MM ALGORITHMS FOR VARIANCE COMPONENT ESTIMATION AND SELECTION IN LOGISTIC LINEAR MIXED MODEL

Liuyi Hu[1], Wenbin Lu[1], Jin Zhou[2] and Hua Zhou[3]

[1]*North Carolina State University,* [2]*University of Arizona, and* [3]*University of California, Los Angeles*

Supplementary Material

S1 Derivation of approximated log-likelihood in (3.2)

$$L_{\mathrm{LA}}(\beta,\sigma) = h(u^* \mid \beta,\sigma^2) - \frac{1}{2}\ln\det\nabla^2\left\{-h(u^*\mid\beta,\sigma^2)\right\},$$

where

$$h(u^*\mid\beta,\sigma^2) = \sum_j\left\{y_j\eta_j^* - \ln\left(1+e^{\eta_j^*}\right)\right\} - \frac{1}{2}\sum_{i=1}^m q_i\ln\sigma_i^2 - \frac{n}{2}\ln 2\pi - \frac{1}{2}\sum_{i=1}^m\frac{\|u_i^*\|_2^2}{\sigma_i^2}.$$

The gradient and Hessian of $h(u\mid\beta,\sigma)$ at $u=u^*$ are

$$\begin{aligned}
\nabla_u h(u\mid\beta,\sigma^2)_{|u=u^*} &= Z^T(y-p^*) - \begin{pmatrix}\sigma_1^{-2}u_1^*\\ \vdots \\ \sigma_m^{-2}u_m^*\end{pmatrix},\\
\nabla_u^2 h(u\mid\beta,\sigma^2)_{|u=u^*} &= -\left\{Z^TW^*Z + \mathrm{blkdiag}(\sigma_1^{-2}I_{q_1},\ldots,\sigma_m^{-2}I_{q_m})\right\},
\end{aligned}$$

where $p^* = (p_1^*,\ldots,p_n^*)^T$ with $p_j^* = e^{\eta_j^*}/(1+e^{\eta_j^*})$ and $W^* = \mathrm{diag}(w^*)$ is a diagonal matrix with entries

$$w_j^* = p_j^*(1 - p_j^*) = \frac{e^{\eta_j^*}}{\left(1 + e^{\eta_j^*}\right)^2}.$$

Therefore,

$$\begin{aligned} L_{\mathrm{LA}}(\beta, \sigma) &= \sum_j \left\{ y_j \eta_j^* - \ln\left(1 + e^{\eta_j^*}\right) \right\} - \frac{1}{2} \sum_{i=1}^m q_i \ln \sigma_i^2 - \frac{n}{2} \ln 2\pi - \frac{1}{2} \sum_{i=1}^m \frac{\|u_i^*\|_2^2}{\sigma_i^2} \\ &\quad - \frac{1}{2} \ln\det\left\{ Z^T W^* Z + \mathrm{blkdiag}(\sigma_1^{-2} I_{q_1}, \ldots, \sigma_m^{-2} I_{q_m}) \right\}. \end{aligned} \tag{S1.1}$$

Using the matrix determinant lemma, we have

$$\begin{aligned} & \ln\det\left\{ Z^T W^* Z + \mathrm{blkdiag}(\sigma_1^{-2} I_{q_1}, \ldots, \sigma_m^{-2} I_{q_m}) \right\} \\ = \; & \ln\det\left(W^{*-1} + \sum_i \sigma_i^2 Z_i Z_i^T \right) + \ln\det\left(\mathrm{blkdiag}(\sigma_1^{-2} I_{q_1}, \ldots, \sigma_m^{-2} I_{q_m}) \right) + \ln\det W^* \\ = \; & \ln\det\left(W^{*-1} + \sum_i \sigma_i^2 Z_i Z_i^T \right) - \sum_{i=1}^m q_i \ln \sigma_i^2 + \ln\det W^*. \end{aligned} \tag{S1.2}$$

Substitute (S1.2) to (S1.1) gives

$$\begin{aligned} L_{\mathrm{LA}}(\beta, \sigma) &= \sum_j \left\{ y_j \eta_j^* - \ln\left(1 + e^{\eta_j^*}\right) \right\} - \frac{1}{2} \sum_{i=1}^m \frac{\|u_i^*\|_2^2}{\sigma_i^2} \\ &\quad - \frac{1}{2} \ln\det\left(W^{*-1} + \sum_i \sigma_i^2 Z_i Z_i^T \right) - \frac{1}{2} \ln\det W^* + \text{constant term}, \end{aligned}$$

where the constant term equals $-\frac{n}{2} \ln 2\pi$.

S2 Derivation of approximated log-likelihood in (3.8)

$$L_{\mathrm{LA}}(\beta, \sigma) = h(u^* \mid \beta, \sigma) - \frac{1}{2} \ln\det \nabla^2 \{-h(u^* \mid \beta, \sigma^2)\},$$

where

$$h(u \mid \beta, \sigma^2) = \sum_j \{y_j \eta_j - \ln(1 + e^{\eta_j})\} - \frac{1}{2}\|u\|_2^2 - \frac{n}{2}\ln 2\pi.$$

The gradient and Hessian $h(u \mid \beta, \sigma)$ at $u = u^*$ are

$$\begin{aligned}
\nabla_u h(u \mid \beta, \sigma)_{|u=u^*} &= D^T Z^T (y - p^*) - u^*, \\
\nabla_u^2 h(u \mid \beta, \sigma)_{|u=u^*} &= -\left(D^T Z^T W^* Z D + I_q\right),
\end{aligned}$$

where $p^* = (p_1^*, \ldots, p_n^*)^T$ with $p_j^* = e^{\eta_j^*}/(1 + e^{\eta_j^*})$ and $W^* = \mathrm{diag}(w^*)$ is a diagonal matrix with entries

$$w_j^* = p_j^*(1 - p_j^*) = \frac{e^{\eta_j^*}}{\left(1 + e^{\eta_j^*}\right)^2}.$$

Using the matrix determinant lemma, we have

$$\ln\det\left(D^T Z^T W^* Z D + I_q\right) = \ln\det\left(W^{*-1} + \sum_i \sigma_i^2 Z_i Z_i^T\right) + \ln\det W^*.$$

Therefore,

$$\begin{aligned}
& L_{\mathrm{LA}}(\beta, \sigma) \\
= & \sum_j \left\{y_j \eta_j^* - \ln\left(1 + e^{\eta_j^*}\right)\right\} - \frac{1}{2}\|u^*\|_2^2 - \frac{n}{2}\ln 2\pi - \frac{1}{2}\ln\det\left(D^T Z^T W^* Z D + I_q\right) \\
= & \sum_j \left\{y_j \eta_j^* - \ln\left(1 + e^{\eta_j^*}\right)\right\} - \frac{1}{2}\|u^*\|_2^2 - \frac{1}{2}\ln\det\left(W^{*-1} + \sum_i \sigma_i^2 Z_i Z_i^T\right) \\
& - \frac{1}{2}\ln\det W^* + \text{constant term},
\end{aligned}$$

where the constant term equals $-\frac{n}{2}\ln 2\pi$.

S3 Proof of ascent property in (3.6)

Proof. From (3.2) the approximated log-likelihood is

$$\begin{aligned} L_{\text{LA}}(\beta, \sigma) &= \sum_j \left\{ y_j \eta_j^* - \ln\left(1 + e^{\eta_j^*}\right) \right\} - \frac{1}{2} \sum_{i=1}^m \frac{\|u_i^*\|_2^2}{\sigma_i^2} \\ &\quad - \frac{1}{2} \ln \det \left(W^{*-1} + \sum_i \sigma_i^2 Z_i Z_i^T \right) - \frac{1}{2} \ln \det W^* + \text{ terms without } \beta, \sigma^2, \end{aligned}$$

where u^* is the maximizer of $h(u \mid \beta, \sigma)$, $\eta^* = X\beta + Zu^*$ and $W^* = \text{diag}(w^*)$ is a diagonal matrix with entries

$$w_j^* = p_j^*(1 - p_j^*) = \frac{e^{\eta_j^*}}{\left(1 + e^{\eta_j^*}\right)^2}.$$

Thus

$$L_{\text{LA}}(\sigma \mid \beta, u^*) = -\frac{1}{2} \sum_{i=1}^m \frac{\|u_i^*\|_2^2}{\sigma_i^2} - \frac{1}{2} \ln \det \left(W^{*-1} + \sum_i \sigma_i^2 Z_i Z_i^T \right) + c,$$

where $c = \sum_j \left\{ y_j \eta_j^* - \ln\left(1 + e^{\eta_j^*}\right) \right\} - \frac{1}{2} \ln \det W^* - \frac{n}{2} \ln 2\pi$ is a constant not involving σ.

The minorization (2.4) leads to the surrogate function of $L_{\text{LA}}(\sigma \mid \beta, u^*)$

$$g(\sigma^2 \mid \sigma^{2(t)}) = -\frac{1}{2} \sum_{i=1}^m \frac{\|u_i^*\|_2^2}{\sigma_i^2} - \frac{1}{2} \sum_{i=1}^m \sigma_i^2 \text{tr} \left\{ \left(\sum_i \sigma_i^{2(t)} Z_i Z_i^T + W^{*-1} \right)^{-1} Z_i Z_i^T \right\} + c^{(t)},$$

where $c^{(t)}$ is a constant irrelevant to optimization. Since $\sigma^{2(t+1)} = (\sigma_1^{2(t+1)}, \ldots, \sigma_m^{2(t+1)})$ with

$$\sigma_i^{2(t+1)} = \left[\frac{\|u_i^*\|_2^2}{\text{tr}\left\{ Z_i^T (\sum_i \sigma_i^{2(t)} Z_i Z_i^T + W^{*-1})^{-1} Z_i \right\}} \right]^{\frac{1}{2}}$$

maximizes the surrogate function $g(\sigma^2 \mid \sigma^{2(t)})$, we have the following inequality satisfied

$$L_{LA}(\sigma^{(t+1)} \mid \beta, u^*) \geq g(\sigma^{2(t+1)} \mid \sigma^{2(t)}) \geq g(\sigma^{2(t)} \mid \sigma^{2(t)}) = L_{LA}(\sigma^{(t)} \mid \beta, u^*).$$

Therefore, the iterates possess the ascent property. $\square$