

## Relationship with nonnegative matrix factorization

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

Let  $G = \{g_{i,\mathbf{m}}\}$  denote the  $I \times |\mathbf{M}|$  matrix, where  $g_{i,\mathbf{m}}$  is the number of mutations whose mutation patters are equal to  $\mathbf{m}$  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  $F = \{f_{k,\mathbf{m}}\}$  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:

$$f_{k,\mathbf{m}} \leftarrow f_{k,\mathbf{m}} \frac{(\tilde{Q}^T G)_{k,\mathbf{m}}}{(\tilde{Q}^T \tilde{Q} F)_{k,\mathbf{m}}}, \quad \tilde{q}_{i,k} \leftarrow \tilde{q}_{i,k} \frac{(GF^T)_{i,k}}{(\tilde{Q} F F^T)_{i,k}},$$

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

$$\begin{aligned} & \text{minimize} \quad \|G - \tilde{Q}F\| \\ & \text{subject to} \quad \sum_{\mathbf{m}} f_{k,\mathbf{m}} = 1, \quad k = 1, \dots, K \\ & \quad \quad \quad f_{k,\mathbf{m}} \geq 0, \quad k = 1, \dots, K, \quad \mathbf{m} \in M \\ & \quad \quad \quad \tilde{q}_{i,k} \geq 0, \quad i = 1, \dots, I, \quad k = 1, \dots, K. \end{aligned} \tag{1}$$

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

$$\begin{aligned} f_{k,\mathbf{m}} & \leftarrow f_{k,\mathbf{m}} \frac{\sum_i \tilde{q}_{i,k} g_{i,\mathbf{m}} / (\tilde{Q}F)_{i,\mathbf{m}}}{\sum_i \tilde{q}_{i,k}}, \\ \tilde{q}_{i,k} & \leftarrow \tilde{q}_{i,k} \frac{\sum_{\mathbf{m}} f_{k,\mathbf{m}} g_{i,\mathbf{m}} / (\tilde{Q}F)_{i,\mathbf{m}}}{\sum_{\mathbf{m}} f_{k,\mathbf{m}}}. \end{aligned}$$

that reduces the Kullback-Liebler Divergence:

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

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

$$\begin{aligned} & \text{minimize} \quad KL(G||\tilde{Q}F) \\ & \text{subject to} \quad f_{k,\mathbf{m}} = \prod_l f_{k,l,m_l}, \quad k = 1, \dots, K, \quad \mathbf{m} \in M \\ & \quad \quad \quad f_{k,l,p} \geq 0, \quad k = 1, \dots, K, \quad \mathbf{m} \in M \\ & \quad \quad \quad \tilde{q}_{i,k} \geq 0, \quad i = 1, \dots, I, \quad k = 1, \dots, K. \end{aligned} \tag{2}$$

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

$$\begin{aligned} & \text{maximize} \quad L(Q, F|G) (= \sum_{i,\mathbf{m}} g_{i,\mathbf{m}} \log(QF)_{i,\mathbf{m}}) \\ & \text{subject to} \quad f_{k,\mathbf{m}} = \prod_l f_{k,l,m_l}, \quad k = 1, \dots, K, \quad \mathbf{m} \in M \\ & \quad \quad \quad f_{k,l,p} \geq 0, \quad k = 1, \dots, K, \quad \mathbf{m} \in M \\ & \quad \quad \quad \sum_k q_{i,k} = 1, \quad i = 1, \dots, I \\ & \quad \quad \quad q_{i,k} \geq 0, \quad i = 1, \dots, I, \quad k = 1, \dots, K. \end{aligned} \tag{3}$$

**Proposition 1** When  $(Q, F) = (Q^*, F^*)$  is an optimal solution of the optimization problem (??), then  $(\tilde{Q}, F) = (R^*Q^*, F^*)$  is an optimal solution of the optimization problem (??). On the other hand, when  $(\tilde{Q}, F) = (\tilde{Q}^*, F^*)$  is an optimal solution of the optimization problem (??), then  $(Q, F) = (R^{*-1}\tilde{Q}^*, F^*)$  is an optimal solution of the optimization problem (??), where  $R^* = \text{diag}(r_1^*, \dots, r_I^*)$ ,  $r_i^* = \sum_{\mathbf{m}} g_{i,\mathbf{m}}$ ,  $i = 1, \dots, I$ .

Proof. This is because

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

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