1 Truncated power series

1.1 Notation

A spline model with subject-specific curves can be modelled as

$$
\begin{aligned}
y_{ijk} = f_i(x_{ijk}) + g_{ij}(x_{ijk}) + \epsilon_{ijk}, \quad & i = 1, \dots, G \text{ (indices of populations)}, \\
& j = 1, \dots, J_i \text{ (indices of subjects in } i^{th} \text{ population)}, \\
& k = 1, \dots, N_{ij} \text{ (indices of datapoints in } j^{th} \text{ subjects, } i^{th} \text{ population)},
\end{aligned}
$$

where $f_i(x)$ and $g_{ij}(x)$ are estimated using d degree penalised spline models with truncated power series bases, and ϵ_{ijk} are i.i.d. Gaussian error $\epsilon_{ijk} \sim N(0, \sigma_\epsilon^2)$. Essentially, each individual curve is modelled as a deviation from its respective population estimate. The mathematical models are

$$
f_i(x) = \sum_{m=0}^{d} \beta_{mi} x^m + \sum_{l=1}^{K} u_{li}(x - \kappa_l)_+^d, \quad g_{ij}(x) = \sum_{m=0}^{d} b_{mij} x^m + \sum_{l=1}^{K} v_{lij}(x - \kappa_l)_+^d,
$$

where the population coefficients for the i^{th} population are modelled as

$$
\begin{aligned}
\boldsymbol{\beta}_i &= (\beta_{0i}, \dots, \beta_{2i})^T \text{ are fixed effects}, \quad \forall i \\
\boldsymbol{u}_i | \boldsymbol{\beta}_i &= (u_{1i}, \dots, u_{Ki})^T \sim TN(\mathbf{0}, \sigma_u^2 \boldsymbol{I}), \quad \forall i
\end{aligned}
$$

and the subject deviations for the j^{th} subject in i^{th} population are modelled as

$$
\begin{aligned}
\boldsymbol{b}_{ij} | \boldsymbol{\beta}_i, \boldsymbol{u}_i &= (b_{0ij}, \dots, b_{dij})^T \sim TN(\mathbf{0}, \boldsymbol{\Sigma}_b), \quad \forall j \\
\boldsymbol{v}_{ij} | \boldsymbol{\beta}_i, \boldsymbol{u}_i &= (v_{1ij}, \dots, v_{Kij})^T \sim TN(\mathbf{0}, \sigma_v^2 \boldsymbol{I}), \quad \forall j.
\end{aligned}
$$

We will refer $\kappa_1, \dots, \kappa_K$ as *inner knots* for the rest of this paper. Our goal is to enforce shape constraints on both $f_i(x)$ and $f_i(x) + g_{ij}(x)$ in the interval $[\kappa_0, \kappa_{K+1}]$ which are usually the endpoints of the predictor range.

The covariance matrix $\boldsymbol{\Sigma}_b$ in the conditional $\boldsymbol{b}_{ij}$ is modelled as a full positive-definite matrix rather than a multiple of identity, to take into account of the potential correlation among the polynomial terms; whereas, the spline terms ($\boldsymbol{u}_i | \boldsymbol{\beta}_i$ and $\boldsymbol{v}_{ij} | \boldsymbol{\beta}_i, \boldsymbol{u}_i$) are treated as mutually independent. The truncation region of the normal distributions are dependent to the regression coefficients, and will be specified in details when discussing the prior distributions.

For brevity sake, we denote the population coefficients as

$$
\boldsymbol{\theta}_i^T = (\boldsymbol{\beta}_i^T, \boldsymbol{u}_i^T), \quad \boldsymbol{\theta}^T = (\boldsymbol{\theta}_1^T, \dots, \boldsymbol{\theta}_G^T)
$$

and subject-specific deviations as

$$
\boldsymbol{\gamma}_{ij}^T = (\boldsymbol{b}_{ij}^T, \boldsymbol{v}_{ij}^T), \quad \boldsymbol{\gamma}_i^T = (\boldsymbol{\gamma}_{i1}^T, \dots, \boldsymbol{\gamma}_{iJ_i}^T), \quad \boldsymbol{\gamma}^T = (\boldsymbol{\gamma}_1^T, \dots, \boldsymbol{\gamma}_G^T)
$$

for the rest of this paper. We also define the design matrix for the j^{th} subject in the i^{th} population as

$$
\boldsymbol{X}_{ij} = \begin{bmatrix}
1 & x_{ij1} & \cdots & x_{ij1}^d & (x_{ij1} - \kappa_1)_+^d & \cdots & (x_{ij1} - \kappa_K)_+^d \\
1 & x_{ij2} & \cdots & x_{ij2}^d & (x_{ij2} - \kappa_1)_+^d & \cdots & (x_{ij2} - \kappa_K)_+^d \\
\vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\
1 & x_{ijN_{ij}} & \cdots & x_{ijN_{ij}}^d & (x_{ijN_{ij}} - \kappa_1)_+^d & \cdots & (x_{ijN_{ij}} - \kappa_K)_+^d
\end{bmatrix},
$$

and the following matrices

$$
\begin{aligned}
\boldsymbol{P}_i &= \begin{bmatrix} \boldsymbol{X}_{i1} \\ \vdots \\ \boldsymbol{X}_{iJ_i} \end{bmatrix} \in \mathbb{R}^{N_i \times (1+d+K)}, & \boldsymbol{P} &= \text{blockdiag}(\boldsymbol{P}_1 \dots \boldsymbol{P}_G) \in \mathbb{R}^{N \times G(1+d+K)}, \\
\boldsymbol{S}_i &= \text{blockdiag}(\boldsymbol{X}_{i1}, \dots, \boldsymbol{X}_{iJ_i}) \in \mathbb{R}^{N_i \times J_i(1+d+K)}, & \boldsymbol{S} &= \text{blockdiag}(\boldsymbol{S}_1, \dots, \boldsymbol{S}_G) \in \mathbb{R}^{N \times J(1+d+K)},
\end{aligned}
$$

where $N_i = \sum_{j=1}^{J_i} N_{ij}$ is the number of samples in i^{th} population, $N = \sum_{i=1}^{G} N_i$ is the total number of samples, and $J = \sum_{i=1}^{G} J_i$ is the total number of subjects, such that the model can be written more concisely as

$$\begin{aligned} \boldsymbol{y}_{ij} &= (y_{ij1}, \ldots, y_{ijN_{ij}})^T = X_{ij}(\boldsymbol{\theta}_i + \gamma_{ij}) + \boldsymbol{\epsilon}_{ij} \,, \\ \boldsymbol{y}_i &= (\boldsymbol{y}_{i1}^T, \ldots, \boldsymbol{y}_{iJ_i}^T)^T = \boldsymbol{P}_i\boldsymbol{\theta}_i + \boldsymbol{S}_i\gamma_i + \boldsymbol{\epsilon}_i \,, \\ \boldsymbol{y} &= (\boldsymbol{y}_1^T, \ldots, \boldsymbol{y}_G^T)^T = \boldsymbol{P\theta} + \boldsymbol{S\gamma} + \boldsymbol{\epsilon} \,. \end{aligned}$$

Lastly, the variance and covariance terms involved in this model (σ_ϵ^2, σ_u^2, σ_v^2, and $\boldsymbol{\Sigma}_b$) are denoted as $\boldsymbol{\tau}$.

1.2 Constraints

Monotonicity for linear and quadratic spline models can be achieved by ensuring both $f_i'(x) \geq 0$ and $f_i'(x) + g_{ij}'(x) \geq 0$ for all $x \in [\kappa_0, \kappa_{K+1}]$, which can be written as a set of linear inequality constraints $\boldsymbol{A\theta}_i \geq \boldsymbol{0}$ and $\boldsymbol{A}(\boldsymbol{\theta}_i + \boldsymbol{\gamma}_{ij}) \geq \boldsymbol{0}$ respectively. For monotonic increasing trend, the constraints matrix $\boldsymbol{A}$ is

$$\boldsymbol{A} = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 1 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 1 & 1 & \cdots & 1 \end{bmatrix}, \quad dim(\boldsymbol{A}) = K+1 \times K+2$$

for linear spline models, and

$$\boldsymbol{A} = \begin{bmatrix} 0 & 1 & 2\kappa_0 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 2\kappa_1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 2\kappa_2 & 2(\kappa_2 - \kappa_1) & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 1 & 2\kappa_{K+1} & 2(\kappa_{K+1} - \kappa_1) & 2(\kappa_{K+1} - \kappa_2) & \cdots & 2(\kappa_{K+1} - \kappa_K) \end{bmatrix}, \quad dim(\boldsymbol{A}) = K+2 \times K+3$$

for quadratic spline models. Monotonic decreasing models can be obtained by negating the corresponding $\boldsymbol{A}$.

1.3 Likelihood

The likelihood is given by

$$p(\boldsymbol{y}|\gamma, \boldsymbol{\theta}, \boldsymbol{\tau}) \propto \exp\left\{-\frac{1}{2\sigma_\epsilon^2}\|\boldsymbol{y} - \boldsymbol{P\theta} - \boldsymbol{S}\gamma\|^2\right\},$$

which, if written as a $\boldsymbol{\theta}$, can be simplified to

$$\begin{aligned} p(\boldsymbol{y}|\gamma, \boldsymbol{\theta}, \boldsymbol{\tau}) &\propto \exp\left\{-\frac{1}{2\sigma_\epsilon^2}\left(\|\boldsymbol{y} - \boldsymbol{P\theta}\|^2 - 2(\boldsymbol{S}\gamma)^T(\boldsymbol{y} - \boldsymbol{P\theta}) + \|\boldsymbol{S}\gamma\|^2\right)\right\} \\ &\propto \exp\left\{-\frac{1}{2\sigma_\epsilon^2}\left(2(\boldsymbol{S}\gamma - \boldsymbol{y})^T\boldsymbol{P\theta} + \boldsymbol{\theta}^T\boldsymbol{P}^T\boldsymbol{P\theta}\right)\right\} \quad (1) \\ &\propto \prod_{i=1}^{G} \exp\left\{-\frac{1}{2\sigma_\epsilon^2}\left(2(\boldsymbol{S}_i\gamma_i - \boldsymbol{y}_i)^T\boldsymbol{P}_i\boldsymbol{\theta}_i + \boldsymbol{\theta}_i^T\boldsymbol{P}_i^T\boldsymbol{P}_i\boldsymbol{\theta}_i\right)\right\}. \quad (2) \end{aligned}$$

Similarly, as a function of $\boldsymbol{\gamma}$:

$$\begin{aligned} p(\boldsymbol{y}|\boldsymbol{\gamma}, \boldsymbol{\theta}, \boldsymbol{\tau}) &\propto \exp\left\{-\frac{1}{2\sigma_\epsilon^2}\left(2(\boldsymbol{P\theta} - \boldsymbol{y})^T\boldsymbol{S\gamma} + \boldsymbol{\gamma}^T\boldsymbol{S}^T\boldsymbol{S\gamma}\right)\right\} \\ &= \prod_{i=1}^{G}\prod_{j=1}^{J_i} \exp\left\{-\frac{1}{2\sigma_\epsilon^2}\left(2(\boldsymbol{X}_{ij}\boldsymbol{\theta}_i - \boldsymbol{y}_{ij})^T\boldsymbol{X}_{ij}\gamma_{ij} + \gamma_{ij}^T\boldsymbol{X}_{ij}^T\boldsymbol{X}_{ij}\gamma_{ij}\right)\right\}. \quad (3) \end{aligned}$$

1.4 Posterior

The above-mentioned model was specified from a frequentist perspective, and its equivalent Bayesian model can be obtained by assigning appropriate priors. Due to the high-dimensional nature of the posterior

$$p(\boldsymbol{\theta}, \boldsymbol{\gamma}, \boldsymbol{\tau}|\boldsymbol{y}) \propto p(\boldsymbol{y}|\boldsymbol{\theta}, \boldsymbol{\gamma}, \boldsymbol{\tau})p(\boldsymbol{\gamma}|\boldsymbol{\theta}, \boldsymbol{\tau})p(\boldsymbol{\theta}|\boldsymbol{\tau})p(\boldsymbol{\tau}), \tag{4}$$

we will sample (4) using a Gibbs sampler. The corresponding conditional distributions are presented.

1.4.1 Subject deviations

The conditional prior $p(\boldsymbol{\gamma}|\boldsymbol{\theta}, \boldsymbol{\tau})$ is modelled as truncated normally distributed, and $\boldsymbol{\gamma}_{11}, \ldots, \boldsymbol{\gamma}_{GJ_G}$ are conditionally independent

$$\begin{aligned} p(\boldsymbol{\gamma}|\boldsymbol{\theta}, \boldsymbol{\tau}) &= \prod_{i=1}^{G} p(\boldsymbol{\gamma}_i|\boldsymbol{\theta}, \boldsymbol{\tau}) \\ &= \prod_{i=1}^{G}\prod_{j=1}^{J_i} p(\boldsymbol{\gamma}_{ij}|\boldsymbol{\theta}, \boldsymbol{\tau}) \\ &\propto \mathbb{1}_{\mathcal{C}_\gamma} \prod_{i=1}^{G}\prod_{j=1}^{J_i} \exp\left\{-\frac{1}{2}\boldsymbol{\gamma}_{ij}^T \boldsymbol{\Sigma}_\gamma^{-1}\boldsymbol{\gamma}_{ij}\right\}, \end{aligned} \tag{5}$$

where $\mathcal{C}_\gamma$ is defined as

$$\mathcal{C}_\gamma \coloneqq \{\boldsymbol{\gamma} \mid \boldsymbol{A}\boldsymbol{\gamma}_{ij} \geq -\boldsymbol{A}\boldsymbol{\theta}_i, \forall i, j\};$$

and $\boldsymbol{\Sigma}_\gamma$ as

$$\boldsymbol{\Sigma}_\gamma = \begin{bmatrix} \boldsymbol{\Sigma}_b & \boldsymbol{0} \\ \boldsymbol{0} & \sigma_v^2 \boldsymbol{I} \end{bmatrix}.$$

The conditional posterior can be obtained by dividing both side of (4) by $p(\boldsymbol{\theta}, \boldsymbol{\tau}|\boldsymbol{y})$, and substituting (3) and (5):

$$\begin{aligned} p(\boldsymbol{\gamma}|\boldsymbol{y}, \boldsymbol{\theta}, \boldsymbol{\tau}) &\propto p(\boldsymbol{y}|\boldsymbol{\gamma}, \boldsymbol{\theta}, \boldsymbol{\tau})\, p(\boldsymbol{\gamma}|\boldsymbol{\theta}, \boldsymbol{\tau}) \\ &\propto \mathbb{1}_{\mathcal{C}_\gamma} \prod_{i=1}^{G}\prod_{j=1}^{J_i} \exp\left\{-\frac{1}{2\sigma_\epsilon^2}\left(2(\boldsymbol{X}_{ij}\boldsymbol{\theta}_i - \boldsymbol{y}_{ij})^T \boldsymbol{X}_{ij}\boldsymbol{\gamma}_{ij} + \boldsymbol{\gamma}_{ij}^T \boldsymbol{X}_{ij}^T \boldsymbol{X}_{ij}\boldsymbol{\gamma}_{ij}\right) - \frac{1}{2}\boldsymbol{\gamma}_{ij}^T \boldsymbol{\Sigma}_\gamma^{-1}\boldsymbol{\gamma}_{ij}\right\} \\ &\propto \mathbb{1}_{\mathcal{C}_\gamma} \prod_{i=1}^{G}\prod_{j=1}^{J_i} \exp\left\{-\frac{1}{2\sigma_\epsilon^2}\left(2(\boldsymbol{X}_{ij}\boldsymbol{\theta} - \boldsymbol{y}_{ij})^T \boldsymbol{X}_{ij}\boldsymbol{\gamma}_{ij} + \boldsymbol{\gamma}_{ij}^T (\boldsymbol{X}_{ij}^T \boldsymbol{X}_{ij} + \sigma_\epsilon^2 \boldsymbol{\Sigma}_\gamma^{-1})\boldsymbol{\gamma}_{ij}\right)\right\}. \end{aligned}$$

We can define $\boldsymbol{N}_{ij}$ as

$$\boldsymbol{N}_{ij} = \left(\boldsymbol{X}_{ij}^T \boldsymbol{X}_{ij} + \sigma_\epsilon^2 \boldsymbol{\Sigma}_\gamma^{-1}\right)^{-1},$$

and write the conditional posterior in its quadratic form

$$p(\boldsymbol{\gamma}|\boldsymbol{y}, \boldsymbol{\theta}, \boldsymbol{\tau}) \propto \mathbb{1}_{\mathcal{C}_\gamma} \prod_{i=1}^{G}\prod_{j=1}^{J_i} \exp\left\{-\frac{1}{2\sigma_\epsilon^2}\left(\boldsymbol{\gamma}_{ij} - \boldsymbol{N}_{ij}\boldsymbol{X}_{ij}^T(\boldsymbol{y}_{ij} - \boldsymbol{X}_{ij}\boldsymbol{\theta}_i)\right)^T \boldsymbol{N}_{ij}^{-1}\left(\boldsymbol{\gamma}_{ij} - \boldsymbol{N}_{ij}\boldsymbol{X}_{ij}^T(\boldsymbol{y}_{ij} - \boldsymbol{X}_{ij}\boldsymbol{\theta}_i)\right)\right\}.$$

Therefore, the individual posteriors $p(\boldsymbol{\gamma}_{ij}|\boldsymbol{y}, \boldsymbol{\theta}, \boldsymbol{\tau})$ are independent from the rest of $\boldsymbol{\gamma}$s, and are normally distributed

$$p(\boldsymbol{\gamma}_{ij}|\boldsymbol{y}, \boldsymbol{\theta}, \boldsymbol{\tau}) \sim \mathrm{TN}\left(\boldsymbol{N}_{ij}\boldsymbol{X}_{ij}^T(\boldsymbol{y}_{ij} - \boldsymbol{X}_{ij}\boldsymbol{\theta}_i), \sigma_\epsilon^2 \boldsymbol{N}_{ij}\right),$$

truncated to the set $\{\boldsymbol{\gamma}_{ij} \mid \boldsymbol{A}\boldsymbol{\gamma}_{ij} \geq -\boldsymbol{A}\boldsymbol{\theta}_i\}$.

1.4.2 Population coefficients

Firstly, we define a symmetrical idempotent matrix $\boldsymbol{K}$

$$\boldsymbol{K} = \begin{bmatrix} \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I} \end{bmatrix}$$

to extract $\boldsymbol{u}_i$ from $\boldsymbol{\theta}_i$, such that

$$\boldsymbol{K\theta}_i = \begin{bmatrix} \boldsymbol{0} \\ \boldsymbol{u}_i \end{bmatrix}, \quad \boldsymbol{\theta}_i^T \boldsymbol{K\theta}_i = \boldsymbol{u}_i^T \boldsymbol{u}_i.$$

We assign a diffuse prior to the $\boldsymbol{\beta}_i$ coefficients and truncated normal to $\boldsymbol{u}_i$, leading to the following improper prior

$$\begin{aligned} p(\boldsymbol{\theta}|\boldsymbol{\tau}) &= \prod_{i=1}^{G} p(\boldsymbol{\theta}_i|\boldsymbol{\tau}) \\ &\propto \prod_{i=1}^{G} p(\boldsymbol{\beta}_i|\boldsymbol{\tau})\, p(\boldsymbol{u}_i|\boldsymbol{\beta}_i, \boldsymbol{\tau}) \\ &\propto \mathbb{1}_{\mathcal{C}_\theta} \prod_{i=1}^{G} \exp\left\{ -\frac{1}{2\sigma_u^2} \boldsymbol{\theta}_i^T \boldsymbol{K\theta}_i \right\} \end{aligned}$$

where $\mathcal{C}_\theta$ is defined as

$$\mathcal{C}_\theta \coloneqq \{\boldsymbol{\theta} \mid \boldsymbol{A\theta}_i \geq \boldsymbol{0}, i = 1, \ldots, G\}.$$

This selection of prior is equivalent to imposing a 'squared' penalty on the spline coefficients. The conditional posterior can be obtained by dividing both sides of (4) by $p(\boldsymbol{\gamma}, \boldsymbol{\tau}|\boldsymbol{y})$

$$\begin{aligned} p(\boldsymbol{\theta}|\boldsymbol{y}, \gamma, \boldsymbol{\tau}) &\propto p(\boldsymbol{y}|\boldsymbol{\theta}, \gamma, \boldsymbol{\tau})\, p(\gamma|\boldsymbol{\theta}, \boldsymbol{\tau})\, p(\boldsymbol{\theta}|\boldsymbol{\tau}) \\ &\propto \mathbb{1}_{\mathcal{C}_\theta} \mathbb{1}_{\mathcal{C}_\gamma} \prod_{i=1}^{G} \exp\left\{ -\frac{1}{2\sigma_\epsilon^2} \left(2(\boldsymbol{S}_i\gamma_i - \boldsymbol{y}_i)^T \boldsymbol{P}_i\boldsymbol{\theta}_i + \boldsymbol{\theta}_i^T \boldsymbol{P}_i^T \boldsymbol{P}_i \boldsymbol{\theta}_i \right) - \frac{1}{2\sigma_u^2} \boldsymbol{\theta}_i^T \boldsymbol{K\theta}_i \right\} \\ &= \mathbb{1}_{\mathcal{C}_\theta} \mathbb{1}_{\mathcal{C}_\gamma} \prod_{i=1}^{G} \exp\left\{ -\frac{1}{2\sigma_\epsilon^2} \left(2(\boldsymbol{S}_i\gamma_i - \boldsymbol{y}_i)^T \boldsymbol{P}_i\boldsymbol{\theta}_i + \boldsymbol{\theta}_i^T \left(\boldsymbol{P}_i^T \boldsymbol{P}_i + \frac{\sigma_\epsilon^2}{\sigma_u^2} \boldsymbol{K} \right) \boldsymbol{\theta}_i \right) \right\}. \end{aligned}$$

If we define $\boldsymbol{M}_i$ as

$$\boldsymbol{M}_i = \left(\boldsymbol{P}_i^T \boldsymbol{P}_i + \frac{\sigma_\epsilon^2}{\sigma_u^2} \boldsymbol{K} \right)^{-1},$$

the conditional posterior can be written in its quadratic form

$$p(\boldsymbol{\theta}|\boldsymbol{y}, \boldsymbol{\gamma}, \boldsymbol{\tau}) \propto \mathbb{1}_{\mathcal{C}_\theta} \mathbb{1}_{\mathcal{C}_\gamma} \prod_{i=1}^{G} \exp\left\{ -\frac{1}{2\sigma_\epsilon^2} \left(\boldsymbol{\theta}_i - \boldsymbol{M}_i \boldsymbol{P}_i^T (\boldsymbol{y}_i - \boldsymbol{S}_i\gamma_i) \right)^T \boldsymbol{M}_i^{-1} \left(\boldsymbol{\theta}_i - \boldsymbol{M}_i \boldsymbol{P}_i^T (\boldsymbol{y}_i - \boldsymbol{S}_i\gamma_i) \right) \right\},$$

and consequently, the posterior for each population is a truncated normal distribution

$$p(\boldsymbol{\theta}_i|\boldsymbol{y}, \boldsymbol{\gamma}, \boldsymbol{\tau}) \sim \mathrm{TN}\left(\boldsymbol{M}_i \boldsymbol{P}_i^T (\boldsymbol{y}_i - \boldsymbol{S}_i\gamma_i),\, \sigma_\epsilon^2 \boldsymbol{M}_i \right)$$

truncated to $\{\boldsymbol{\theta}_i \mid \boldsymbol{A\theta}_i \geq \boldsymbol{0}, \boldsymbol{A\theta}_i \geq -\boldsymbol{A\gamma}_{ij} \text{ for } j = 1, \ldots, J_i\}$.

1.4.3 Variances

The variance terms σ_ϵ^2, σ_u^2, and σ_v^2 are all assigned an inverse Gamma prior $IG(a, b)$, hence:

$$\begin{aligned} p(\sigma_u^2|\boldsymbol{y}, \boldsymbol{\theta}, \boldsymbol{\gamma}) &\propto p(\boldsymbol{\theta}|\boldsymbol{\tau}) p(\boldsymbol{\tau}) \\ &\propto (\sigma_u^2)^{-\frac{GK}{2} - a - 1} \exp\left\{ -\frac{\boldsymbol{u}^T\boldsymbol{u} + b}{2\sigma_u^2} \right\} \\ &\sim \mathrm{IG}(0.5\, GK + a, 0.5\, \boldsymbol{u}^T\boldsymbol{u} + b), \end{aligned}$$

where $\boldsymbol{u}^T = (\boldsymbol{u}_1^T, \ldots, \boldsymbol{u}_G^T) \in \mathbb{R}^{1\times GK}$; and,

$$\begin{aligned}
p(\sigma_v^2|\boldsymbol{y},\boldsymbol{\theta},\boldsymbol{\gamma}) &\propto p(\boldsymbol{\gamma}|\boldsymbol{\theta},\boldsymbol{\tau})p(\boldsymbol{\tau}) \\
&\propto (\sigma_v^2)^{-\frac{JK}{2}-a-1} \exp\left\{-\frac{\boldsymbol{v}^T\boldsymbol{v}+b}{2\sigma_v^2}\right\} \\
&\sim \mathrm{IG}(0.5\,JK + a, 0.5\,\boldsymbol{v}^T\boldsymbol{v} + b),
\end{aligned}$$

where $\boldsymbol{v}^T = (\boldsymbol{v}_{11}^T, \boldsymbol{v}_{12}^T, \ldots, \boldsymbol{v}_{GJ_G}^T) \in \mathbb{R}^{1\times JK}$ and J the total number of subjects $J = \sum_{i=1}^G J_i$; and,

$$\begin{aligned}
p(\sigma_\epsilon^2|\boldsymbol{y},\boldsymbol{\theta},\boldsymbol{\gamma}) &\propto p(\boldsymbol{y}|\boldsymbol{\theta},\boldsymbol{\gamma},\boldsymbol{\tau})p(\boldsymbol{\tau}) \\
&\propto (\sigma_\epsilon^2)^{-\frac{N}{2}-a-1} \exp\left\{-\frac{\|\boldsymbol{y}-\boldsymbol{P\theta}-\boldsymbol{S\gamma}\|^2+b}{2\sigma_\epsilon^2}\right\} \\
&\sim \mathrm{IG}(0.5N + a, 0.5\,\|\boldsymbol{y}-\boldsymbol{P\theta}-\boldsymbol{S\gamma}\|^2 + b),
\end{aligned}$$

where N is the total number of samples $N = \sum_{i=1}^G \sum_{j=1}^{J_i} N_{ij}$. The covariance matrix $\boldsymbol{\Sigma}_b$ is assigned with an inverse-Wishart prior $W^{-1}(d, \boldsymbol{\Phi})$, hence:

$$\begin{aligned}
p(\boldsymbol{\Sigma}_b|\boldsymbol{y},\ldots) &\propto p(\boldsymbol{\theta}|\boldsymbol{\tau})p(\boldsymbol{\tau}) \\
&\propto det(\boldsymbol{\Sigma}_b)^{-\frac{J}{2}} \exp\left\{-\frac{1}{2}\sum_{i=1}^G\sum_{j=1}^{J_i} tr(\boldsymbol{b}_{ij}\boldsymbol{b}_{ij}^T\boldsymbol{\Sigma}_b^{-1})\right\} p(\boldsymbol{\Sigma}_b) \qquad \because \boldsymbol{b}_{ij}^T\boldsymbol{\Sigma}_b^{-1}\boldsymbol{b}_{ij} = tr(\boldsymbol{b}_{ij}^T\boldsymbol{\Sigma}_b^{-1}\boldsymbol{b}_{ij}) \\
&\propto det(\boldsymbol{\Sigma}_b)^{-\frac{1}{2}(J+d+dim(\boldsymbol{\Sigma}_b)+1)} \exp\left\{-\frac{1}{2}\sum_{i=1}^G\sum_{j=1}^{J_i} tr(\boldsymbol{b}_{ij}\boldsymbol{b}_{ij}^T\boldsymbol{\Sigma}_b^{-1}) - \frac{1}{2}tr(\boldsymbol{\Phi}\boldsymbol{\Sigma}_b^{-1})\right\} \\
&\sim W^{-1}(d+J, \boldsymbol{\Phi} + \boldsymbol{BB}^T),
\end{aligned}$$

where $\boldsymbol{B} = (\boldsymbol{b}_{11}, \ldots, \boldsymbol{b}_{GJ_G}) \in \mathbb{R}^{d\times J}$.