# Supplementary Note 1

|  |
| --- |
| Algorithm 1. SNMF dimensionality reduction algorithm |
| Input: The matrix $A$ and the initial value of $W$ $(W\_{init})$ and the rank $k$ of $A.$Output: the non-negative factorization of $A$, $A≈ WH$ 1: repeat2: solving $\min\_{H(:,i)\geq 0}\left‖\left[\begin{array}{c}W\\α\_{1×k}\end{array}\right]H\left(:,i\right)-\left[\begin{array}{c}A\left(:,i\right)\\0\end{array}\right]\right‖\_{2}^{2},i=1,2,…,n$ in parallel;3: solving $\min\_{W(j,:)\geq 0}∥W\left(j,:\right)\left[\begin{array}{c}H,β\_{k×1}\end{array}\right]-\left[\begin{array}{c}A\left(j,:\right),0\end{array}\right]∥\_{2}^{2},j=1,2,…,m $in parallel;4: until the relative difference in the object function between two adjacent loops is less than some threshold ($1×10^{-5}$ in this paper) or the maximum specified number of iterations is reached. |

$$\begin{array}{c}\begin{array}{c}\min\_{W,H\geq 0}∥A-WH∥\_{F}^{2}+α^{2}\sum\_{i=1}^{n}∥H\left(:,i\right)∥\_{1}^{2}+β^{2}\sum\_{j=1}^{m}∥W\left(j,:\right)∥\_{1}^{2}\end{array}\#\left(1\right)\end{array}$$

The detailed process of alternate iteration algorithm used to solve mode (1) was described in algorithm 1. Here, we depict more details of step 2 in Algorithm 1. After $W$ is fixed, the optimization question (1) is equivalent to mode (2).

 $\begin{array}{c}\min\_{H\geq 0}∥A-WH∥\_{F}^{2}+α^{2}\sum\_{i=1}^{n}∥H\left(:,i\right)∥\_{1}^{2}\#\left(2\right)\end{array}$

Notice that

$$\begin{matrix}&∥A-WH∥\_{F}^{2}+α^{2}\sum\_{i=1}^{n}∥H\left(:,i\right)∥\_{1}^{2} \\&=\sum\_{i=1}^{n}(∥WH(:,i)-A(:,i)∥\_{2}^{2}+α^{2}∥H(:,i)∥\_{1}^{2})\\&=\sum\_{i=1}^{n}\left‖\left[\begin{array}{c}W\\α\_{1×k}\end{array}\right]H(:,i)-\left[\begin{array}{c}A(:,i)\\0\end{array}\right]\right‖\_{2}^{2} \end{matrix}$$

where $α\_{1×k}$ is a vector that all elements are $α$. Therefore, solving the problem (2) is equivalent to solve n non-negativity-constrained least squares problems (3) in parallel, i.e.

$\begin{array}{c}\min\_{H\left(:,i\right)\geq 0}\left‖\left[\begin{array}{c}W\\α\_{1×k}\end{array}\right]H\left(:,i\right)-\left[\begin{array}{c}A\left(:,i\right)\\0\end{array}\right]\right‖\_{2}^{2}\#\left(3\right)\end{array}$It is easy to show that the problem (3) is equivalent to

$\begin{array}{c}\min\_{x\geq 0}∥C\_{k×k}x-b\_{k×1}∥\_{2}^{2}\#\left(4\right)\end{array}$where $C\_{k×k}=\left[\begin{array}{c}W\\α\_{1×k}\end{array}\right]^{T}\left[\begin{array}{c}W\\α\_{1×k}\end{array}\right],b\_{k×1}=\left[\begin{array}{c}W\\α\_{1×k}\end{array}\right]^{T}\left[\begin{array}{c}A(:,i)\\0\end{array}\right]$. Compared with model (3), model (4) can be solved with less computer time and less memory by using the method proposed by Lawson et al (Gentleman 1976). The rank of $A$ and the initial value of $W$ were estimated by the algorithm reported in WEDGE (Hu *et al.* 2019).

# Supplementary Note 2

The Gaussian mixture model is the sum of a set of Gaussian density functions (McLachlan *et al.* 2019), i.e.

$\begin{array}{c}\begin{array}{c}p\left(λ\right)=\sum\_{i=1}^{M}\left[w\_{i}N\left(μ\_{i},Σ\_{i}\right)\right]\end{array}\#\left(5\right)\end{array}$where $x$ is a D-dimensional continuous-valued feature vector, and $λ=\{(w\_{i},μ\_{i},Σ\_{i})|i=1\~M\}$($w\_{i}$ is the mixed weight). $N\left(μ\_{i},Σ\_{i}\right)$ is a D-variate Gaussian density function, expressed asl

$$\begin{array}{c}N\left(μ\_{i},Σ\_{i}\right)= \frac{1}{2π^{\frac{D}{2}}\left‖Σ\_{i}\right‖^{\frac{1}{2}}}e^{-\frac{1}{2}(x-μ\_{i})^{T}Σ\_{i}^{-1}\left(x-μ\_{i}\right)}\#\left(6\right)\end{array}$$

# Supplementary Note 3

**Theorem** Let $L=diag(L\_{1}, L\_{2}, …, L\_{k})$ be a block diagonal matrix with $L\_{i}$ being a square matrix of size $l\_{i}×l\_{i}$ whose diagonal elements are $l\_{i}-1$ and the other elements are $-1.$ Then $L$is semi-definite whose eigenvalues are 0 with multiplicity $k$, and $l\_{i}$ with multiplicity $l\_{i}-1, i=1,2,…,k.$ Furthermore, the eigen-space associated with the eigenvalue 0 is spanned by $1\_{1},1\_{2},…,1\_{k}$, here $1\_{i}=(0,…0,1,…,1,0,…,0)$ is an indicator vector whose entries are 1 at indices from $l\_{1}+…+l\_{i-1}+1$ to $l\_{1}+…+l\_{i}$, and 0 at other indices.

 **Proof:**

It is easy to show that

$\left|λI-L\right|=П\_{i=1}^{k}\left|λI-L\_{i}\right|$=$П\_{i=1}^{k}λ\left(λ-l\_{i}\right)^{l\_{i}-1}=λ^{k}П\_{i=1}^{k}\left(λ-l\_{i}\right)^{l\_{i}-1}.$

The first part of the conclusion follows immediately.

For the second part, since the sum of each row of $L\_{i}$ is zero, clearly the indicator vector $1\_{i}$ is an eigenvector associated with the eigenvalue 0, $i=1,2,…,k.$ On the other hand, the multiplicity of the eigenvalue 0 is $k$, thus $1\_{1},1\_{2},…,1\_{k}$ span the eigenspace of the eigenvalue 0.

References

Gentleman, W. M., 1976 Solving Least Squares Problems (Charles L. Lawson and Richard J. Hanson). SIAM Rev. 18: 518–520.

Hu, Y., B. Li, N. Liu, P. Cai, F. Chen *et al.*, 2019 WEDGE: recovery of gene expression values for sparse single-cell RNA-seq datasets using matrix decomposition. bioRxiv 864488.

McLachlan, G. J., S. X. Lee, and S. I. Rathnayake, 2019 Finite Mixture Models. Annu. Rev. Stat. Its Appl. 6: 355–378.