Additional file 1

Mathematical Derivation of the Update Rule for Factor V

We derive the multiplicative update rules for $\mathbf{V}$, as given in Equation **3**, which are guaranteed to decrease the objective function $F(\mathbf{S}, \mathbf{V})$ for $\mathbf{S}$ fixed. Specifically, we adopt the same strategy used in [1, 2] that introduced an auxiliary function in the Expectation-Maximization algorithm. The purpose of the auxiliary function f is to offer an upper approximation of F that coincides with F at the point $\mathbf{V}^h$, which represents the current iterate in the algorithm. Consequently, the matrix

$$\mathbf{V}^{h+1} = \arg\min_{\mathbf{V}} f(\mathbf{V}, \mathbf{V}^h)$$

satisfies the relationship:

$$F(\mathbf{V}^{h+1}) \leq f(\mathbf{V}^{h+1}, \mathbf{V}^h) \leq f(\mathbf{V}^h, \mathbf{V}^h) = F(\mathbf{V}^h).$$

Since the minimum of f can be computed efficiently, this approach ensures a monotonic algorithm for minimizing F:

$$\mathbf{V}^1 \to \mathbf{V}^2 \to \dots,$$

with $F(\mathbf{V}^{h+1}) \leq F(\mathbf{V}^h)$ for all $h \geq 1$.

We denote $F(\mathbf{S}, \mathbf{V}) = F(\mathbf{V})$ for $\mathbf{S}$ fixed and express it as the sum of the following terms:

$$F(\mathbf{V}) = F_1(\mathbf{V}) + F_2(\mathbf{V}) + \lambda_2 F_3(\mathbf{V}) + \lambda_2 F_4(\mathbf{V})$$

where the functions F_i are defined as follows:

- $F_1(\mathbf{V}) = \sum_{i=1}^{n} \sum_{j=1}^{m} -X_{ij} \log(USV^\top)_{ij}$,
- $F_2(\mathbf{V}) = \sum_{b=1}^{r} \sum_{j=1}^{m} V_{jb} \big(\sum_{i=1}^{n} (US)_{ib} + \lambda_1 \big)$,
- $F_3(\mathbf{V}) = tr(\mathbf{V}\mathbf{V}^\top \mathbf{V}\mathbf{V}^\top)$,
- $F_4(\mathbf{V}) = -tr(\mathbf{V}\mathbf{V}^\top)$.

In fact, we have

$$\begin{aligned}
\lambda_2 \|\mathbf{V}^\top \mathbf{V} - \mathbb{I}_r\|_F^2 &= \lambda_2 \operatorname{tr}\left((\mathbf{V}^\top \mathbf{V} - \mathbb{I}_r)^\top (\mathbf{V}^\top \mathbf{V} - \mathbb{I}_r) \right) \\
&= \lambda_2 \operatorname{tr}\left((\mathbf{V}^\top \mathbf{V}\mathbf{V}^\top \mathbf{V} - 2\mathbf{V}^\top \mathbf{V} + \mathbb{I}_r) \right) \\
&= \lambda_2 \operatorname{tr}(\mathbf{V}^\top \mathbf{V}\mathbf{V}^\top \mathbf{V}) - 2\lambda_2 \operatorname{tr}(\mathbf{V}^\top \mathbf{V}) + \lambda_2 r \\
&= \lambda_2 \operatorname{tr}(\mathbf{V}\mathbf{V}^\top \mathbf{V}\mathbf{V}^\top) - 2\lambda_2 \operatorname{tr}(\mathbf{V}\mathbf{V}^\top) + \lambda_2 r,
\end{aligned}$$

where, in the last step, we utilize the property that the trace of a product of matrices is invariant under cyclic permutations of the factors.

To derive the update in Equation **3** , for each term of F, we construct an auxiliary function. The sum of these auxiliary functions serves as an auxiliary function for the original objective function, thereby enabling the derivation of the multiplicative update. For simplicity, we use the notation V_{jb} instead of $V_{bj}^{\top}$ to denote the transposed elements. This notation is adopted to improve readability.

- The function

$$f_1(\mathbf{V}, \mathbf{V}^h) = -\sum_{i=1}^{n}\sum_{b=1}^{r}\sum_{j=1}^{m} X_{ij} \frac{(US)_{ib} V_{jb}^h}{\sum_{a=1}^{r}(US)_{ia} V_{ja}^h} \times \left(\log \big((US)_{ib} V_{jb} \big) - \log \left(\frac{(US)_{ib} V_{jb}^h}{\sum_{a=1}^{r}(US)_{ia} V_{ja}^h} \right) \right)$$

  is an auxiliary function for $F_1(\mathbf{V})$.
- The function

$$f_2(\mathbf{V}, \mathbf{V}^h) = \sum_{b=1}^{r}\sum_{j=1}^{m} \frac{l_b}{2V_{jb}^h} V_{jb}^2 + \frac{l_b}{2} V_{jb}^h$$

  where $l_b = \sum_{i=1}^{n}(US)_{ib} + \lambda_1 \geq 0$,is an auxiliary function for $F_2(\mathbf{V})$.
- The function

$$f_3(\mathbf{V}, \mathbf{V}^h) = \sum_{b=1}^{r}\sum_{j=1}^{m} \frac{(V^h (V^h)^{\top} V^h)_{jb}}{(V^h)_{jb}^3} V_{jb}^4$$

  is an auxiliary function for $F_3(\mathbf{V})$.
- The function

$$f_4(\mathbf{V}, \mathbf{V}^h) = \sum_{b=1}^{r}\sum_{j=1}^{m} \frac{V_{jb}^4}{(V_{jb}^h)^2} - 3V_{jb}^2 + (V_{jb}^h)^2$$

  is an auxiliary function for $F_4(\mathbf{V})$.

Based on these derivations, we define the function

$$f(\mathbf{V}, \mathbf{V}^h) = f_1(\mathbf{V}, \mathbf{V}^h) + f_2(\mathbf{V}, \mathbf{V}^h) + \lambda_2 (f_3(\mathbf{V}, \mathbf{V}^h) + f_4(\mathbf{V}, \mathbf{V}^h))$$

which acts as an auxiliary function for $F(\mathbf{V})$.

The final step is to establish that the update in Equation **3** achieves the minimum of f.

The minimum of $f(\mathbf{V}, \mathbf{V}^h)$ with respect to $\mathbf{V}$ is determined by applying the Karush-Kuhn-Tucker conditions, which require setting the gradient to zero:

$$\frac{\partial f(\mathbf{V}, \mathbf{V}^h)}{\partial V_{jb}} = -\frac{\sum_i X_{ij}(US)_{ib}V^h_{jb}}{\sum_a (US)_{ia}V^h_{ja}} \frac{1}{V_{jb}} + \left(\sum_i (US)_{ib} + \lambda_1\right) \frac{V_{jb}}{V^h_{jb}}$$
$$+ 4\lambda_2 \frac{\left(V^h (V^h)^\top V^h\right)_{jb}}{(V^h_{jb})^3} V^3_{jb} + \lambda_2 \left(\frac{4V^3_{jb}}{(V^h_{jb})^2} - 6V_{jb}\right) = 0$$

which, for $V_{jb} \neq 0$, can be simplified to

$$-\frac{\sum_i X_{ij}(US)_{ib}V^h_{jb}}{(US(V^h)^\top)_{ij}} + \left(\frac{\sum_i (US)_{ib} + \lambda_1}{V^h_{jb}}\right) V^2_{jb}$$
$$+4\lambda_2 \frac{\left(V^h (V^h)^\top V^h\right)_{jb}}{(V^h_{jb})^3} V^4_{jb} + \lambda_2 \frac{4V^4_{jb}}{(V^h_{jb})^2} - 6\lambda_2 V^2_{jb} = 0.$$

Let A_{jb}, B_{jb}, and C_{jb} be defined as follows:

- $A_{jb} = \frac{4\lambda_2}{(V^h_{jb})^2} \left[\frac{\left(V^h (V^h)^\top V^h\right)_{jb}}{V^h_{jb}} + 1\right]$,
- $B_{jb} = \frac{\sum_i (US)_{ib} + \lambda_1}{V^h_{jb}} - 6\lambda_2$,
- $C_{jb} = -\sum_i \frac{X_{ij}(US)_{ib}V^h_{jb}}{(US(V^h)^\top)_{ij}}$.

The gradient can then be rewritten as:

$$A_{jb}V^4_{jb} + B_{jb}V^2_{jb} + C_{jb} = 0. \qquad \text{(Biquadratic)}$$

The polynomial (Biquadratic) has four roots, but only one non-negative root, given by

$$V^*_{jb} = V_{jb}\sqrt{\frac{-B_{jb} + \sqrt{\Delta_{jb}}}{2A^*_{jb}}},$$

where $\Delta_{jb} = B^2_{jb} - 4A_{jb}C_{jb} \geq B^2_{jb} =$ since $A_{jb} \geq 0$ and $C_{jb} \leq 0$. This demonstrates that the update in attains the minimum of f.

References

[1] Lee, D., Seung, H.: Algorithms for Non-negative Matrix Factorization. *Advances in Neural Information Processing Systems*, **13** (2001).

[2] Esposito, F., Gillis, N., Del Buono, N.: Orthogonal joint sparse NMF for microarray data analysis. *Journal of Mathematical Biology*, **79**(1), 223–247 (2019). https://doi.org/10.1007/s00285-019-01355-2