

Supplementary Material for DAME

Bomin Kim

October 11, 2017

1 Modeling Framework

1.1 Model Formulation

Our interest is simultaneously modeling the sequence of $N \times N$ time-varying symmetric matrices $\mathbf{Y} = \{Y(t), t \in \mathcal{T}\}$ having entries y_{ijt} , where $\mathcal{T} = \{t_1, \dots, t_T\}$ denote the time grid on which networks are observed. We define the model as

$$y_{ijt} = \mathbf{X}_{ijt}^T \boldsymbol{\beta}_t + z_{ijt} \quad t \in \mathcal{T}, \quad (1)$$

independently for each $i = 2, \dots, N$ and $j = 1, \dots, i-1$, where $X_{ijt} = [X_{ijt}^1, \dots, X_{ijt}^P]^T$ is a P -dimensional vector of edge covariates, $\boldsymbol{\beta}_t = [\beta_{1t}, \dots, \beta_{Pt}]^T$ is the corresponding dynamic coefficients, and z_{ijt} is the unobserved random effects. Following the latent factor model, we also model z_{ijt} to account for potential dependencies in the additive and multiplicative form

$$z_{ijt} = \theta_{it} + \theta_{jt} + \mathbf{u}_{it}^T D_t \mathbf{u}_{jt} + \epsilon_{ijt}, \quad (2)$$

where θ_{it} and θ_{jt} are the additive nodal random effects of i and j , respectively, $\mathbf{u}_{it}^T D_t \mathbf{u}_{jt}$ is the multiplicative random effect with D_t denoting the $R \times R$ diagonal matrix (i.e. $D_t = \text{diag}(d_{1t}, \dots, d_{Rt})$) and $\mathbf{u}_{it} = [u_{i1t}, \dots, u_{iRt}]^T$ denoting the vector of latent coordinates of node i , and ϵ_{ijt} is the noise.

1.2 Prior Specification

We assume independent Gaussian process priors for the parameters $\{\beta_p(\cdot)\}_{p=1}^P$, $\{\theta_i(\cdot)\}_{i=1}^N$, $\{d_r(\cdot)\}_{r=1}^R$ and $\{\{u_{ir}(\cdot)\}_{i=1}^N\}_{r=1}^R$. We also assign additional hierarchy to include a shrinkage parameters $\{\tau_p^\beta\}_{p=1}^P$, τ^θ , and $\{\tau_r^u\}_{r=1}^R$ using Inverse-Gamma (IG) priors. Specifically, we let

1. $\beta_p(\cdot) \sim \text{GP}(0, \tau_p^\beta c_\beta)$ for $p = 1, \dots, P$, where $\tau_p^\beta \sim \text{IG}(a_\beta, b_\beta)$ and $c_\beta(t, t') = f(\kappa_p^\beta, |t - t'|)$;
2. $\theta_i(\cdot) \sim \text{GP}(0, \tau^\theta c_\theta)$ for $i = 1, \dots, N$, where $\tau^\theta \sim \text{IG}(a_\theta, b_\theta)$ and $c_\theta(t, t') = f(\kappa^\theta, |t - t'|)$;
3. $d_r(\cdot) \sim \text{GP}(0, c_d)$ for $r = 1, \dots, R$, where $c_d(t, t') = f(\kappa^d, |t - t'|)$;
4. $u_{ir}(\cdot) \sim \text{GP}(0, \tau_r^u I_T)$ for $i = 1, \dots, N$ and $r = 1, \dots, R$, where $\tau_r^u \sim \text{IG}(a_u, b_u)$;
(i.e., $u_{ti}(\cdot) \sim \text{MVN}_R(0, \text{diag}(\tau_1^u, \dots, \tau_r^u))$ for $i = 1, \dots, N$)
5. $\epsilon_{ij} \sim \text{N}(0, \sigma_e^2)$, where $\sigma_e^2 \sim \text{IG}(a_\sigma, b_\sigma)$,

where $\beta_p(\cdot) = [\beta_p(t_1), \dots, \beta_p(t_T)]^T$ and the rest of vectors with (\cdot) defined in the same manner, and I_T is $T \times T$ identity matrix. Since the multiplicative random effect is measured by $u_i(t)^T D(t) u_j(t)$, we assign temporal correlation structure to the diagonals $d_r(\cdot)$ and shrinkage parameter τ_u^r to the latent coordinates $u_{ir}(\cdot)$, in order to prevent from non-identifiability issue. To specify the form of Gaussian process, we use the standard Exponential covariance function:

$$f(\kappa, |t - t'|) = \exp\left(-\frac{|t - t'|}{\kappa}\right),$$

where κ is the positive length parameter. We choose this parameterization such that $\kappa \approx 0$ implies independence across time and larger κ implies higher correlation between timepoints. To estimate κ , we use the uninformative prior $\kappa \sim \text{Uniform}(0.001, 10)$.

1.3 Posterior Computation

Posterior computation is performed via simple Gibbs sampler to update the vector of time-varying regression coefficients and the vector of additive and multiplicative latent factors. Using conjugate prior distributions, this section provides a Gibbs sampling scheme that approximates the joint posterior distribution over the variables $\boldsymbol{\beta} = \{\beta_p(\cdot)\}_{p=1}^P$, $\boldsymbol{\theta} = \{\theta_i(\cdot)\}_{i=1}^N$, $\mathbf{d} = \{d_r(\cdot)\}_{r=1}^R$, $\mathbf{u} = \{\{u_{ir}(\cdot)\}_{i=1}^N\}_{r=1}^R$,

$\boldsymbol{\tau}^\beta = \{\tau_p^\beta\}_{p=1}^P, \tau^\theta, \boldsymbol{\tau}^u = \{\tau_r^u\}_{r=1}^R, \boldsymbol{\kappa}^\beta = \{\kappa_p^\beta\}_{p=1}^P, \kappa^\theta, \kappa^d$, and σ_e^2 , given the observed data (\mathbf{Y}, \mathbf{X}) and the hyperparameters $(a_\beta, b_\beta, a_\theta, b_\theta, a_u, b_u, a_\sigma, b_\sigma)$. Therefore, our inference goal is to draw samples from the posterior distribution

$$\begin{aligned}
& P(\boldsymbol{\beta}, \boldsymbol{\theta}, \mathbf{d}, \mathbf{u}, \boldsymbol{\tau}^\beta, \tau^\theta, \boldsymbol{\tau}^u, \boldsymbol{\kappa}^\beta, \kappa^\theta, \kappa^d, \sigma_e^2 | \mathbf{Y}, \mathbf{X}, a_\beta, b_\beta, a_\theta, b_\theta, a_u, b_u, a_\sigma, b_\sigma) \\
& \propto P(\boldsymbol{\beta}, \boldsymbol{\theta}, \mathbf{d}, \mathbf{u}, \boldsymbol{\tau}^\beta, \tau^\theta, \boldsymbol{\tau}^u, \boldsymbol{\kappa}^\beta, \kappa^\theta, \kappa^d, \sigma_e^2, \mathbf{Y} | \mathbf{X}, a_\beta, b_\beta, a_\theta, b_\theta, a_u, b_u, a_\sigma, b_\sigma) \\
& = P(\boldsymbol{\tau}^\beta | a_\beta, b_\beta) P(\boldsymbol{\kappa}^\beta) P(\boldsymbol{\beta} | \mathbf{X}, \boldsymbol{\tau}^\beta, \boldsymbol{\kappa}^\beta) P(\tau^\theta | a_\theta, b_\theta) P(\kappa^\theta) P(\boldsymbol{\theta} | \tau^\theta, \kappa^\theta) P(\kappa^d) P(\mathbf{d} | \kappa^d) P(\boldsymbol{\tau}^u | a_u, b_u) P(\mathbf{u} | \boldsymbol{\tau}^u) \\
& \quad \times P(\sigma_e^2 | a_\sigma, b_\sigma) P(\mathbf{Y} | \mathbf{X}, \boldsymbol{\beta}, \boldsymbol{\theta}, \mathbf{d}, \mathbf{u}, \sigma_e^2)
\end{aligned} \tag{3}$$

All of the $P()$ can be easily obtained from their prior specifications in Section 1.2, except the last one $P(\mathbf{Y} | \mathbf{X}, \boldsymbol{\beta}, \boldsymbol{\theta}, \mathbf{d}, \mathbf{u}, \sigma_e^2)$. To obtain this, we introduce the noise array \mathbf{E} . Let $\mathbf{E} = [E(t_1), \dots, E(t_T)]^T$ denote the noise array over the time grid, where the $(i, j)^{th}$ entry of the matrix $E(t)$ is defined as $E_{ij}(t) = y_{ij}(t) - (\sum_{p=1}^P X_{ij}^{pT}(t) \beta_p(t) + \theta_i(t) + \theta_j(t) + u_i(t)^T D(t) u_j(t))$. Given that the distribution of the observed network $\mathbf{Y} = \{Y(t_1), \dots, Y(t_T)\}$ conditional on all the parameters can be written as

$$P(\mathbf{Y} | \mathbf{X}, \boldsymbol{\beta}, \boldsymbol{\theta}, \mathbf{d}, \mathbf{u}, \sigma_e^2) \propto \prod_{t=1}^T \prod_{i>j} (\sigma_e^2)^{-\frac{1}{2}} \exp\{-\frac{1}{2\sigma_e^2} \|E_{ij}(t)\|^2\}. \tag{4}$$

Now, we want to sequentially update each parameter $(\boldsymbol{\beta}, \boldsymbol{\theta}, \mathbf{d}, \mathbf{u}, \boldsymbol{\tau}^\beta, \tau^\theta, \boldsymbol{\tau}^u, \boldsymbol{\kappa}^\beta, \kappa^\theta, \kappa^d, \sigma_e^2)$ from its full conditional distributions. Although we derive full conditionals of τ 's using Normal-inverse Gamma conjugacy (when assuming κ fixed), we still use Metropolis-Hastings algorithm for multivariate sampling of the two Gaussian process parameters (τ, κ) due to their high correlation (when assuming κ not fixed and thus estimable).

1. Sample $\sigma_e^2 \sim \text{IG}(\frac{T \cdot N(N-1)}{4} + a_\sigma, \frac{1}{2} \sum_{t=1}^T \sum_{i>j} (E_{ij}(t))^2 + b_\sigma)$;
2. For each $p \in \{1, \dots, P\}$ in random order, sample $\beta_p(\cdot)$ as follows:

$$(a) \text{ (Fixed } \kappa_p^\beta) \text{ Sample } \tau_p^\beta \sim \text{IG}(\frac{T}{2} + a_\beta, \frac{1}{2} \beta_p'(\cdot) c_\beta^{-1} \beta_p(\cdot) + b_\beta);$$

$$(a') \text{ (Estimate } \kappa_p^\beta) \text{ Sample } (\tau_p^\beta, \kappa_p^\beta) \text{ using Metropolis-Hastings algorithm}$$

(b) Sample $\beta_p(\cdot) \sim N_T(\tilde{\mu}_{\beta_p}, \tilde{\Sigma}_{\beta_p})$, where

$$\tilde{\Sigma}_{\beta_p} = \left((\tau_p^\beta c_\beta)^{-1} + \frac{\text{diag}(\{\sum_{i>j} X_{ij}^{p2}(t)\}_{t_1}^{t_T})}{\sigma_e^2} \right)^{-1} \text{ and } \tilde{\mu}_{\beta_p} = \left(\frac{\{\sum_{i>j} (E_{ij}^{-p}(t) X_{ij}^p(t))\}_{t_1}^{t_T}}{\sigma_e^2} \right) \tilde{\Sigma}_{\beta_p}$$

$$(\text{NOTE: } E_{ij}^{-p}(t) = E_{ij}(t) + X_{ij}^{pT}(t) \beta_p(t))$$

3. Sample $\theta_i(\cdot)$ as follows:

(a) (Fixed κ^θ) Sample $\tau^\theta \sim \text{IG}(\frac{T \cdot N}{2} + a_\theta, \frac{1}{2} \sum_{i=1}^N \theta'_i(\cdot) c_\theta^{-1} \theta_i(\cdot) + b_\theta)$;

(a') (Estimate κ^θ) Sample $(\tau^\theta, \kappa^\theta)$ using Metropolis-Hastings algorithm

(b) For each $i \in \{1, \dots, N\}$ in random order, sample $\theta_i(\cdot) \sim N_T(\tilde{\mu}_{\theta_i}, \tilde{\Sigma}_{\theta_i})$, where

$$\tilde{\Sigma}_{\theta_i} = \left((\tau^\theta c_\theta)^{-1} + \frac{(N-1)I_T}{\sigma_e^2} \right)^{-1} \text{ and } \tilde{\mu}_{\theta_i} = \left(\frac{\{\sum_{i=i, j \neq i} E_{ij}^{-i}(t)\}_{t_1}^{t_T}}{\sigma_e^2} \right) \tilde{\Sigma}_{\theta_i}$$

$$(\text{NOTE: } E_{ij}^{-i}(t) = E_{ij}(t) + \theta_i(t))$$

4. For each $r \in \{1, \dots, R\}$ in random order, sample $d_r(\cdot)$ as follows:

(a) (Estimate κ^d) Sample κ^d using (univariate) Metropolis-Hastings algorithm

(b) Sample $d_r(\cdot) \sim N_T(\tilde{\mu}_{d_r}, \tilde{\Sigma}_{d_r})$, where

$$\tilde{\Sigma}_{d_r} = \left(c_d^{-1} + \frac{\text{diag}(\{\sum_{i>j} (u_{ir}(t) u_{jr}(t))^2\}_{t_1}^{t_T})}{\sigma_e^2} \right)^{-1} \text{ and } \tilde{\mu}_{d_r} = \left(\frac{\{\sum_{i>j} (E_{ij}^{-r}(t) u_{ir}(t) u_{jr}(t))\}_{t_1}^{t_T}}{\sigma_e^2} \right) \tilde{\Sigma}_{d_r}$$

$$(\text{NOTE: } E_{ij}^{-u}(t) = E_{ij}(t) + u_{i[-r]}(t)^T D_{[-r, -r]}(t) u_{j[-r]}(t);)$$

5. For each $t \in \{1, \dots, T\}$ and $i \in \{1, \dots, N\}$ in random order, sample $u_i(t)$ as follows:

(a) Sample $\tau_r^u \sim \text{IG}(\frac{T \cdot N}{2} + a_u, \frac{1}{2} \sum_{i=1}^N u'_{ir}(\cdot) u_{ir}(\cdot) + b_u)$;

(b) Sample $u_i(t) \sim N_R(\tilde{\mu}_{u_i(t)}, \tilde{\Sigma}_{u_i(t)})$, where

$$\tilde{\Sigma}_{u_{ir}(t)} = \left(\text{diag}(\boldsymbol{\tau}^u)^{-1} + \frac{\sum_{j \neq i} D(t) u_j(t) u'_j(t) D(t)}{\sigma_e^2} \right)^{-1} \text{ and } \tilde{\mu}_{u_{ir}(t)} = \left(\frac{\{\sum_{i=i, j \neq i} (E_{ij}^{-u}(t) u'_j(t) D(t))'\}_{r1}^R}{\sigma_e^2} \right) \tilde{\Sigma}_{u_{ir}(t)}$$

$$(\text{NOTE: } E_{ij}^{-u}(t) = E_{ij}(t) + u_i^T(t) D(t) u_j(t);)$$

Note that between every steps 1-5, $\mathbf{E} = [E(t_1), \dots, E(t_T)]^T$ has to be calculated again using the previously updated values, so that the next update is conditioned on the current values of all the other parameters. Except Gaussian process length parameters $(\kappa^\beta, \kappa^\theta, \kappa^d)$, we know all the full conditional distributions, we achieve easy implementation and fast computation via Markov Chain Monte Carlo (MCMC) with Gibbs sampling. For κ 's, we use the simplest case of Metropolis-Hastings algorithm using Normal proposals.

2 Appendix: Derivation

2.1 Noise error variance σ_e^2

$$\begin{aligned}
P(\sigma_e^2 | \mathbf{Y}, a_\sigma, b_\sigma) &\propto P(\mathbf{Y} | \mathbf{X}, \boldsymbol{\beta}, \boldsymbol{\theta}, \mathbf{d}, \mathbf{u}, \sigma_e^2) P(\sigma_e^2 | a_\sigma, b_\sigma) \\
&\propto \prod_{t=1}^T \prod_{i>j} (\sigma_e^2)^{-\frac{1}{2}} \exp\left\{-\frac{1}{2\sigma_e^2} \|y_{ij}(t) - \left(\sum_{p=1}^P X'_{ij}(t)\beta_p(t) + \theta_i(t) + \theta_j(t) + u_i(t)^T D(t)u_j(t)\right)\|^2\right\} \\
&\quad \times (\sigma_e^2)^{-a_\sigma-1} \exp\left\{\frac{1}{\sigma_e^2} b_\sigma\right\} \\
&= (\sigma_e^2)^{-\frac{T}{2} \cdot \frac{N(N-1)}{2} - a_\sigma - 1} \\
&\quad \times \exp\left\{-\frac{1}{\sigma_e^2} \left(\frac{1}{2} \sum_{t=1}^T \sum_{i>j} \|y_{ij}(t) - \left(\sum_{p=1}^P X'_{ij}(t)\beta_p(t) + \theta_i(t) + \theta_j(t) + u_i(t)^T D(t)u_j(t)\right)\|^2 + b_\sigma\right)\right\} \\
&\sim \text{IG}\left(\frac{T \cdot N(N-1)}{4} + a_\sigma, \frac{1}{2} \sum_{t=1}^T \sum_{i>j} (E_{ij}(t))^2 + b_\sigma\right)
\end{aligned}$$

2.2 Gaussian Process variance parameter τ 's

1. τ_p^β , for $p = 1, \dots, P$:

$$\begin{aligned}
P(\tau_p^\beta | \beta_p, a_\beta, b_\beta) &\propto P(\beta_p | \tau_p^\beta, \kappa^\beta) \times P(\tau_p^\beta | a_\beta, b_\beta) \\
&\propto |\tau_p^\beta c_\beta|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2} \left(\beta_p'(\cdot) (\tau_p^\beta c_\beta)^{-1} \beta_p(\cdot)\right)\right\} \times (\tau_p^\beta)^{-a_\beta-1} \exp\left\{-\frac{1}{\tau_p^\beta} b_\beta\right\} \\
&\propto (\tau_p^\beta)^{-\frac{T}{2} - a_\beta - 1} \exp\left\{-\frac{1}{\tau_p^\beta} \left(\frac{1}{2} \beta_p'(\cdot) c_\beta^{-1} \beta_p(\cdot) + b_\beta\right)\right\} \\
&\sim \text{IG}\left(\frac{T}{2} + a_\beta, \frac{1}{2} \beta_p'(\cdot) c_\beta^{-1} \beta_p(\cdot) + b_\beta\right)
\end{aligned}$$

2. τ^θ :

$$\begin{aligned}
P(\tau^\theta | \boldsymbol{\theta}, a_\theta, b_\theta) &\propto \prod_{i=1}^N P(\theta_i | \tau^\theta, \kappa^\theta) \times P(\tau^\theta | a_\theta, b_\theta) \\
&\propto \prod_{i=1}^N |\tau^\theta c_\theta|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2} \left(\theta'_i(\cdot) (\tau^\theta c_\theta)^{-1} \theta_i(\cdot) \right)\right\} \times (\tau^\theta)^{-a_\theta-1} \exp\left\{-\frac{1}{\tau^\theta} b_\theta\right\} \\
&\propto (\tau^\theta)^{-\frac{T \cdot N}{2} - a_\theta - 1} \exp\left\{-\frac{1}{\tau^\theta} \left(\frac{1}{2} \sum_{i=1}^N \theta'_i(\cdot) c_\theta^{-1} \theta_i(\cdot) + b_\theta \right)\right\} \\
&\sim \text{IG}\left(\frac{T \cdot N}{2} + a_\theta, \quad \frac{1}{2} \sum_{i=1}^N \theta'_i(\cdot) c_\theta^{-1} \theta_i(\cdot) + b_\theta\right)
\end{aligned}$$

3. τ_r^u , for $r = 1, \dots, R$:

$$\begin{aligned}
P(\tau_r^u | \mathbf{u}_r, a_u, b_u) &\propto \prod_{i=1}^N P(u_{ir} | \tau_r^u) \times P(\tau_r^u | a_u, b_u) \\
&\propto \prod_{i=1}^N |\tau_r^u c_d|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2} \left(u'_{ir}(\cdot) (\tau_r^u I_T)^{-1} u_{ir}(\cdot) \right)\right\} \times (\tau_r^u)^{-a_u-1} \exp\left\{-\frac{1}{\tau_r^u} b_u\right\} \\
&\propto (\tau_r^u)^{-\frac{T \cdot N}{2} - a_u - 1} \exp\left\{-\frac{1}{\tau_r^u} \left(\frac{1}{2} \sum_{i=1}^N u'_{ir}(\cdot) u_{ir}(\cdot) + b_u \right)\right\} \\
&\sim \text{IG}\left(\frac{T \cdot N}{2} + a_u, \quad \frac{1}{2} \sum_{i=1}^N u'_{ir}(\cdot) u_{ir}(\cdot) + b_u\right)
\end{aligned}$$

2.3 Fixed and random effects

1. $\beta_p(\cdot)$, for $p = 1, \dots, P$:

$$\begin{aligned}
P(\beta_p(\cdot) | \mathbf{Y}, \mathbf{X}, \kappa_p^\beta, \tau_p^\beta) &\propto P(\mathbf{Y} | \mathbf{X}, \beta_p(\cdot), \boldsymbol{\beta}_{[-p]}, \boldsymbol{\theta}, \mathbf{d}, \mathbf{u}, \sigma_e^2) \times P(\beta_p(\cdot) | \kappa_p^\beta, \tau_p^\beta) \\
&\propto \prod_{i>j} \exp\left\{-\frac{1}{2\sigma_e^2} \|E_{ij}^{-p}(\cdot) - X_{ij}^p(\cdot) \beta_p(\cdot)\|^2\right\} \times \exp\left\{-\frac{1}{2} \left(\beta'_p(\cdot) (\tau_p^\beta c_\beta)^{-1} \beta_p(\cdot) \right)\right\} \\
&\quad \text{by letting } E_{ij}^{-p}(t) = y_{ij}(t) - \left(\sum_{p \neq p} X_{ij}^p(t) \beta_p(t) + \theta_i(t) + \theta_j(t) + u_i(t)^T D(t) u_j(t) \right) \\
&\propto \exp\left\{-\frac{1}{2\sigma_e^2} \left(\sum_{i>j} -2(E_{ij}^{-p}(\cdot) X_{ij}^p(\cdot))' \beta_p(\cdot) + \beta'_p(\cdot) (\text{diag}(\sum_{i>j} X_{ij}^{p \ 2}(\cdot))) \beta_p(\cdot) \right)\right\} \times \exp\left\{-\frac{1}{2} \left(\beta'_p(\cdot) (\tau_p^\beta c_\beta)^{-1} \beta_p(\cdot) \right)\right\} \\
&\propto \exp\left\{-\frac{1}{2} \left(\beta'_p(\cdot) \left((\tau_p^\beta c_\beta)^{-1} + \frac{\text{diag}(\sum_{i>j} X_{ij}^{p \ 2}(\cdot))}{\sigma_e^2} \right) \beta_p(\cdot) - \frac{2}{\sigma_e^2} \left(\sum_{i>j} (E_{ij}^{-p}(\cdot) X_{ij}^p(\cdot))' \beta_p(\cdot) \right) \right)\right\} \\
&\sim \text{MVN}_T\left(\left((\tau_p^\beta c_\beta)^{-1} + \frac{\text{diag}(\sum_{i>j} X_{ij}^{p \ 2}(\cdot))}{\sigma_e^2} \right)^{-1} \frac{\sum_{i>j} (E_{ij}^{-p}(\cdot) X_{ij}^p(\cdot))}{\sigma_e^2}, \left((\tau_p^\beta c_\beta)^{-1} + \frac{\text{diag}(\sum_{i>j} X_{ij}^{p \ 2}(\cdot))}{\sigma_e^2} \right)^{-1} \right)
\end{aligned}$$

2. $\theta_i(\cdot)$, for $i = 1, \dots, N$:

$$\begin{aligned}
p(\theta_i(\cdot)|\mathbf{Y}, \kappa^\theta, \tau^\theta) &\propto \prod_{i=i, j \neq i} p(\mathbf{Y}|\mathbf{X}, \theta_i(\cdot), \boldsymbol{\beta}, \boldsymbol{\theta}_{[-i]}, \mathbf{d}, \mathbf{u}, \sigma_e^2) \times p(\theta_i(\cdot)|\kappa^\theta, \tau^\theta) \\
&\propto \prod_{i=i, j \neq i} \exp\left\{-\frac{1}{2\sigma_e^2} \|E_{ij}^{-i}(\cdot) - \theta_i(\cdot)\|^2\right\} \times \exp\left\{-\frac{1}{2} \left(\theta_i'(\cdot)(\tau^\theta c_\theta)^{-1} \theta_i(\cdot)\right)\right\} \\
&\quad \text{by letting } E_{ij}^{-i}(t) = y_{ij}(t) - \left(\sum_{p=1}^P X_{ij}^p(t) \beta_p(t) + \theta_j(t) + u_i(t)^T D(t) u_j(t)\right) \\
&\propto \exp\left\{-\frac{1}{2\sigma_e^2} \left(\sum_{i=i, j \neq i} -2(E_{ij}^{-i}(\cdot))' \theta_i(\cdot) + \theta_i'(\cdot) \left(\sum_{i=i, j \neq i} I_T \theta_i(\cdot)\right)\right)\right\} \times \exp\left\{-\frac{1}{2} \left(\theta_i'(\cdot)(\tau^\theta c_\theta)^{-1} \theta_i(\cdot)\right)\right\} \\
&\propto \exp\left\{-\frac{1}{2} \left(\theta_i'(\cdot) \left((\tau^\theta c_\theta)^{-1} + \frac{(N-1)I_T}{\sigma_e^2}\right) \theta_i(\cdot) - \frac{2}{\sigma_e^2} \left(\sum_{i=i, j \neq i} (E_{ij}^{-i}(\cdot))' \theta_i(\cdot)\right)\right)\right\} \\
&\sim \text{MVN}_T\left(\left((\tau^\theta c_\theta)^{-1} + \frac{(N-1)I_T}{\sigma_e^2}\right)^{-1} \frac{\sum_{i=i, j \neq i} E_{ij}^{-i}(\cdot)}{\sigma_e^2}, \left((\tau^\theta c_\theta)^{-1} + \frac{(N-1)I_T}{\sigma_e^2}\right)^{-1}\right)
\end{aligned}$$

3. $d_r(\cdot)$, for $r = 1, \dots, R$:

$$\begin{aligned}
P(d_r(\cdot)|\mathbf{Y}, \kappa^d) &\propto P(\mathbf{Y}|\mathbf{X}, d_r(\cdot), \boldsymbol{\beta}, \boldsymbol{\theta}, \mathbf{d}_{[-r]}, \mathbf{u}, \sigma_e^2) \times P(d_r(\cdot)|\kappa^d) \\
&\propto \prod_{i>j} \exp\left\{-\frac{1}{2\sigma_e^2} \|E_{ij}^{-r}(\cdot) - d_r(\cdot) u_{ir}(\cdot) u_{jr}(\cdot)\|^2\right\} \times \exp\left\{-\frac{1}{2} \left(d_r'(\cdot) c_d^{-1} d_r(\cdot)\right)\right\} \\
&\quad \text{by letting } E_{ij}^{-r}(t) = y_{ij}(t) - \left(\sum_{p=1}^P X_{ij}^p(t) \beta_p(t) + \theta_i(t) + \theta_j(t) + u_{i[-r]}(t)^T D_{[-r, -r]}(t) u_{j[-r]}(t)\right) \\
&\propto \exp\left\{-\frac{1}{2\sigma_e^2} \left(\sum_{i>j} -2(E_{ij}^{-r}(\cdot) u_{ir}(\cdot) u_{jr}(\cdot))' d_r(\cdot) + d_r'(\cdot) \left(\text{diag}\left(\sum_{i>j} (u_{ir}(\cdot) u_{jr}(\cdot))^2\right) d_r(\cdot)\right)\right)\right\} \\
&\quad \times \exp\left\{-\frac{1}{2} \left(d_r'(\cdot) c_d^{-1} d_r(\cdot)\right)\right\} \\
&\propto \exp\left\{-\frac{1}{2} \left(d_r'(\cdot) \left(c_d^{-1} + \frac{\text{diag}(\sum_{i>j} (u_{ir}(\cdot) u_{jr}(\cdot))^2)}{\sigma_e^2}\right) d_r(\cdot) - \frac{2}{\sigma_e^2} \left(\sum_{i>j} (E_{ij}^{-r}(\cdot) u_{ir}(\cdot) u_{jr}(\cdot))' d_r(\cdot)\right)\right)\right\} \\
&\sim \text{MVN}_T\left(\left(c_d^{-1} + \frac{\text{diag}(\sum_{i>j} (u_{ir}(\cdot) u_{jr}(\cdot))^2)}{\sigma_e^2}\right)^{-1} \frac{\sum_{i>j} (E_{ij}^{-r}(\cdot) u_{ir}(\cdot) u_{jr}(\cdot))}{\sigma_e^2}, \left(c_d^{-1} + \frac{\text{diag}(\sum_{i>j} (u_{ir}(\cdot) u_{jr}(\cdot))^2)}{\sigma_e^2}\right)^{-1}\right)
\end{aligned}$$

4. $u_i(t)$, for $i = 1, \dots, N$; (while other variables sample T -length vectors, $u_i(t)$ is R -length vector):

$$\begin{aligned}
P(u_i(t)|Y, \tau^u) &\propto P(Y(t)|u_i(t), \beta, \theta, d, u_{[-t]}, \sigma_e^2) \times P(u_i(t)|\tau^u) \\
&\propto \prod_{i=i, j \neq i} \exp\left\{-\frac{1}{2\sigma_e^2} \|E_{ij}^{-u}(t) - u_j(t)^T D(t) u_i(t)\|^2\right\} \times \exp\left\{-\frac{1}{2} \left(u_i'(t) \text{diag}(\tau^u) u_i(t)\right)\right\} \\
&\quad \text{by letting } E_{ij}^{-u}(t) = y_{ij}(t) - \left(\sum_{p=1}^P X_{ij}^p(t) \beta_p(t) + \theta_i(t) + \theta_j(t)\right) \text{ and } \text{diag}(\tau^u) = \text{diag}(\tau_1^u, \dots, \tau_r^u) \\
&\propto \exp\left\{-\frac{1}{2\sigma_e^2} \left(\sum_{i=i, j \neq i} -2(E_{ij}^{-u}(t) u_j'(t) D(t)) u_i(t) + u_i'(t) \left(\sum_{j \neq i} D(t) u_j(t) u_j'(t) D(t)\right) u_i(t)\right)\right\} \\
&\quad \times \exp\left\{-\frac{1}{2} \left(u_i'(t) \text{diag}(\tau^u) u_i(t)\right)\right\} \\
&\propto \exp\left\{-\frac{1}{2} \left(u_i'(t) \left(\text{diag}(\tau^u)^{-1} + \frac{\sum_{j \neq i} D(t) u_j(t) u_j'(t) D(t)}{\sigma_e^2}\right) u_i(t) - \frac{2}{\sigma_e^2} \left(\sum_{i=i, j \neq i} (E_{ij}^{-u}(t) u_j'(t) D(t)) u_i(t)\right)\right)\right\} \\
&\sim \text{MVN}_R\left(\left(\text{diag}(\tau^u)^{-1} + \frac{\sum_{j \neq i} D(t) u_j(t) u_j'(t) D(t)}{\sigma_e^2}\right)^{-1} \frac{\sum_{i=i, j \neq i} (E_{ij}^{-u}(t) u_j'(t) D(t))'}{\sigma_e^2}, \right. \\
&\quad \left. \left(\text{diag}(\tau^u)^{-1} + \frac{\sum_{j \neq i} D(t) u_j(t) u_j'(t) D(t)}{\sigma_e^2}\right)^{-1}\right)
\end{aligned}$$

2.4 Metropolis-Hastings algorithm for Gaussian process parameters

For Gaussian process parameters, we use the Metropolis-Hastings algorithm with a proposal density Q being the multivariate Gaussian distribution, with a diagonal covariance matrix multiplied by σ_Q^2 (proposal distribution variance parameters set by the user), centered on the current values of (τ, κ) . Since both τ and κ have to be positive, we use

$$(\tau', \kappa') \sim \exp(MVN_2(\log(\tau, \kappa), \sigma_Q^2 I_2)).$$

Under the symmetric proposal distribution as above, we cancel out Q-ratio and then accept the new proposed value (τ', κ') with probability equal to:

$$\text{Acceptance Probability} = \begin{cases} \frac{P(\tau^{x'}, \kappa^{x'} | x(\cdot), a_x, b_x)}{P(\tau^x, \kappa^x | x(\cdot), a_x, b_x)} & \text{if } < 1 \\ 1 & \text{else} \end{cases}, \quad (5)$$

where x is the variable of interest (e.g. β_p, θ, d) and $x(\cdot)$ is the T -length vector. Then the log of acceptance ratio we have is:

$$\log(\text{Acceptance Probability}) = \min \left(\log \left(\frac{P(\tau^{x'}, \kappa^{x'} | x(\cdot), a_x, b_x)}{P(\tau^x, \kappa^x | x(\cdot), a_x, b_x)} \right), 0 \right). \quad (6)$$

Use the log of acceptance ratio, if the log of a sample from Uniform(0,1) is less than the log-acceptance probability (6), we accept the proposed value. Otherwise, we reject. Below are the derivation of acceptance ratio for each of the variables.

1. $(\tau_p^\beta, \kappa_p^\beta)$, for $p = 1, \dots, P$:

$$\begin{aligned} \frac{P(\tau_p^{\beta'}, \kappa_p^{\beta'} | \beta_p(\cdot), a_\beta, b_\beta)}{P(\tau_p^\beta, \kappa_p^\beta | \beta_p(\cdot), a_\beta, b_\beta)} &= \frac{P(\tau_p^{\beta'}, \kappa_p^{\beta'}, \beta_p(\cdot) | a_\beta, b_\beta)}{P(\tau_p^\beta, \kappa_p^\beta, \beta_p(\cdot) | a_\beta, b_\beta)} \\ &= \frac{P(\tau_p^{\beta'} | a_\beta, b_\beta) P(\kappa_p^{\beta'}) P(\beta_p(\cdot) | \tau_p^{\beta'}, \kappa_p^{\beta'})}{P(\tau_p^\beta | a_\beta, b_\beta) P(\kappa_p^\beta) P(\beta_p(\cdot) | \tau_p^\beta, \kappa_p^\beta)} \\ &= \frac{\text{dinvgamma}(\tau_p^{\beta'}, a_\beta, b_\beta) \times \text{dmvnorm}(\beta_p(\cdot), \mathbf{0}, \tau_p^{\beta'} c(\kappa_p^{\beta'}))}{\text{dinvgamma}(\tau_p^\beta, a_\beta, b_\beta) \times \text{dmvnorm}(\beta_p(\cdot), \mathbf{0}, \tau_p^\beta c(\kappa_p^\beta))} \end{aligned} \quad (7)$$

2. $(\tau^\theta, \kappa^\theta)$

$$\begin{aligned} \frac{P(\tau^{\theta'}, \kappa^{\theta'} | \boldsymbol{\theta}(\cdot), a_\theta, b_\theta)}{P(\tau^\theta, \kappa^\theta | \boldsymbol{\theta}(\cdot), a_\theta, b_\theta)} &= \frac{P(\tau^{\theta'}, \kappa^{\theta'}, \boldsymbol{\theta}(\cdot) | a_\theta, b_\theta)}{P(\tau^\theta, \kappa^\theta, \boldsymbol{\theta}(\cdot) | a_\theta, b_\theta)} \\ &= \frac{P(\tau^{\theta'} | a_\theta, b_\theta) P(\kappa^{\theta'}) \prod_{i=1}^N P(\theta_i(\cdot) | \tau^{\theta'}, \kappa^{\theta'})}{P(\tau^\theta | a_\theta, b_\theta) P(\kappa^\theta) P(\prod_{i=1}^N \theta_i(\cdot) | \tau^\theta, \kappa^\theta)} \\ &= \frac{\text{dinvgamma}(\tau^{\theta'}, a_\theta, b_\theta) \times \prod_{i=1}^N \text{dmvnorm}(\theta_i(\cdot), \mathbf{0}, \tau^{\theta'} c(\kappa^{\theta'}))}{\text{dinvgamma}(\tau^\theta, a_\theta, b_\theta) \times \prod_{i=1}^N \text{dmvnorm}(\theta_i(\cdot), \mathbf{0}, \tau^\theta c(\kappa^\theta))} \end{aligned} \quad (8)$$

3. κ^d

$$\begin{aligned} \frac{P(\kappa^{d'} | \mathbf{d}(\cdot))}{P(\kappa^d | \mathbf{d}(\cdot))} &= \frac{P(\kappa^{d'}, \mathbf{d}(\cdot))}{P(\kappa^d, \mathbf{d}(\cdot))} \\ &= \frac{P(\kappa^{d'}) \prod_{r=1}^R P(d_r(\cdot) | \kappa^{d'})}{P(\kappa^d) \prod_{r=1}^R P(d_r(\cdot) | \kappa^d)} \\ &= \frac{\prod_{r=1}^R \text{dmvnorm}(d_r(\cdot), \mathbf{0}, c(\kappa^{d'}))}{\prod_{r=1}^R \text{dmvnorm}(d_r(\cdot), \mathbf{0}, c(\kappa^d))} \end{aligned} \quad (9)$$