quad \left. + \frac{\hat{\eta}_i'\Omega^{-1}\hat{\eta}_i}{2} - \frac{1}{2}\log\frac{|\Omega|}{\sigma^2}\left|\Omega^{-1}\sigma^2 + f'(\hat{\eta}_i,\theta)^T f'(\hat{\eta}_i,\theta)\right|\right]\\
&= -2\left[-\frac{n_i}{2}\log(2\pi\sigma^2) - \frac{(y_i - f(\hat{\eta}_i,\theta))^2 + \hat{\eta}_i'\Lambda^T\Lambda\hat{\eta}_i}{2\sigma^2}\right.\\
&\quad \left. + \log|\Lambda| - \frac{1}{2}\log\left|\Lambda^T\Lambda + f'(\hat{\eta}_i,\theta)^T f'(\hat{\eta}_i,\theta)\right|\right]\\
&= -2\left[\ell_{LAi}\right]
\end{aligned}
$$

where ℓ_{LAi} is the log-likelihood of the modified Laplacian approximation for the ith individual (Eq. 7.19 in Ref. [8]) and Λ is defined as a relative precision factor so that $\frac{\Omega^{-1}}{1/\sigma^2} = \Lambda^T\Lambda$ in Ref. [8].

5. NONMEM and Splus codes for numerical example

NONMEM:

```
$PROBLEM OBJECTIVE FUNCTION CALCULATION
$INPUT ID TIME DV DOSE
$DATA SIMSET.csv IGNORE=C
$PRED
KE=THETA(1)*EXP(ETA(1))
IPRE=10*EXP(-KE*TIME)
Y=IPRE+ERR(1)
;Y=IPRE*(1+ERR(1));FOR PROPORTIONAL RESIDUAL ERROR
;Y=IPRE*EXP(ERR(1));FOR EXPONENTIAL RESIDUAL ERROR
$THETA
(0.5 FIX);~KE
$OMEGA
0.04 FIX;~KEETA
$SIGMA
0.1 FIX
$ESTIMATION MAXEVAL=0 NSIG=3 METHOD=0 NOABORT PRINT=10
$TABLE ID TIME ETA1 IPRE NOPRINT ONEHEADER FILE=OUT.FIT
```

Splus:

```
## import NONMEM output file which contains ID TIME ETA1 IPRE
data<-read.table(…)
ke<-0.5
omega<-0.04
eps<-0.1
#define function f()
f<-
function(x, y)
{
10*exp(-ke*exp(x)*y)
```

```
}
#define the differential function of f()
fp<-
function(x, y)
{
10*exp(-ke*exp(x)*y)*(-ke*y)*exp(x)
}
data$fp<-fp(data$ETA1, data$TIME)
data$f<-f(0, data$TIME)
#data$f<-f(data$ETA1, data$TIME)#for FOCE with interaction
sum<-0
for (i in 1:10) {
data1<-data[data$ID==i,]
#residual var-cov matrix for additive error
cov<-data1$fp%*%t(data1$fp)*omega+diag(2)*eps
#for proportional error
#cov<-data1$fp%*%t(data1$fp)*omega+diag(data1$f**2)*eps
ginv<-solve(cov)#inverse matrix of residual var-cov matrix
sec<-t(data1$DV-data1$IPRE+data1$fp*data1$ETA1)%*%ginv%*%
   (data1$DV-data1$IPRE+data1$fp*data1$ETA1)
frs<-determinant(cov, logarithm=T)$modulus[[1]]
sum1<-sec+frs
sum<-sum+sum1
}
sum
```

References

1. Sheiner LB, Rosenberg B, Melmon KL (1972) Modelling of individual pharmacokinetics for computer aided drug dosage. Comput Biomed Res 5:441–459
2. Beal SL, Sheiner LB (1980) The NONMEM system. Am Statistician 34:118
3. Beal SL, Sheiner LB (1998) NONMEM users guide
4. Mentré F (2005) History and new developments in estimation methods for nonlinear mixed-effects model, Population Approach Group in Europe (PAGE) 2005 meeting.
5. Wolfinger RD (1993) Laplace's approximation for nonlinear mixed models. Biometrika 80:791–795
6. Davidian M, Giltinan DM (1995) Nonlinear models for repeated measurement data. Chapman & Hall, First Edition
7. Vonesh EF, Chinchilli VM (1997) Linear and nonlinear models for the analysis of repeated measures. Marcel Dekker, New York
8. Pinheiro JC, Bates DM (2000) Mixed effects models in S and S-plus. Springer Verlag