Subspace Structure-aware Spectral Clustering for Robust Subspace Clustering: Supplementary Material

1. Derivation of Eq. (10)

In this section, we consider the following problem:

$$\max_{\Sigma_1,\ldots,\Sigma_K \in \mathcal{S}} \sum_i^N \log \sum_k^K G_{ik} \mathcal{N}(\boldsymbol{x}_i; 0, \Sigma_k), \tag{23}$$

where $\mathcal{S} = \{\Sigma | \Sigma \in \mathcal{R}^{D \times D}, \Sigma = \Sigma^T \text{ and } \Sigma \succeq \sigma I\}$. In the following, we will derive that the maximizers of Eq. (23) $\hat{\Sigma}_1, \ldots, \hat{\Sigma}_K$ can be analytically solved as follows:

$$\hat{\Sigma}_k = U_k \mathrm{Diag}(\max(\boldsymbol{d}_k, \sigma \mathbf{1}_D)) U_k^T, \tag{24}$$

where $\frac{1}{\sum_i G_{ik}} \sum_i G_{ik} \boldsymbol{x}_i \boldsymbol{x}_i^T = U_k \mathrm{Diag}(\boldsymbol{d}_k) U_k^T$. Since $G_{ij} \in \{0, 1\}$ and $G\mathbf{1}_K = \mathbf{1}_N$, Eq. (23) can be represented as follows:

$$\begin{aligned}
&\max_{\Sigma_1,\ldots,\Sigma_K \in \mathcal{S}} \sum_i^N \log \sum_k^K G_{ik} \mathcal{N}(\boldsymbol{x}_i; 0, \Sigma_k) \\
=& \max_{\Sigma_1,\ldots,\Sigma_K \in \mathcal{S}} \sum_i^N \sum_k^K G_{ik} \log \mathcal{N}(\boldsymbol{x}_i; 0, \Sigma_k) \\
=& \max_{\Sigma_1,\ldots,\Sigma_K \in \mathcal{S}} \sum_k^K \sum_i^N G_{ik} \log \mathcal{N}(\boldsymbol{x}_i; 0, \Sigma_k) \\
=& \max_{\Sigma_1,\ldots,\Sigma_K \in \mathcal{S}} \sum_k^K \sum_{i \in \mathcal{I}_k} \log \mathcal{N}(\boldsymbol{x}_i; 0, \Sigma_k) \\
=& \sum_k^K \max_{\Sigma_k \in \mathcal{S}} \sum_{i \in \mathcal{I}_k} \log \mathcal{N}(\boldsymbol{x}_i; 0, \Sigma_k),
\end{aligned} \tag{25}$$

where $\mathcal{I}_k = \{i | 1 \leq i \leq N, G_{ik} = 1\}$. From Eq. (25), for deriving that Eq. (24) maximizes Eq. (23), it is sufficient to derive that the maximizer $\hat{\Sigma}_*$ of the following problem can be analytically solved by $\hat{\Sigma}_* = U_* \mathrm{Diag}(\max(\boldsymbol{d}_*, \sigma \mathbf{1}_D)) U_*^T$, where $\frac{1}{m_*} \sum_{i \in \mathcal{I}_*} \boldsymbol{x}_i \boldsymbol{x}_i^T = U_* \mathrm{Diag}(\boldsymbol{d}_*) U_*^T$ and $m_* = \#\mathcal{I}_*$:

$$\max_{\Sigma_* \in \mathcal{S}} \sum_{i \in \mathcal{I}_*} \log \mathcal{N}(\boldsymbol{x}_i; 0, \Sigma_*), \tag{26}$$

First, we consider the one-dimensional case (i.e., the case of $D = 1$). In this case, Eq. (26) can be represented as follows:

$$\begin{aligned}
&\max_{\Sigma_* \geq \sigma} \sum_{i \in \mathcal{I}_*} \log \mathcal{N}(\boldsymbol{x}_i; 0, \Sigma_*) \\
=& \max_{\Sigma_* \geq \sigma} -\frac{m}{2} \log(2\pi\Sigma_*) - \frac{m}{2\Sigma_*} d_*,
\end{aligned} \tag{27}$$

where $d_* = \frac{1}{m_*}\sum_i x_j^2$. Also, note that Σ_* and $\Sigma_*^\dagger$ are both scalar values since $D = 1$. Since $\Sigma_* \geq \sigma$, it can be easily shown that the maximizer of Eq. (27) is $\hat{\Sigma}_* = \max(d_*, \sigma)$.

Next, we consider Eq. (26). In this case, we can derive that the mazimizer of Eq. (26) is $\hat{\Sigma}_* = U_*\mathrm{Diag}(\max(\boldsymbol{d}_*, \sigma\mathbf{1}_D))U_*^T$ by first transforming data points with a matrix U_*^T so that each dimension of data points is decorrelated, then computing a covariance matrix $\mathrm{Diag}(\max(\boldsymbol{d}_k, \sigma\mathbf{1}_D))$ by solving Eq. (27) for each dimension, and finally re-transforming a computed matrix $\mathrm{Diag}(\max(\boldsymbol{d}_k, \sigma\mathbf{1}_D))$ with a matrix U_*.

2. Derivation of Eq. (11)

By applying singular value decomposition to $\mathrm{Diag}(\boldsymbol{g}_*)X$, we have:

$$\mathrm{Diag}(\boldsymbol{g}_*)X = V_*\Lambda_*U_*^T, \tag{28}$$

where Λ_* is a matrix such that $\Lambda_* \in \mathcal{R}^{N\times D}$ and $(\Lambda_*)_{ij} = (\boldsymbol{\lambda}_*)_i$ if $i = j$, otherwise 0, and V_* is a matrix such that $V_* \in \mathcal{R}^{N\times N}$ and $V_*^TV_* = I$. Since a uncentered covariance matrix $\frac{1}{m}X^T\mathrm{Diag}(\boldsymbol{g}_*)\mathrm{Diag}(\boldsymbol{g}_*)X = U_*\mathrm{Diag}(\boldsymbol{d}_*)U_*^T$, we have:

$$\begin{aligned}
&U_*\mathrm{Diag}(\boldsymbol{d}_*)U_*^T\\
=&\frac{1}{m_*}X^T\mathrm{Diag}(\boldsymbol{g}_*)\mathrm{Diag}(\boldsymbol{g}_*)X\\
=&\frac{1}{m_*}U_*\mathrm{Diag}(\boldsymbol{\lambda}_k)^2U_*^T,
\end{aligned} \tag{29}$$

therefore

$$(\lambda_*^2)_d = m_*(\boldsymbol{d}_*)_d. \tag{30}$$

First, by substituting $\hat{\Sigma}_* = U_*\mathrm{Diag}(\max(\boldsymbol{d}_*, \sigma\mathbf{1}_D))U_*^T$ into Eq. (26), we have:

$$\begin{aligned}
&\sum_{i\in\mathcal{I}_*}\log\mathcal{N}(\boldsymbol{x}_i; 0, \Sigma_*)\\
=&-\sum_d^D\frac{m_*}{2}\log(2\pi\max(\sigma, (\boldsymbol{d}_*)_d)) + \frac{m_*}{2\max(\sigma, (\boldsymbol{d}_*)_d)}(\boldsymbol{d}_*)_d\\
=&-m_*\sum_d^D\log\max(1, \sqrt{\frac{(\boldsymbol{d}_*)_d}{\sigma}}) + \frac{1}{2}\min(1, \frac{(\boldsymbol{d}_*)_d}{\sigma}) + \frac{1}{2}\log 2\pi\sigma\\
=&-m_*\sum_d^D\log\max(1, \sqrt{\frac{(\boldsymbol{d}_*)_d}{\sigma}}) + \frac{1}{2}\min(1, \frac{(\boldsymbol{d}_*)_d}{\sigma}) + \frac{1}{2}\log 2\pi\sigma\\
=&-m_*\sum_d^D\log\max(1, \frac{(\boldsymbol{\lambda}_*)_d}{\rho_*}) + \frac{1}{2}\min(1, \frac{(\boldsymbol{\lambda}_*)_d^2}{\rho_*^2}) + \frac{1}{2}\log 2\pi\sigma\\
=&-\sum_d^D\mathrm{f}_{\rho_*}((\boldsymbol{\lambda}_*)_d) + const.,
\end{aligned} \tag{31}$$

where $\rho_* = \sqrt{m_*\sigma}$ and $\mathrm{f}_{\rho_*}(\lambda) = m_*\log\max(1, \frac{\lambda}{\rho_*}) + m_*\frac{1}{2}\min(1, \frac{\lambda^2}{\rho_*^2})$.

From Eq. (25) and Eq. (31), we have:

$$\begin{aligned}
&\mathrm{r}(G, X) \\
&= -\max_{\Sigma_1,\ldots,\Sigma_K\in\mathcal{S}} \sum_i^N \log \sum_k^K G_{ik}\mathcal{N}(\boldsymbol{x}_i; 0, \Sigma_k) \\
&= -\sum_k^K \max_{\Sigma_k\in\mathcal{S}} \sum_{i\in\mathcal{I}_k} \log\mathcal{N}(\boldsymbol{x}_i; 0, \Sigma_k) \\
&= -\sum_k^K \sum_d^D \mathrm{f}_{\rho_k}((\boldsymbol{\lambda}_k)_d) + const.
\end{aligned} \tag{32}$$

3. Derivation of Eq. (17)

Because we can derive the maximizer of the problem $\max_{\Sigma_1,\ldots,\Sigma_K\in\mathcal{S}} \sum_i^N \sum_k^K q(Z_{ik}=1)\log\mathcal{N}(\boldsymbol{x}_i; 0, \Sigma_k)$ by following steps written in section 1, we omit derivation of this.

4. Proof of Proposition 1

Since we assume $\eta = 1$ and $\Sigma_1, \ldots, \Sigma_K$ are fixed, we can rewrite Eq. (13) as follows:

$$\max_{G\in\mathcal{H}} \sum_i^N \log \sum_k^K G_{ik}C_{ik}, \tag{33}$$

where $C_{ik} = \mathcal{N}(\boldsymbol{x}_i; 0, \Sigma_k)$, which is constant. For each i, if there is only a single value that equals to $\max_k C_{ik}$ in $C_{i1}, \ldots, C_{iK}$, the optimal solution $\hat{G} \in \mathcal{H}$ satisfies $\hat{G}_{ik} =< k = \underset{k'\in\{1,\ldots,K\}}{\arg\max}\ \hat{G}_{ik'} >= G^\dagger_{ik}$, and even if it is not satisfied, $k^\dagger = \underset{k'\in\{1,\ldots,K\}}{\arg\max}\ \hat{G}_{ij'}$ satisfies $C_{ik^\dagger} = \max_k C_{ik}$, hence $\sum_k^K \hat{G}_{ik}C_{ik} = \sum_k^K G^\dagger_{ik}C_{ik}$ holds. Therefore, $G^\dagger$ obtained by Eq. (21) is also the optimal solution of Eq. (13).

5. The Optimal Solutions of Soft Spectral Clustering Problem

In this section, we prove that an assignment matrix G is always one of the optimal solutions of $\epsilon(G, M) = \sum_{l=1}^K \frac{\boldsymbol{g}_l^T M\boldsymbol{g}_l}{\boldsymbol{g}_l^T D\boldsymbol{g}_l}$ if it can be factorized in the form of $G = \mathbf{1}_N\boldsymbol{v}^T$, where $\boldsymbol{v}$ is a vector in the $(K-1)$-simplex and satisfies $(\boldsymbol{v})_i > 0$ for all i. We also assume that an affinity matrix M satisfies $M = M^T$ and $M_{ij} \geq 0$.

Since $D_{ii} = \sum_j M_{ij}$ and $D_{ij} = 0$ if $i \neq j$, we have:

$$\begin{aligned}
\boldsymbol{g}^T M\boldsymbol{g} - \boldsymbol{g}^T D\boldsymbol{g} &= \sum_{i,j} \boldsymbol{g}_i\boldsymbol{g}_j M_{ij} - \sum_{i,j} \boldsymbol{g}_i^2 M_{ij} \\
&= -\sum_{i,j} (\boldsymbol{g}_i - \boldsymbol{g}_j)^2 M_{ij} \leq 0
\end{aligned} \tag{34}$$

Specifically, Eq. (34) is maximized if $(\boldsymbol{g})_i = (\boldsymbol{g})_j$ holds for all (i, j).

By Eq. (34), we have:

$$\begin{aligned}
\epsilon(G, M) &= \sum_{l=1}^K \frac{\boldsymbol{g}_l^T M\boldsymbol{g}_l}{\boldsymbol{g}_l^T D\boldsymbol{g}_l} \\
&= K + \sum_{l=1}^K \frac{\boldsymbol{g}_l^T M\boldsymbol{g}_l - \boldsymbol{g}_l^T D\boldsymbol{g}_l}{\boldsymbol{g}_l^T D\boldsymbol{g}_l} \\
&= K - \sum_{l=1}^K \frac{\sum_{i,j}((\boldsymbol{g}_l)_i - (\boldsymbol{g}_l)_j)^2 M_{ij}}{\boldsymbol{g}_l^T D\boldsymbol{g}_l} \\
&\leq K
\end{aligned} \tag{35}$$

Specifically, $\epsilon(G, M)$ is maximized if $(\boldsymbol{g}_l)_i = (\boldsymbol{g}_l)_j$ holds for all (i, j, l). Therefore, when $G = \mathbf{1}_N \boldsymbol{v}^T$ and $\boldsymbol{v}$ is in the $(K-1)$-simplex and satisfies $(\boldsymbol{v})_i > 0$ for all i, it is one of the optimal solutions of $\epsilon(G, M) = \sum_{l=1}^{K} \frac{\boldsymbol{g}_l^T M \boldsymbol{g}_l}{\boldsymbol{g}_l^T D \boldsymbol{g}_l}$.