Derivation of EM algorithm

The complete log-likelihood including missing data \( \{z_i\} \) for the proposed model is

\[
\sum_{i=1}^{I} \sum_{j=1}^{J} \sum_{k=1}^{K} I(z_{i,j} = k) \left( \sum_{l=1}^{L} \log f_{k,l,x_{i,j,l}} + \log q_{i,k} \right).
\]

Here, we introduce the variable for conditional probability for \( z_{i,j} \) given the parameters and the mutation features \( x_{i,j} \), \( \theta_{i,k,m} = \Pr(z_{i,j} = k \mid x_{i,j} = m, \{f_{k,l}\}, \{q_i\}) \).

Note that this conditional probability just depends on the value of mutation feature \( m = (m_1, \ldots, m_L) \), not on the index \( j \). Then, the expected complete log-likelihood augmented by Lagrange multipliers is calculated as

\[
\sum_{i=1}^{I} \sum_{m} g_i,m \sum_{k=1}^{K} \theta_{i,k,m} \left( \sum_{l=1}^{L} \log f_{k,l,m_j} + \log q_{i,k} \right) + \sum_{k=1}^{K} \sum_{l=1}^{L} \tau_{k,l} (1 - \sum_{p=1}^{M_l} f_{k,l,p}) + \sum_{i=1}^{I} \rho_i (1 - \sum_{k=1}^{K} q_{i,k}).
\]

Differentiating it leads to the following stationary equations:

\[
\sum_{i=1}^{I} \sum_{m:m_i=p} g_i,m \theta_{i,k,m} - \tau_{k,l} f_{k,l,p} = 0, \quad (p = 1, \cdots, M_i, k = 1, \cdots, K, l = 1, \cdots, L),
\]

\[
\sum_{m} g_i,m \theta_{i,k,m} - \rho_i q_{i,k} = 0, \quad (k = 1, \cdots, K, i = 1, \cdots, I).
\]

Then, by eliminating Lagrange multipliers, updating rules can be obtained.