In Gibbs sampling algorithm, the samples are generated from full conditional posteriors of each variable. The full conditional probabilities of variables U ,C, L are represented as $p(U|X, C, L)$, $p(C|X, U, L)$, $p(L|X, U, C)$ which are proportional to joint distribution $p(X, U, C, L)$. In each of these probabilities, the intended variable is conditioned on fixed values of other variables and the observation set X. Sampling from these distributions is a hard task in problem. In Metropolis-Hastings algorithm it is not necessary to generate samples from full conditional distributions. Instead a proposal distribution is chosen based on the current value of random variables in a special application. In this case we can use a combination of Gibbs sampling and Metropolis-Hastings algorithms which is called Metropolis-within-Gibbs algorithm. This algorithm which performs the MAP inference to reach the best values of random variables,is represented in algorithm (1).

The proposed distribution for sampling membership values is a Dirichlet distribution using α parameter.

$$u_n^t \sim Dirichlet(\alpha = 1_S) \tag{1}$$

As mentioned before if we set $\alpha = 1$ the Dirichlet distribution will be symmetric. So the ratio of the proposal density values is equal to 1 and the sampled membership value in time t for data point x_n is evaluated by equation (**??**) and is accepted with probability a_u which is based on the ratio of target density for proposed and current value.

$$\begin{aligned}
&p(x_n, u_n|C, L) = p(x_n|u_n, C, L) \times p(u_n|\alpha) \\
&\propto \prod_{s=1}^{S} \exp\{-\frac{1}{2}(x_n - c_s)^T L^T u_{ns}^m L(x_n - c_s)\} \\
&\times \prod_{i=1}^{N} \exp\{-\frac{1}{2}\left(\frac{(u_n.u_i)}{||u_n||||u_i||} - 0.5\right)(x_n - x_i)^T L^T L(x_n - x_i)\} \times \prod_{s=1}^{S} u_{ns}^{\alpha_s - 1}
\end{aligned} \tag{2}$$

$$a_u = \min\{1, \frac{p(x_n, u_n^t|C, L)}{p(x_n, u_n|C, L)}\} \tag{3}$$

The proposal distribution for sampling cluster centers is a gaussian distribution which its mean is the current markov chain state and

$$c_s^t \sim N(c_s, \frac{1}{\delta}\Sigma_x) \tag{4}$$

$$\begin{aligned}
&\tilde{p}(X, c_s|U, L) = \tilde{p}(X|U, L, c_s) \times p(c_s|\mu_x, \Sigma_x) \\
&\propto \exp\{-\frac{1}{2}\sum_{n=1}^{N}(x_n - c_s)^T L^T u_{ns}^m L(x_n - c_s)\} \\
&\times exp\{-\frac{1}{2}(c_s - \mu_x)^T \Sigma_x^{-1}(c_s - \mu_x)\}
\end{aligned} \tag{5}$$

$$a_c = \min\{1, \frac{p(X, c_s^t|U, L)}{p(X, c_s|U, L)}\} \tag{6}$$

$$L_k^t \sim N(0, diag^{-1}(\nu)) \tag{7}$$

$$\begin{aligned}
&p(X, L|C, U) = p(X|C, U, L) \times p(L|\nu) \\
&\propto \exp\{-\frac{1}{2}\sum_{n=1}^{N}\sum_{s=1}^{S}(x_n - c_s)^T L^T u_{ns}^m L(x_n - c_s)\} \\
&\times \exp\{-\frac{1}{2}\sum_{n=1}^{N}\sum_{i=1}^{N}\left(\frac{(u_n.u_i)}{||u_n||||u_i||} - 0.5\right)(x_n - x_i)^T L^T L(x_n - x_i)\} \\
&\times exp\{-\frac{1}{2}\sum_{k=1}^{p}(L_k)^T diag(\nu)(L_k)\}
\end{aligned} \tag{8}$$

$$a_L = \min\{1, \frac{p(X, L^t|C, U)}{p(X, L|C, U)}\} \tag{9}$$

Algorithm 1 Inference algorithm for GBML model

Input: Data matrix X, fuzzifier$m$, Dirichlet parameterα, number of clusters S; Gamma hyperparameters a and b, number of sampling iterations N_{iter};

Output: MAP estimates for membership values U^*, cluster centers C^* and transformation matrix L^*;

initialize hyperparameters a^*,b^*,α and set μ_x, Σ_x according to (**??**) and (**??**);
sample initial $u_n \sim Dirichlet(\alpha = 1_S)$ for all $n = \{1, ..., N\}$;
sample initial $c_s \sim N(\mu_x, \Sigma_x)$ for all $s = \{1, ..., S\}$;
sample initial $L = I_p$;
set MAP samples to current states $u_n^* \leftarrow u_n$,$c_s^* \leftarrow c_s$,$L_k^* \leftarrow L_k$ for all $n = \{1, ..., N\}$, $s = \{1, ..., S\}$, $k = \{1, ..., p\}$
for $iter = 1...N_{iter}$ **do**
    /* Sample $U \sim p(U|X, C, L) \propto p(X, C, U, L)$ */
    for $n = \{1, ..., N\}$ **do**
        sample proposed new membership vector u_n^t from ()
        accept $(u_n \leftarrow u_n^t)$ proposal with probability a_u from ()
        if $p(x_n, u_n^t|C^*, L^*) > p(x_n, u_n^*|C^*, L^*)$ using () **then**
            $u_n^* \leftarrow u_n^t$
    /* Sample $C \sim p(C|X, U, L) \propto p(X, C, U, L)$ */
    for $s = \{1, ..., S\}$ **do**
        sample proposed new cluster center c_s^t from ()
        accept proposal $(c_s \leftarrow c_s^t)$ with probability a_c from ()
        if $p(X, c_s^t|U^*, L^*) > p(X, c_s^*|U^*, L^*)$ using () **then**
            $c_s^* \leftarrow c_s^t$
    /* Sample $L \sim p(L|X, C, U) \propto p(X, C, U, L)$ */
    for $k = \{1, ..., p\}$ **do**
        sample proposed new cluster center L_k^t from ()
        for $q = \{1, ..., p\}$ **do**
            if $q == k$ **then** $L_q^t = L_k^t$
        accept proposal $(L_k \leftarrow L_k^t)$ with probability a_L from ()
        if $p(X, L^t|C^*, U^*) > p(X, L^*|C^*, U^*)$ using () **then**
            $L_k^* \leftarrow L_k^t$
/* check full sample for new maximum likelihood */
if $p(X, C, U, L) > p(X, C^*, U^*, L^*)$ using () **then**
    $U^* \leftarrow U$
    $C^* \leftarrow C$