# Appendix of "Partial Clustering Ensemble"

## Appendix A: Proof of Theorem 1

The $\mathbf{H}$-subproblem is:

$$\min_{\mathbf{H}^T\mathbf{H}=\mathbf{I}} \quad tr(\mathbf{H}^T\mathbf{D}\mathbf{H}) - 2tr(\mathbf{H}^T\mathbf{C}), \tag{1}$$

where $\mathbf{D} = \mathbf{V}^2 + \gamma\mathbf{I}$ and $\mathbf{C} = \gamma\mathbf{Y}\mathbf{R}^T + \mathbf{V}^2\sum_{i=1}^m \alpha_i \mathbf{Y}^{(i)}\mathbf{R}^{(i)}$. Let $\mathbf{H}^t$ denote the value of $\mathbf{H}$ in the $t$-th iteration. Given a step size $\eta > 0$, we denote $\mathbf{M} = [\eta(\mathbf{D}\mathbf{H}^t - \mathbf{C}), -\eta\mathbf{H}^t]$ and $\mathbf{N} = [\mathbf{H}^t, \mathbf{D}\mathbf{H}^t - \mathbf{C}]^T$. The following Theorem provides an update formula of $\mathbf{H}^{t+1}$:

**Theorem 1.** *Suppose $\mathbf{H}^t$, $\mathbf{M}$ and $\mathbf{N}$ be defined as before, if ${\mathbf{H}^t}^T\mathbf{H}^t = \mathbf{I}$, update $\mathbf{H}^{t+1}$ as follows:*

$$\mathbf{H}^{t+1} = \mathbf{H}^t - \mathbf{M}\mathbf{N}\mathbf{H}^t - \mathbf{M}(\mathbf{I}+\mathbf{N}\mathbf{M})^{-1}(\mathbf{N}\mathbf{H}^t - \mathbf{N}\mathbf{M}\mathbf{N}\mathbf{H}^t). \tag{2}$$

*Then, ${\mathbf{H}^{t+1}}^T\mathbf{H}^{t+1} = \mathbf{I}$, and this updating is in a descent direction of Eq.*(1). *Since Eq.*(1) *has a lower bound, the iteration method converges. Moreover, it can converge to a stable point.*

*Proof.* According to Woodbury identity, we have

$$\begin{aligned}\mathbf{H}^{t+1} =& \mathbf{H}^t - \mathbf{M}\mathbf{N}\mathbf{H}^t - \mathbf{M}(\mathbf{I}+\mathbf{N}\mathbf{M})^{-1}\mathbf{N}\mathbf{H}^t + \mathbf{M}(\mathbf{I}+\mathbf{N}\mathbf{M})^{-1}\mathbf{N}\mathbf{M}\mathbf{N}\mathbf{H}^t \\ =& (\mathbf{I} - \mathbf{M}(\mathbf{I}+\mathbf{N}\mathbf{M})^{-1}\mathbf{N})(\mathbf{I}-\mathbf{M}\mathbf{N})\mathbf{H}^t \\ =& (\mathbf{I}+\mathbf{M}\mathbf{N})^{-1}(\mathbf{I}-\mathbf{M}\mathbf{N})\mathbf{H}^t\end{aligned} \tag{3}$$

Let $\mathbf{Q} = \frac{1}{\eta}\mathbf{M}\mathbf{N}$, we have

$$\mathbf{H}^{t+1} = (\mathbf{I}+\eta\mathbf{Q})^{-1}(\mathbf{I}-\eta\mathbf{Q})\mathbf{H}^t \tag{4}$$

We first prove that ${\mathbf{H}^{t+1}}^T\mathbf{H}^{t+1} = \mathbf{I}$. Let us take a closer look at $\mathbf{Q}$:

$$\mathbf{Q} = \frac{1}{\eta}\mathbf{M}\mathbf{N} = \mathbf{D}\mathbf{H}^t\mathbf{H}^{tT} - \mathbf{C}\mathbf{H}^{tT} - \mathbf{H}^t(\mathbf{D}\mathbf{H}^t - \mathbf{C})^T \tag{5}$$

It is easy to verify that $\mathbf{Q}$ is a skew-symmetric matrix, i.e., $\mathbf{Q} = -\mathbf{Q}^T$. Then, we compute ${\mathbf{H}^{t+1}}^T\mathbf{H}^{t+1}$:

$$\begin{aligned}
{\mathbf{H}^{t+1}}^T\mathbf{H}^{t+1} =& \mathbf{H}^{tT}\left(\mathbf{I}-\eta\mathbf{Q}\right)^T\left(\left(\mathbf{I}+\eta\mathbf{Q}\right)^T\right)^{-1}\left(\mathbf{I}+\eta\mathbf{Q}\right)^{-1}\left(\mathbf{I}-\eta\mathbf{Q}\right)\mathbf{H}^t \\
=& \mathbf{H}^{tT}\left(\mathbf{I}+\eta\mathbf{Q}\right)\left(\mathbf{I}-\eta\mathbf{Q}\right)^{-1}\left(\mathbf{I}+\eta\mathbf{Q}\right)^{-1}\left(\mathbf{I}-\eta\mathbf{Q}\right)\mathbf{H}^t \\
=& \mathbf{H}^{tT}\left(\mathbf{I}+\eta\mathbf{Q}\right)\left(\left(\mathbf{I}+\eta\mathbf{Q}\right)\left(\mathbf{I}-\eta\mathbf{Q}\right)\right)^{-1}\left(\mathbf{I}-\eta\mathbf{Q}\right)\mathbf{H}^t.
\end{aligned} \tag{6}$$

Furthermore, we have

$$\left(\mathbf{I}+\eta\mathbf{Q}\right)\left(\mathbf{I}-\eta\mathbf{Q}\right) = \mathbf{I}-\eta^2\mathbf{Q}\mathbf{Q} = \left(\mathbf{I}-\eta\mathbf{Q}\right)\left(\mathbf{I}+\eta\mathbf{Q}\right). \tag{7}$$

Taking it back to Eq.(6), we have

$$\begin{aligned}
{\mathbf{H}^{t+1}}^T\mathbf{H}^{t+1} =& \mathbf{H}^{tT}\left(\mathbf{I}+\eta\mathbf{Q}\right)\left(\left(\mathbf{I}-\eta\mathbf{Q}\right)\left(\mathbf{I}+\eta\mathbf{Q}\right)\right)^{-1}\left(\mathbf{I}-\eta\mathbf{Q}\right)\mathbf{H}^t \\
=& \mathbf{H}^{tT}\left(\mathbf{I}+\eta\mathbf{Q}\right)\left(\mathbf{I}+\eta\mathbf{Q}\right)^{-1}\left(\mathbf{I}-\eta\mathbf{Q}\right)^{-1}\left(\mathbf{I}-\eta\mathbf{Q}\right)\mathbf{H}^t \\
=& \mathbf{H}^{tT}\mathbf{H}^t \\
=& \mathbf{I}.
\end{aligned}$$

Then we prove that updating $\mathbf{H}^{t+1}$ by Eq.(2) is in a descent direction. To prove it, we first provide the following lemma:

**Lemma 1.** *Given the objective function* $\mathcal{J}(\mathbf{H}^{t+1}) = tr({\mathbf{H}^{t+1}}^T\mathbf{D}\mathbf{H}^{t+1}) - 2tr({\mathbf{H}^{t+1}}^T\mathbf{C})$ *defined in Eq.*(1)*, if we update* $\mathbf{H}^{t+1}$ *by Eq.*(2)*, we have:*

$$\left.\frac{\partial\mathcal{J}(\mathbf{H}^{t+1})}{\partial\eta}\right|_{\eta=0} = -2\|\mathbf{Q}\|_F^2 \le 0. \tag{8}$$

*Proof.* According to the chain rule, we have

$$\frac{\partial\mathcal{J}(\mathbf{H}^{t+1})}{\partial\eta} = tr\left(\left(\frac{\partial\mathcal{J}(\mathbf{H}^{t+1})}{\partial\mathbf{H}^{t+1}}\right)^T\frac{\partial\mathbf{H}^{t+1}}{\partial\eta}\right) \tag{9}$$

When $\eta = 0$, $\mathbf{H}^{t+1} = \mathbf{H}^t$, and $\left.\frac{\partial\mathcal{J}(\mathbf{H}^{t+1})}{\partial\mathbf{H}^{t+1}}\right|_{\eta=0} = 2(\mathbf{D}\mathbf{H}^t - \mathbf{C})$, $\left.\frac{\partial\mathbf{H}^{t+1}}{\partial\eta}\right|_{\eta=0} = -2\mathbf{Q}\mathbf{H}^t$.

On one hand, we have

$$\begin{aligned}
\left.\frac{\partial\mathcal{J}(\mathbf{H}^{t+1})}{\partial\eta}\right|_{\eta=0} =& -4tr\left((\mathbf{D}\mathbf{H}^t-\mathbf{C})^T\mathbf{Q}\mathbf{H}^t\right) \\
=& -4tr\left((\mathbf{D}\mathbf{H}^t-\mathbf{C})^T(\mathbf{D}\mathbf{H}^t-\mathbf{C}) - (\mathbf{D}\mathbf{H}^t-\mathbf{C})^T\mathbf{H}^t(\mathbf{D}\mathbf{H}^t-\mathbf{C})^T\mathbf{H}^t\right)
\end{aligned} \tag{10}$$

On the other hand, we have

$$\begin{aligned}
\|\mathbf{Q}\|_F^2 =& tr(\mathbf{Q}^T\mathbf{Q}) \\
=& tr\left(\left((\mathbf{D}\mathbf{H}^t-\mathbf{C})\mathbf{H}^{tT} - \mathbf{H}^t(\mathbf{D}\mathbf{H}^t-\mathbf{C})^T\right)^T\left((\mathbf{D}\mathbf{H}^t-\mathbf{C})\mathbf{H}^{tT} - \mathbf{H}^t(\mathbf{D}\mathbf{H}^t-\mathbf{C})^T\right)\right) \\
=& 2tr\left((\mathbf{D}\mathbf{H}^t-\mathbf{C})^T(\mathbf{D}\mathbf{H}^t-\mathbf{C}) - (\mathbf{D}\mathbf{H}^t-\mathbf{C})^T\mathbf{H}^t(\mathbf{D}\mathbf{H}^t-\mathbf{C})^T\mathbf{H}^t\right)
\end{aligned} \tag{11}$$

Therefore, we have $\left.\frac{\partial\mathcal{J}(\mathbf{H}^{t+1})}{\partial\eta}\right|_{\eta=0} = -2\|\mathbf{Q}\|_F^2 \le 0$. □

Lemma 1 shows that if $\mathbf{H}$ moves a small step $\Delta\eta > 0$ in the update direction, the objective function $\mathcal{J}$ will have a change $-2\|\mathbf{Q}\|_F^2\Delta\eta$ and since $-2\|\mathbf{Q}\|_F^2 \leq 0$, the objective function $\mathcal{J}$ will decrease. Thus the update direction is a descent direction. Moreover, since $\mathbf{H}$ is an orthogonal matrix whose elements are all bounded, the objective function Eq.(1) has a lower bound, and the algorithm will converge.

To prove that it will converge to a stable point, we introduce the following lemma which shows the first-order optimality condition of the objective function:

**Lemma 2.** *Let $\mathcal{L} = tr(\mathbf{H}^T\mathbf{D}\mathbf{H}) - 2tr(\mathbf{H}^T\mathbf{C}) - tr(\mathbf{\Lambda}(\mathbf{H}^T\mathbf{H} - \mathbf{I}))$ be the Lagrangian function of our objective function, where $\mathbf{\Lambda}$ is the Lagrangian multiplier, then $\frac{\partial\mathcal{L}}{\partial\mathbf{H}} = \mathbf{0}$ if and only if $\mathbf{Q} = \mathbf{0}$, so $\mathbf{Q} = \mathbf{0}$ is the first-order optimality condition of our objective function.*

*Proof.* Set the partial derivative of $\mathcal{L}$ w.r.t. $\mathbf{H}$ to zero:

$$\frac{\partial\mathcal{L}}{\partial\mathbf{H}} = 2\left(\mathbf{D}\mathbf{H} - \mathbf{C} - \mathbf{H}\mathbf{\Lambda}\right) = \mathbf{0}. \tag{12}$$

By multiplying both sides of Eq.(12) by $\mathbf{H}^T$ and applying the constraint $\mathbf{H}^T\mathbf{H} = \mathbf{I}$, we can solve $\mathbf{\Lambda}$ as $\mathbf{\Lambda} = \mathbf{H}^T(\mathbf{D}\mathbf{H} - \mathbf{C})$. Note that $\mathbf{H}^T\mathbf{H}$ is symmetric, and its corresponding Lagrangian multiplier $\mathbf{\Lambda}$ is also symmetric. So we rewrite $\mathbf{\Lambda}$ as $\mathbf{\Lambda} = (\mathbf{D}\mathbf{H} - \mathbf{C})^T\mathbf{H}$. Putting it back into Eq.(12), we obtain

$$\frac{\partial\mathcal{L}}{\partial\mathbf{H}} = 2\left(\mathbf{D}\mathbf{H}\mathbf{H}^T - \mathbf{C}\mathbf{H}^T - \mathbf{H}(\mathbf{D}\mathbf{H} - \mathbf{C})^T\right)\mathbf{H} = 2\mathbf{Q}\mathbf{H}. \tag{13}$$

On one hand, we have $\frac{\partial\mathcal{L}}{\partial\mathbf{H}} = 2\mathbf{Q}\mathbf{H}$, so if $\mathbf{Q} = \mathbf{0}$, then $\frac{\partial\mathcal{L}}{\partial\mathbf{H}} = \mathbf{0}$.

On the other hand, if $\frac{\partial\mathcal{L}}{\partial\mathbf{H}} = \mathbf{0}$, i.e., $\left(\mathbf{D}\mathbf{H}\mathbf{H}^T - \mathbf{C}\mathbf{H}^T - \mathbf{H}(\mathbf{D}\mathbf{H} - \mathbf{C})^T\right)\mathbf{H} = \mathbf{0}$. Let $\mathbf{Z} = \mathbf{D}\mathbf{H} - \mathbf{C}$, then we have $\mathbf{Z} = \mathbf{H}\mathbf{Z}^T\mathbf{H}$ due to $\mathbf{H}^T\mathbf{H} = \mathbf{I}$. Thus,

$$\mathbf{Z} = \mathbf{H}\mathbf{Z}^T\mathbf{H} = \mathbf{H}(\mathbf{H}\mathbf{Z}^T\mathbf{H})^T\mathbf{H} = \mathbf{H}\mathbf{H}^T\mathbf{Z} \tag{14}$$

Taking the transposition of both sides, we have $\mathbf{Z}^T = \mathbf{Z}^T\mathbf{H}\mathbf{H}^T$. Then we obtain

$$\mathbf{H}\mathbf{Z}^T = \mathbf{H}\mathbf{Z}^T\mathbf{H}\mathbf{H}^T = \mathbf{Z}\mathbf{H}^T \tag{15}$$

which means $\mathbf{Z}\mathbf{H}^T - \mathbf{H}\mathbf{Z}^T = \mathbf{0}$. Note that $\mathbf{Q} = \mathbf{Z}\mathbf{H}^T - \mathbf{H}\mathbf{Z}^T$, so $\mathbf{Q} = \mathbf{0}$. In summary, $\mathbf{Q} = \mathbf{0}$ is the first-order optimality condition. □

Now, get back to Theorem 1. The algorithm converges when $\frac{\partial\mathcal{J}(\mathbf{H}^{t+1})}{\partial\eta}\Big|_{\eta=0} = 0$, which means $\mathbf{H}$ cannot move a small step in the descent direction to make the objective function decreases. Since $\frac{\partial\mathcal{J}(\mathbf{H}^{t+1})}{\partial\eta}\Big|_{\eta=0} = -2\|\mathbf{Q}\|_F^2$, $\|\mathbf{Q}\|_F^2 = 0$, i.e., $\mathbf{Q} = \mathbf{0}$. Due to Lemma 2, it satisfies the first-order optimality condition, so the algorithm converges to a stable point. □

# Appendix B: Proof of Theorem 2

The $\boldsymbol{\alpha}$-subproblem is:

$$\min_{\boldsymbol{\alpha}} \quad \boldsymbol{\alpha}^T \mathbf{G} \boldsymbol{\alpha} - 2\mathbf{f}^T \boldsymbol{\alpha}, \qquad (16)$$
$$s.t. \quad 0 \le \alpha_i \le 1, \quad \sum_{i=1}^{m} \alpha_i = 1.$$

where the $(i,j)$-th element of $\mathbf{G}$ is $G_{ij} = tr(\mathbf{R}^{(i)T}\mathbf{Y}^{(i)T}\mathbf{V}^2\mathbf{Y}^{(j)}\mathbf{R}^{(j)})$ and the $i$-th element of vector $\mathbf{f}$ is $f_i = tr(\mathbf{R}^{(i)T}\mathbf{Y}^{(i)T}\mathbf{V}^2\mathbf{H})$. Then, we have the following Theorem about its convexity:

**Theorem 2.** *Eq.*(16) *is a convex quadratic programming.*

*Proof.* Obviously, Eq.(16) is a quadratic programming, and the constraint is a convex set. To prove it is a convex quadratic programming, we just need to prove that $\mathbf{G}$ is a positive semi-definite matrix. Given any non-zero vector $\mathbf{x} \in \mathbb{R}^m$, we compute:

$$\begin{aligned}
\mathbf{x}^T\mathbf{G}\mathbf{x} &= \sum_{i,j=1}^{m} x_i G_{ij} x_j \qquad (17)\\
&= \sum_{i,j=1}^{m} x_i x_j tr(\mathbf{R}^{(i)T}\mathbf{Y}^{(i)T}\mathbf{V}^2\mathbf{Y}^{(j)}\mathbf{R}^{(j)})\\
&= tr\left(\sum_{i=1}^{m} x_i \mathbf{R}^{(i)T}\mathbf{Y}^{(i)T}\mathbf{V}^2 \sum_{j=1}^{m} x_j \mathbf{Y}^{(j)}\mathbf{R}^{(j)}\right)\\
&= tr\left(\left(\sum_{i=1}^{m} x_i \mathbf{Y}^{(i)}\mathbf{R}^{(i)}\right)\left(\sum_{i=1}^{m} x_i \mathbf{Y}^{(i)}\mathbf{R}^{(i)}\right)^T \mathbf{V}^2\right)
\end{aligned}$$

Denoting $\mathbf{A} = \sum_{i=1}^{m} x_i \mathbf{Y}^{(i)}\mathbf{R}^{(i)}$, we have

$$\mathbf{x}^T\mathbf{G}\mathbf{x} = tr(\mathbf{A}\mathbf{A}^T diag(\mathbf{v})^2) = \sum_{p=1}^{m} v_p^2 \|\mathbf{A}_{p\cdot}\|_2^2 \ge 0. \qquad (18)$$

Therefore, $\mathbf{G}$ is a positive semi-definite matrix, and thus Eq.(16) is convex quadratic programming. □

# Appendix C: Proof of Theorem 3

The $\mathbf{R}^{(i)}$-subproblem is:

$$\min_{\mathbf{R}^{(i)T}\mathbf{R}^{(i)}=\mathbf{I}} \quad tr(\mathbf{K}\mathbf{R}^{(i)}), \qquad (19)$$

where $\mathbf{K} = \sum_{j:j\neq i} \alpha_j \mathbf{R}^{(j)T}\mathbf{Y}^{(j)T}\mathbf{V}^2\mathbf{Y}^{(i)} - \mathbf{H}^T\mathbf{V}^2\mathbf{Y}^{(i)}$.

The following Theorem provides its global optima:

**Theorem 3.** *Supposing the singular value decomposition (SVD) of* $-\mathbf{K}^T$ *is* $-\mathbf{K}^T = \mathbf{U}\boldsymbol{\Sigma}\mathbf{S}^T$*, then the global optima of Eq.*(19) *is* $\mathbf{R}^{(i)} = \mathbf{U}\mathbf{S}^T$.

*Proof.* Denote $\mathbf{W} = -\mathbf{K}^T$ and we have its SVD is $\mathbf{W} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{S}^T$. Notice that to minimize $tr(\mathbf{K}\mathbf{R}^{(i)})$ is equivalent to maximize $tr(\mathbf{W}^T\mathbf{R}^{(i)})$. Since $\mathbf{R}^{(i)}$ is an orthogonal matrix, its SVD is $\mathbf{R}^{(i)} = \mathbf{R}^{(i)} * \mathbf{I} * \mathbf{I}$.

According to Von Neumanns trace inequality, we have

$$\begin{aligned} tr(\mathbf{W}^T\mathbf{R}^{(i)}) \leq & tr(\boldsymbol{\Sigma}\mathbf{I}) \\ = & tr(\boldsymbol{\Sigma}\mathbf{U}^T\mathbf{U}\mathbf{S}^T\mathbf{S}) \\ = & tr(\mathbf{S}\boldsymbol{\Sigma}\mathbf{U}^T\mathbf{U}\mathbf{S}^T) \\ = & tr(\mathbf{W}^T\mathbf{U}\mathbf{S}^T) \end{aligned} \tag{20}$$

Obviously, the equality holds when $\mathbf{R}^{(i)} = \mathbf{U}\mathbf{S}^T$. Therefore, the global optima of Eq.(19) is $\mathbf{R}^{(i)} = \mathbf{U}\mathbf{S}^T$. □