Appendix for "Flexible semiparametric analysis of longitudinal genetic studies by reduced rank smoothing"

Yuanjia Wang*
Columbia University, New York, USA
Chiahui Huang
Columbia University, New York, USA
Yixin Fang
Georgia State University, Atlanta, USA
Qiong Yang
Boston University, Boston, USA
Runze Li
The Pennsylvania State University at University Park, University Park, USA

A.1 Connection of penalized splines and mixed effects model

To explore the well-known connection between penalized splines and mixed effects model (Ruppert et al. 2003, Wand 2003), we write the model (1) in a matrix form. Let n denote the number of family, let n_i denote the number of subjects in the ith family, and let T_{ij} denote the number of measurements on subject (i, j). Let Y_i denote the vector of trait measurements from all subjects in family i, that is, $(y_{ijh})_{j=1,\cdots,n_i;h=1,\cdots,T_{ij}}$. Let $\mathbf{t}_i$ denote the corresponding assessment time points, and let ε_i denote the corresponding residual effects. Let $X_i = (x_{i11}, \cdots, x_{in_iT_{in_i}})^T$, and let $\mathbf{1}_n$ denote a vector of n ones. Define the design matrices

$$\begin{aligned} W_i &= (\mathbf{1}, \mathbf{t}_i, \cdots, \mathbf{t}_i^q, (\mathbf{t}_i - \tau_1)_+^q, \cdots, (\mathbf{t}_i - \tau_M)_+^q), \\ Z_i &= (\mathbf{1}, \mathrm{diag}(\mathbf{1}_{T_{i1}}, \cdots, \mathbf{1}_{T_{in_i}})), \end{aligned}$$

* *Address for Correspondence:* Yuanjia Wang, Department of Biostatistics, Mailman School of Public Health, Columbia University, New York, NY 10032, USA, Email: yuanjia.wang@columbia.edu

and set $u_i = (\alpha_i, \gamma_{i1}, \cdots, \gamma_{in_i})^T$, then the model (1) can be expressed as

$$Y_i = X_i\beta + W_i\eta + Z_i u_i + \varepsilon_i, \tag{A.1}$$

where β is a vector of fixed effects and u_i are random effects including family-specific shared environmental effect and subject-specific polygenic effect, and ε_i are the residuals. The covariance structure of the random effects can be expressed as

$$D_i = \mathrm{Cov}(u_i, u_i^T) = \mathrm{diag}(\sigma_\alpha^2, 2K^i\sigma_\gamma^2),$$

where K^i is the known kinship matrix with the (h, l) element being the kinship coefficient between subject (i, h) and (i, l) (i.e., K^i_{hl}). We write the covariance matrix of ε_i in a variance-correlation form, that is,

$$E_i = V_i R_i(\theta) V_i, \tag{A.2}$$

where $V_i = \mathrm{diag}(\sigma_{i11}, \cdots, \sigma_{in_i T_{in_i}})$, and $R_i(\theta)$ is the correlation matrix with elements $(\rho(t_{ijh}, t_{ikl}; \theta))$. Then the covariance matrix for Y_i is

$$\Sigma_i = Z_i D_i Z_i^T + E_i.$$

Let $\Sigma = \mathrm{diag}(\Sigma_1, \cdots, \Sigma_n)$, let $X = \mathrm{diag}(X_1, \cdots, X_n)$, let $W = \mathrm{diag}(W_1, \cdots, W_n)$, and let $Y = (Y_1^T, \cdots, Y_n^T)^T$. Then the penalized likelihood (6) in a matrix form is

$$-\frac{1}{2}\log|\Sigma| - \frac{1}{2}(Y - X\beta - W\eta)^T\Sigma^{-1}(Y - X\beta - W\eta) - \frac{1}{2}\lambda\eta^T J\eta. \tag{A.3}$$

Here $J = \mathrm{diag}(\mathbf{0}_q, \mathbf{1}_M)$ is a penalty matrix implying the spline coefficients $\eta_{q+1}, \cdots, \eta_{q+M}$ are penalized. Given variance components Σ, the fixed effects β and η can be solved by maximizing the penalized likelihood (A.3). The solution $\hat{\beta}, \hat{\eta}$ solves

$$\begin{pmatrix} X^T\Sigma^{-1}X & X^T\Sigma^{-1}W \\ W^T\Sigma^{-1}X & W^T\Sigma^{-1}W + \lambda J \end{pmatrix} \begin{pmatrix} \beta \\ \eta \end{pmatrix} = \begin{pmatrix} X^T\Sigma^{-1}Y \\ W^T\Sigma^{-1}Y \end{pmatrix}. \tag{A.4}$$

Wand (2003) shows that there is a connection of the solution to (A.4) with a linear mixed model which we now describe. Define the design matrices

$$B_i = (1, \mathbf{t}_i, \cdots, \mathbf{t}_i^q), \quad U_i = ((\mathbf{t}_i - \tau_1)_+^q, \cdots, (\mathbf{t}_i - \tau_M)_+^q),$$

and set $\eta^P = (\eta_0, \cdots, \eta_q)^T$, $\eta^Q = (\eta_{q+1}, \cdots, \eta_{q+M})^T$. Then the mixed effects model yielding equivalent solution to (6) is

$$\begin{aligned} Y_i &= X_i\beta + B_i\eta^P + Z_iu_i + U_i\eta^Q + \varepsilon_i, \\ u_i &\sim N(0, D_i), \quad \eta^Q \sim N(0, \sigma_\eta^2 I_M), \quad \varepsilon_i \sim N(0, \sigma_\varepsilon^2), \end{aligned} \tag{A.5}$$

In other words, the spline coefficients $\eta^Q = (\eta_{q+1}, \cdots, \eta_{q+M})^T$ are modeled as independent random effects with the same variance and therefore are shrunk towards zero. The smoothing parameter can be estimated by the ratio of the two variance components, that is,

$$\lambda = \sigma_\varepsilon^2/\sigma_\eta^2. \tag{A.6}$$

When using smoothing parameter (A.6), the solution to (A.4) is identical to the best linear unbiased predictor (BLUP) from the linear mixed model which we described above (Wand 2003). That is, let $\hat{\eta}_1, \cdots, \hat{\eta}_q$ denote the estimates of the fixed coefficients, and let $\hat{\eta}_{q+1}, \cdots, \hat{\eta}_{q+M}$ denote the BLUP estimates of the random coefficients. Then the fitted value of the mean function η at time t_{ijk} is

$$\hat{\eta}(t_{ijk}) = \hat{\eta}_0 + \hat{\eta}_1 t_{ijk} + \cdots + \hat{\eta}_q t_{ijk}^q + \hat{\eta}_{q+1}(t_{ijk} - \tau_1)_+^q \cdots + \hat{\eta}_{q+M}(t_{ijk} - \tau_M)_+^q. \tag{A.7}$$

A.2 Estimating time-varying genetic effect

In this section, we expand the mixed effects model in Appendix A.1 to handle time-varying QTL genetic effect in model (4). Let $\mathbf{g}_i = (g_{ij}\mathbf{1}_{n_{ij}}^T)_{j=1,\cdots,n_i}^T$, where g_{ij} is the genotype of subject j from family i. Let $\mathbf{w}_i = (w_{ijh})_{j=1,\cdots,n_i, h=1,\cdots,T_{ij}}$ denote the time-varying covariate with varying-coefficient. Define the design matrices

$$C_i^1 = (\mathbf{g}_i, \mathbf{t}_i \circ \mathbf{g}_i, \cdots, \mathbf{t}_i^q \circ \mathbf{g}_i), C_i^2 = (\mathbf{w}_i, \mathbf{t}_i \circ \mathbf{w}_i, \cdots, \mathbf{t}_i^q \circ \mathbf{w}_i)$$

and

$$F_i^1 = ((\mathbf{t}_i - \tau_1)_+^q \circ \mathbf{g}_i, \cdots, (\mathbf{t}_i - \tau_M)_+^q \circ \mathbf{g}_i), F_i^2 = ((\mathbf{t}_i - \tau_1)_+^q \circ \mathbf{w}_i, \cdots, (\mathbf{t}_i - \tau_M)_+^q \circ \mathbf{w}_i),$$

where "$\circ$ " denote Hadamard (element by element) product. For given variance components, the penalized log likelihood (based on the marginal likelihood) of β, η and ξ is

$$-\frac{1}{2}\log|\Sigma| - \frac{1}{2}r'\Sigma^{-1}r - \frac{1}{2}\lambda_1\eta' J\eta - \frac{1}{2}\lambda_2\xi' J\xi - \frac{1}{2}\lambda_3\theta' J\theta,$$

where $r = Y - X\beta - W\eta - S_1\xi - S_2\theta$, $S_1 = (C^1, F^1)$ and $S_2 = (C^2, F^2)$. A mixed effects model similar to (A.5) can be written as

$$\begin{aligned} Y_i &= X_i\beta + B_i\eta^P + C_i^1\xi^P + C_i^1\theta^P + Z_iu_i + U_i\eta^Q + F_i^1\xi^Q + F_i^2\theta^Q + \varepsilon_i, \\ u_i &\sim N(0, D_i), \quad \eta^Q \sim N(0, \sigma_\eta^2 I_M), \quad \xi^Q \sim N(0, \sigma_\xi^2 I_M), \quad \theta^Q \sim N(0, \sigma_\theta^2 I_M) \quad \varepsilon_i \sim N(0, \sigma_\varepsilon^2), \end{aligned} \tag{A.8}$$

where σ_η^2 controls smoothness of the baseline function, σ_ξ^2 controls smoothness of the QTL genetic effect function and σ_θ^2 controls smoothness of other varying coefficients. Similar to the smoothing parameter for the baseline function, the smoothing parameter for the genetic effect function can be estimated as $\lambda_1 = \sigma_\varepsilon^2/\sigma_\eta^2$, $\lambda_2 = \sigma_\varepsilon^2/\sigma_\xi^2$, $\lambda_3 = \sigma_\varepsilon^2/\sigma_\theta^2$.

Let $\hat{\xi}_0, \cdots, \hat{\xi}_q, \hat{\theta}_0, \cdots, \hat{\theta}_q$ denote the estimates of the fixed coefficients, and let $\hat{\xi}_{q+1}, \cdots, \hat{\xi}_{q+M}$, $\hat{\theta}_{q+1}, \cdots, \hat{\theta}_{q+M}$ denote the BLUP estimates of the random coefficients. Then the fitted value of the time-varying genetic function $\beta_g(t)$ at time t_{ijh} is

$$\hat{\beta}_g(t_{ijh}) = \hat{\xi}_0 + \hat{\xi}_1 t_{ijh} + \cdots + \hat{\xi}_q t_{ijh}^q + \hat{\xi}_{q+1}(t_{ijh} - \tau_1)_+^q \cdots + \hat{\xi}_{q+M}(t_{ijh} - \tau_M)_+^q, \tag{A.9}$$

and the estimated other varying-coefficient

$$\hat{\theta}(t_{ijh}) = \hat{\theta}_0 + \hat{\theta}_1 t_{ijh} + \cdots + \hat{\theta}_q t_{ijh}^q + \hat{\theta}_{q+1}(t_{ijh} - \tau_1)_+^q \cdots + \hat{\theta}_{q+M}(t_{ijh} - \tau_M)_+^q. \tag{A.10}$$