Relationship with nonnegative matrix factorization

First, for ease of explanation, let assume that the full representation" representation \((L = 1)\) is used. Suppose that each \(m\) has unique appropriate index from 1 to \(|M| = \prod_{l=1}^{L} M_l\) (the number of possible mutation patterns), so that \(m\) can be indices of matrices.

Let \(G = \{g_{i,m}\}\) denote the \(I \times |M|\) matrix, where \(g_{i,m}\) is the number of mutations whose mutation patterns are equal to \(m\) in the \(i\)-th cancer genome. Nonnegative matrix factorization aims to find low rank decomposition, \(G \sim \tilde{Q}F\), where \(\tilde{Q} = \{\tilde{q}_{i,k}\}\) and \(F = \{f_{k,m}\}\) are nonnegative matrix, and row vectors of \(F\) are often restricted to be sum to one. We used the notation \(\tilde{Q}\) instead of \(Q\) to represent that the row vectors of \(\tilde{Q}\) are not normalized to sum to one in general.

For solving NMF, the previous study (Lee et al. 2000) used the following updating rule:

\[
\begin{align*}
f_{k,m} &\leftarrow f_{k,m} \frac{(\tilde{Q}^T G)_{k,m}}{(\tilde{Q}^T \tilde{Q})_{k,m}}, \\
\tilde{q}_{i,k} &\leftarrow \tilde{q}_{i,k} \frac{(GF^T)_{i,k}}{(QFF)^T}_{i,k},
\end{align*}
\]

that reduces the Euclidean distance \(||G - \tilde{Q}F||\). Therefore, the optimization problem for the existing approach is

\[
\begin{align*}
\text{minimize} & \quad ||G - \tilde{Q}F|| \\
\text{subject to} & \quad \sum_m f_{k,m} = 1, \ k = 1, \ldots, K \\
 & \quad f_{k,m} \geq 0, \ k = 1, \ldots, K, \ m \in M \\
 & \quad \tilde{q}_{i,k} \geq 0, \ i = 1, \ldots, I, \ k = 1, \ldots, K.
\end{align*}
\]

(1)

On the other hand, there is another type of updating rule:

\[
\begin{align*}
f_{k,m} &\leftarrow f_{k,m} \frac{\sum_i \tilde{q}_{i,k} g_{i,m}/(\tilde{Q}F)_{i,m}}{\sum_i \tilde{q}_{i,k}}, \\
\tilde{q}_{i,k} &\leftarrow \tilde{q}_{i,k} \frac{\sum_m f_{k,m} g_{i,m}/(\tilde{Q}F)_{i,m}}{\sum_m f_{k,m}},
\end{align*}
\]

that reduces the Kullback-Liebler Divergence:

\[
KL(G\|\tilde{Q}F) = \sum_{i,m} \left( g_{i,m} \log \frac{g_{i,m}}{(\tilde{Q}F)_{i,m}} - g_{i,m} + (\tilde{Q}F)_{i,m} \right)
\]

In general cases including the independent representation, there is restrictions \(f_{k,m} = \prod_l f_{k,l,m_l}\) by smaller set of parameters. Let us consider the following optimization problem with the Kullback-Liebler Divergence and the restrictions on \(F\):

\[
\begin{align*}
\text{minimize} & \quad KL(G\|\tilde{Q}F) \\
\text{subject to} & \quad f_{k,m} = \prod_l f_{k,l,m_l}, \ k = 1, \ldots, K, \ m \in M \\
 & \quad f_{k,l,m} \geq 0, \ k = 1, \ldots, K, \ m \in M \\
 & \quad \tilde{q}_{i,k} \geq 0, \ i = 1, \ldots, I, \ k = 1, \ldots, K.
\end{align*}
\]

(2)
In fact, this is equivalent to the proposed method, whose optimization problem can be written as:

\[
\text{maximize} \quad L(Q, F|G) = \sum_{i,m} g_{i,m} \log(QF)_{i,m}
\]

subject to

\[

t_{k,m} = \prod_l f_{k,l,m}, \quad k = 1, \ldots, K, \quad m \in M
\]

\[
f_{k,l,m} \geq 0, \quad k = 1, \ldots, K, \quad m \in M
\]

\[
\sum_i q_{i,k} = 1, \quad i = 1, \ldots, I
\]

\[
q_{i,k} \geq 0, \quad i = 1, \ldots, I, \quad k = 1, \ldots, K.
\]

(3)

**Proposition 1** When \((Q, F) = (Q^*, F^*)\) is an optimal solution of the optimization problem (3), then \((\hat{Q}, \hat{F}) = (R^*Q^*, F^*)\) is an optimal solution of the optimization problem (2). On the other hand, when \((\hat{Q}, \hat{F}) = (\hat{Q}^*, F^*)\) is an optimal solution of the optimization problem (2), then \((Q, F) = (R^{*-1}\hat{Q}^*, F^*)\) is an optimal solution of the optimization problem (3), where 

\[
R^* = \text{diag}(r_1^*, \ldots, r_I^*), \quad r_i^* = \sum_m g_{i,m}, \quad i = 1, \ldots, I.
\]

**Proof.** This is because

\[
KL(G||\hat{Q}F) = -\sum_i \left( (\sum_m g_{i,m}) \log \hat{r}_i - \hat{r}_i \right) - L(Q, F|G) + \text{(constant value)},
\]

where \(Q\) is row-normalized matrix for \(\hat{Q}\), \(\hat{r}_i = \sum_k q_{i,k}\) for each \(i\), and \((\sum_m g_{i,m}) \log \hat{r}_i - \hat{r}_i\) takes its maximum at \(\hat{r}_i = r_i^*\). \(\square\)