APE Note: This deliverable consists of a mathematical description of the model that I have developed during my APE, adapted from a paper by Guhaniyogi and Spencer (2021) [1]. This model will be used to estimate effects of certain behavioral covariates on high-dimensional, longitudinal brain scan outcomes in subjects with Aphasia. In order to implement this Bayesian model using Gibbs sampling, I derived the conditional posteriors for each model parameter under the stated priors (shown below). These posteriors serve as the basis of our sampling algorithm which fits the high-dimensional (tensor) regression; the next deliverable consists of an implementation of this algorithm in R, including the main functions and a brief demonstration.

1 Bayesian Tensor Response Regression (BTRR)

Suppose we have a set of tensor response variables $Y_n \in \mathbb{R}^{p_1 \times \cdots \times p_D}$ and K scalars predictors, $x_{1n}, \ldots, x_{Kn}$, $n \in [1, N]$. Furthermore, suppose each observation contains a subject identifier (ID), indexed by $i \in [1, M]$ for M subjects. In order to model the effects on each outcome element ($Y_{n,i}(v),\ v \in [1, p_1 \ldots p_D]$) and include subject-level effects, we use the following model:

$$
\begin{aligned}
Y_{n,i} &= B_i + \Gamma_1 x_{1n} + \ldots + \Gamma_K x_{Kn} + E_n \\
\Gamma_k &= \sum_{r=1}^{R} \gamma_{1,k}^{(r)} \circ \ldots \circ \gamma_{D,k}^{(r)},\ k \in [1, K] \\
B_i &= \sum_{r=1}^{R'} \beta_{1,i}^{(r)} \circ \ldots \circ \beta_{K,i}^{(r)} \\
E_n(v) &\sim \mathcal{N}(0, \sigma^2),\ v \in [1, p_1 \ldots p_D]
\end{aligned}
$$

for rank-R tensor-valued coefficients $\Gamma_1, \ldots, \Gamma_K$, subject-specific tensor intercept B_i, and error tensor E_n, which is assumed to have zero mean and equal variance across each element v. Therefore, we have the following likelihood:

$$
Y_{n,i}(v) \sim N\left(B_i(v) + \sum_{k=1}^{K} \Gamma_k(v) x_{kn},\ \sigma^2\right),\ v \in [1, p_1 \ldots p_D]
$$

<u>Priors</u>:

$$
\begin{aligned}
&\sigma^2 \sim \text{Inverse Gamma}(a_\sigma, b_\sigma) \\
&\gamma_{d,k}^{(r)} \sim \mathcal{N}(0, \tau_k W_{dr,k}),\ d \in [1, D],\ r \in [1, R],\ k \in [1, K] \\
&W_{dr,k,j} \sim \text{Exp}(\lambda_{dr,k}/2),\ j \in [1, p_d] \\
&\lambda_{dr,k} \sim \text{Gamma}(a_\lambda, b_\lambda) \\
&\tau_k \sim \text{Gamma}(a_\tau, b_\tau) \\
&\beta_{d,i}^{(r)} \sim \mathcal{N}(0, W'_{di}),\ i \in [1, M] \\
&W'_{di,j} \sim \text{Exp}(\lambda'_{di}/2),\ j \in [1, p_d] \\
&\lambda'_{di} \sim \text{Gamma}(a'_\lambda, b'_\lambda)
\end{aligned}
$$

where $W_{dr,k}, W'_{di} \in \mathbb{R}^{p_d \times p_d}$ are diagonal matrices with the jth diagonal entry denoted by $W_{dr,k,j}$ and $W'_{di,j}$, respectively.

Posterior Derivations

Let $X_n = \begin{bmatrix} x_{1n} & \dots & x_{Kn} \end{bmatrix}$. Let Ω_n be the subset of voxel indices $v = 1, \dots, p_1 \dots p_D$ which are available for the nth observation Y_n. Denote $v_n = |\Omega_n|$ to be the number of voxels in Y_n; let $\Omega_{n,dj}$ be the subset of voxel indices $v = 1, \dots, \frac{p_1 \dots p_D}{p_d}$ corresponding to the sub-tensor obtained by fixing the dth dimension of Y_n at index $j \in [1, p_d]$.

1. Noise Variance: σ^2

$$P(\sigma^2|Y_{ni}, X_n, \dots, n = 1, \dots, N) \propto \prod_{n=1}^{N} P(Y_{ni}|\sigma^2, \dots)P(\sigma^2)$$

$$\propto \left(\prod_{n=1}^{N} \prod_{v \in \Omega_n} \sigma^{-1} \right) \exp\left[-\frac{1}{2\sigma^2} \sum_{n=1}^{N} \sum_{v \in \Omega_n} [Y_n(v) - B_i(v) - x_{1n}\Gamma_1(v) - \dots - x_{1K}\Gamma_K(v)]^2 \right] \left(\sigma^2\right)^{-a_\sigma - 1} \exp\left[-\frac{b_\sigma}{\sigma^2} \right]$$

$$= \left(\sigma^2\right)^{-\left(a_\sigma + \frac{v_1 + \dots + v_N}{2}\right) - 1} \exp\left[-\frac{\frac{1}{2}\sum_{n=1}^{N} \sum_{v \in \Omega_n} [Y_n(v) - B_i(v) - x_{1n}\Gamma_1(v) - \dots - x_{Kn}\Gamma_K(v)]^2}{\sigma^2} \right]$$

$$\implies \sigma^2|Y_n, \dots \sim \text{IG}\left(a_\sigma + \frac{(v_1 + \dots + v_N)}{2},\ b_\sigma + \frac{1}{2} \sum_{n=1}^{N} \sum_{v \in \Omega_n} [Y_n(v) - B_i(v) - x_{1n}\Gamma_1(v) - \dots - x_{Kn}\Gamma_K(v)]^2 \right)$$

2. Tensor margins: $\gamma_{d,k}^{(r)}$

For the kth variable, the rth rank, and the dth dimension, first remove parameters from other variables (k^*), other ranks (r^*), and subject-specific intercepts (B_i).

$$Y_{nk} = Y_{n,i} - B_i - \Gamma_1 X_{n,1} - \dots - \Gamma_K X_{n,K}$$

$$Y_{nkr} = Y_{nk} - \sum_{\substack{r^*=1 \\ r^* \neq r}}^{R} \gamma_{1,k}^{(r^*)} \circ \dots \circ \gamma_{1,k}^{(r^*)}$$

Let $C = \gamma_1 \circ \dots \circ \gamma_{d-1} \circ \gamma_{d+1} \circ \dots \circ \gamma_D$, let $s_{dr,k,j} = \tau W_{dr,k,j}$, and denote $Y_{nkr}(j, v')$ as the element of Y_{nkr} given from j and v', with $j \in [1, p_d]$ and $v' \in \Omega_{n,dj}$ (voxel index of sub-tensor). Then we have,

$$
\begin{aligned}
&P\left(\gamma_{d,k}^{(r)}|Y_n,\dots\right) \propto \prod_{n=1}^{N} P\left(Y_{nkr}|\gamma_{d,k}^{(r)},\dots\right) P\left(\gamma_{d,k}^{(r)}\right) \\
&\propto \exp\left[-\frac{1}{2\sigma^2}\sum_{n=1}^{N}\sum_{j=1}^{p_d}\sum_{v'\in\Omega_{n,dj}}\left(Y_{nkr,j}-\gamma_{d,k,j}^{(r)}\ X_{n,k}C(v')\right)^2\right]\exp\left[-\frac{1}{2}\sum_{j=1}^{p_d}\frac{1}{s_{dr,k,j}}\gamma_{d,k,j}^{(r)2}\right] \\
&\propto \exp\left[-\frac{1}{2\sigma^2}\sum_{j=1}^{p_d}\left(\sum_{n=1}^{N}\sum_{v'\in\Omega_{n,dj}}\gamma_{d,k,j}^{(r)2}\ X_{n,k}^2C^2(v')-2\gamma_{d,k,j}^{(r)}X_{n,k}Y_{nkr}(j,v')C(v')+\frac{\sigma^2}{s_{dr,k,j}}\gamma_{d,k,j}^{(r)2}\right)\right] \\
&= \exp\left[-\frac{1}{2\sigma^2}\sum_{j=1}^{p_d}\left(\gamma_{d,k,j}^{(r)2}\left(\sum_{n=1}^{N}\sum_{v'\in\Omega_{n,dj}}X_{n,k}^2C^2(v')+\frac{\sigma^2}{s_{dr,k,j}}\right)-2\gamma_{d,k,j}^{(r)}\sum_{n=1}^{N}X_{n,k}\sum_{v'\in\Omega_{n,dj}}Y_{nkr}(j,v')C(v')\right)\right] \\
&\propto \exp\left[-\frac{1}{2\sigma^2}\sum_{j=1}^{p_d}A_j\left(\gamma_{d,k,j}^{(r)}-\frac{B_j}{A_j}\right)^2\right],\ A_j=\sum_{n}\sum_{v'\in\Omega_{n,dj}}X_{n,k}^2C^2(v')+\frac{\sigma^2}{s_{dr,k,j}},\ B_j=\sum_{n}X_{n,k}\sum_{v'\in\Omega_{n,dj}}Y_{nkr}(j,v')C(v') \\
&\implies \gamma_{d,k}^{(r)}|Y_n,\dots\sim\mathcal{N}\left(\text{diag}^{-1}(A)B,\ \sigma^2\text{diag}^{-1}(A)\right), A=[A_1,\dots,A_{p_d}]^T, B=[B_1,\dots,B_{p_d}]^T
\end{aligned}
$$

3. Tensor margin covariance (diagonal element): $W_{dr,k,j}$

$$
\begin{aligned}
P\left(W_{dr,k,j}|\gamma_{d,k,j}^{(r)},\tau_k,\lambda_{dr,k}\right) &\propto P\left(\gamma_{d,k,j}^{(r)}|W_{dr,k,j},\tau_k\right)P\left(W_{dr,k,j}|\lambda_{dr,k}\right)\\
&\propto \left(W_{dr,k,j}\right)^{-\frac{1}{2}}\exp\left[-\frac{\left(\gamma_{d,k,j}^{(r)}\right)^2}{2\tau_k W_{dr,k,j}}\right]\exp\left[-\frac{\lambda_{dr,k}W_{dr,k,j}}{2}\right]\\
&= \left(W_{dr,k,j}\right)^{-\frac{1}{2}}\exp\left[-\frac{1}{2}\left(\lambda_{dr,k}W_{dr,k,j}+\frac{\left(\gamma_{d,k,j}^{(r)}\right)^2}{\tau_k W_{dr,k,j}}\right)\right]\\
\implies W_{dr,k,j}|\gamma_{d,k,j}^{(r)},\ldots &\sim \text{gIG}\left(\mu=\frac{1}{2},\ \chi=\frac{\left(\gamma_{d,k,j}^{(r)}\right)^2}{\tau_k},\ \psi=\lambda_{dr,k}\right)
\end{aligned}
$$

4. Rate parameter: $\lambda_{dr,k}$

$$
\begin{aligned}
P(\lambda_{dr,k}|W_{dr,k},a_\lambda,b_\lambda) &\propto P(W_{dr,k}|\lambda_{dr,k})P(\lambda_{dr,k}|a_\lambda,b_\lambda)\\
&\propto \left(\frac{\lambda_{dr,k}}{2}\right)^{p_d}\exp\left[-\left(\frac{\lambda_{dr,k}}{2}\right)\sum_{j=1}^{p_d}W_{dr,k,j}\right]\left(\lambda_{dr,k}\right)^{(a_\lambda-1)}\exp(-b_\lambda\lambda_{dr,k})\\
&\propto \left(\lambda_{dr,k}\right)^{a_\lambda+p_d-1}\exp\left[-\left(b_\lambda+\frac{1}{2}\operatorname{Tr}(W_{dr,k})\right)\lambda_{dr,k}\right]\\
\implies \lambda_{dr,k}|W_{dr,k},\ldots &\sim \text{Gamma}\left(a'_\lambda=a_\lambda+p_d,\ b'_\lambda=b_\lambda+\frac{1}{2}\operatorname{Tr}(W_{dr,k})\right)
\end{aligned}
$$

5. Global scale factor: τ_k

$$
\begin{aligned}
&P\left(\tau_k|\gamma_{d,k}^{(r)},W_{dr,k},a_\tau,b_\tau,d\in[1,D],r\in[1,R]\right)\\
&\propto \prod_{r=1}^{R}\prod_{d=1}^{D}P\left(\gamma_{d,k}^{(r)}|\tau_k,\ldots\right)P\left(\tau_k|a_\tau,b_\tau\right)\\
&\propto \tau^{-\frac{R(p_1+\ldots+p_D)}{2}}\exp\left[-\frac{1}{2\tau_k}\sum_{r=1}^{R}\sum_{d=1}^{D}\left(\gamma_{d,k}^{(r)T}W_{dr,k}^{-1}\gamma_{d,k}^{(r)}\right)\right]\tau_k^{(a_\tau-1)}\exp\left[-b_\tau\tau_k\right]\\
&= \tau_k^{a_\tau-\frac{R(p_1+\ldots+p_D)}{2}-1}\exp\left[-\frac{1}{2}\left(\frac{\sum_{r=1}^{R}\sum_{d=1}^{D}\gamma_{d,k}^{(r)T}W_{dr,k}^{-1}\gamma_{d,k}^{(r)}}{\tau_k}\right)-\frac{1}{2}(2b_\tau\tau_k)\right]\\
&\implies \tau_k\sim\text{gIG}\left(\mu=a_\tau-\frac{R(p_1+\ldots+p_D)}{2},\ \chi=\sum_{r=1}^{R}\sum_{d=1}^{D}\gamma_{d,k}^{(r)T}W_{dr,k}^{-1}\gamma_{d,k}^{(r)},\ \psi=2b_\tau\right)
\end{aligned}
$$

6. Random intercept tensor margins: $\beta_{d,i}^{(r)}$

Let X_{ni} denote the covariate observations for the ith subject. Remove overall effects (Γ's) and other rank effects r^*:

$$\widetilde{Y_{ni}} = Y_{n,i} - B_i - \Gamma_1 X_{ni,1} - \ldots - \Gamma_K X_{ni,K}$$
$$\widetilde{Y_{nir}} = \widetilde{Y_{ni}} - \sum_{\substack{r^*=1\\ r^*\neq r}}^{R'} \beta_{1,i}^{(r^*)} \circ \ldots \circ \beta_{1,i}^{(r^*)}$$

Let $C' = \beta_1 \circ \ldots \circ \beta_{d-1} \circ \beta_{d+1} \circ \ldots \circ \beta_D$ and denote $\widetilde{Y_{nir}}(j, v')$ as the element of $\widetilde{Y_{nir}}$ given from j and v', with $j \in [1, p_d]$ and $v' \in \Omega_{n,dj}$ (voxel index of sub-tensor). Lastly let N_i be the number of observations for the ith subject. Then we have,

$$P\left(\beta_{d,i}^{(r)}|Y_n, \ldots\right) \propto \prod_{n=1}^{N_i} P\left(\widetilde{Y_{nir}}|\beta_{d,i}^{(r)}, \ldots\right) P\left(\beta_{d,i}^{(r)}\right)$$
$$\propto \exp\left[-\frac{1}{2\sigma^2}\sum_{n=1}^{N_i}\sum_{j=1}^{p_d}\sum_{v'\in\Omega_{n,dj}}\left(\widetilde{Y_{nir}}(j,v') - \beta_{d,i,j}^{(r)}\ C'(v')\right)^2\right]\exp\left[-\frac{1}{2}\sum_{j=1}^{p_d}\frac{\beta_{d,i,j}^{(r)2}}{W_{dr,i,j}}\right]$$
$$\propto \exp\left[-\frac{1}{2\sigma^2}\sum_{j=1}^{p_d}\left(\sum_{n=1}^{N_i}\sum_{v'\in\Omega_{n,dj}}\beta_{d,i,j}^{(r)2}C'^2(v') - 2\beta_{d,i,j}^{(r)}\widetilde{Y_{nir}}(j,v')C(v') + \frac{\sigma^2}{W_{dr,i,j}}\beta_{d,i,j}^{(r)2}\right)\right]$$
$$= \exp\left[-\frac{1}{2\sigma^2}\sum_{j=1}^{p_d}\left(\beta_{d,i,j}^{(r)2}\left(\sum_{n=1}^{N_i}\sum_{v'\in\Omega_{n,dj}}C'^2(v') + \frac{\sigma^2}{W_{dr,i,j}}\right) - 2\beta_{d,i,j}^{(r)}\sum_{n=1}^{N_i}\sum_{v'\in\Omega_{n,dj}}\widetilde{Y_{nir}}(j,v')C'(v')\right)\right]$$
$$\propto \exp\left[-\frac{1}{2\sigma^2}\sum_{j=1}^{p_d}A'_j\left(\beta_{d,i,j}^{(r)} - \frac{B'_j}{A'_j}\right)^2\right],\ A'_j = \sum_{n=1}^{N_i}\sum_{v'\in\Omega_{n,dj}}C'^2(v') + \frac{\sigma^2}{W_{dr,i,j}},\ B'_j = \sum_{n=1}^{N_i}\sum_{v'\in\Omega_{n,dj}}\widetilde{Y_{nir}}(j,v')C'(v')$$
$$\implies \beta_{d,i}^{(r)}|Y_{ni}, \ldots \sim \mathcal{N}\left(\mathrm{diag}^{-1}(A')B',\ \sigma^2\mathrm{diag}^{-1}(A')\right),\ A' = [A'_1, \ldots, A'_{p_d}]^T, B' = [B'_1, \ldots, B'_{p_d}]^T$$

7. Random intercept tensor margin covariance (diagonal element): $W'_{dr,i,j}$

$$P\left(W'_{dr,i,j}|\beta_{d,i,j}^{(r)}, \lambda'_{dr,i}\right) \propto P\left(\beta_{d,i,j}^{(r)}|W'_{dr,i,j}\right)P\left(W'_{dr,i,j}|\lambda'_{dr,i}\right)$$
$$\propto (W'_{dr,i,j})^{-\frac{1}{2}}\exp\left[-\frac{\left(\beta_{d,i,j}^{(r)}\right)^2}{2W'_{dr,i,j}}\right]\exp\left[-\frac{\lambda'_{dr,i}W'_{dr,i,j}}{2}\right]$$
$$= (W'_{dr,i,j})^{-\frac{1}{2}}\exp\left[-\frac{1}{2}\left(\lambda'_{dr,i}W'_{dr,i,j} + \frac{\left(\beta_{d,i,j}^{(r)}\right)^2}{W'_{dr,i,j}}\right)\right]$$
$$\implies W'_{dr,i,j}|\beta_{d,i,j}^{(r)}, \ldots \sim \mathrm{gIG}\left(\mu = \frac{1}{2},\ \chi = \left(\beta_{d,i,j}^{(r)}\right)^2,\ \psi = \lambda'_{dr,i}\right)$$

8. Random intercept Rate parameter: $\lambda'_{dr,i}$

$$
\begin{aligned}
&P(\lambda'_{dr,i}|W'_{dr,i}, a'_\lambda, b'_\lambda) \propto P(W'_{dr,i}|\lambda'_{dr,i})P(\lambda'_{dr,i}|a'_\lambda, b'_\lambda) \\
&\propto \left(\frac{\lambda'_{dr,i}}{2}\right)^{p_d} \exp\left[-\left(\frac{\lambda'_{dr,i}}{2}\right)\sum_{j=1}^{p_d} W'_{dr,i,j}\right] \left(\lambda'_{dr,i}\right)^{(a'_\lambda - 1)} \exp(-b'_\lambda \lambda'_{dr,i}) \\
&\propto \left(\lambda'_{dr,i}\right)^{a'_\lambda + p_d - 1} \exp\left[-\left(b'_\lambda + \frac{1}{2}\operatorname{Tr}(W'_{dr,i})\right)\lambda'_{dr,i}\right] \\
&\implies \lambda'_{dr,i}|W'_{dr,i}, \ldots \sim \text{Gamma}\left(a_\lambda{}'' = a'_\lambda + p_d,\ b_\lambda{}'' = b'_\lambda + \frac{1}{2}\operatorname{Tr}(W'_{dr,i})\right)
\end{aligned}
$$

References

[1] Rajarshi Guhaniyogi and Daniel Spencer. "Bayesian Tensor Response Regression with an Application to Brain Activation Studies". In: *Bayesian Analysis* (2021), pp. 1–29.