Supplementary Material for 'Reduction Techniques for Graph-based Convex Clustering'

Proofs

The Lagrangian Dual of TCC

Consider problem (5). Let $\mathbf{Y} = \widetilde{\mathbf{X}} - \widetilde{\mathbf{D}}\mathbf{B}$, and the Lagrangian function is

$$L(\mathbf{Y}, \mathbf{B}; \lambda\boldsymbol{\Theta}) = \frac{1}{2}\|\mathbf{Y}\|_F^2 + \lambda\|\mathbf{W}\mathbf{B}\|_{1,q} + \langle \lambda\boldsymbol{\Theta}, \widetilde{\mathbf{X}} - \widetilde{\mathbf{D}}\mathbf{B} - \mathbf{Y}\rangle$$
$$= \frac{1}{2}\|\mathbf{Y}\|_F^2 + \langle \lambda\boldsymbol{\Theta}, \widetilde{\mathbf{X}} - \mathbf{Y}\rangle - \langle \lambda\boldsymbol{\Theta}, \widetilde{\mathbf{D}}\mathbf{B}\rangle + \lambda\|\mathbf{W}\mathbf{B}\|_{1,q}, \quad (20)$$

where $\lambda\boldsymbol{\Theta} \in \mathbb{R}^{n\times p}$ is the Lagrangian multiplier. To find the dual, we need to solve the following problems:

$$\min_{\mathbf{Y}} f_1(\mathbf{Y}) = \frac{1}{2}\|\mathbf{Y}\|_F^2 - \langle \lambda\boldsymbol{\Theta}, \mathbf{Y}\rangle,$$
$$\min_{\mathbf{B}} f_2(\mathbf{B}) = -\langle \lambda\boldsymbol{\Theta}, \widetilde{\mathbf{D}}\mathbf{B}\rangle + \lambda\|\mathbf{W}\mathbf{B}\|_{1,q}.$$

By setting $\frac{\partial f_1(\mathbf{Y})}{\partial \mathbf{Y}} = 0$, we obtain

$$\mathbf{Y}^* = \lambda\boldsymbol{\Theta}.$$

Therefore,

$$\min_{\mathbf{Y}} f_1(\mathbf{Y}) = f_1(\mathbf{Y}^*) = -\frac{\lambda^2}{2}\left\|\boldsymbol{\Theta} - \frac{\widetilde{\mathbf{X}}}{\lambda}\right\|_F^2 + \|\widetilde{\mathbf{X}}\|_F^2.$$

Consider $f_2(\mathbf{B})$. Let $\boldsymbol{\beta}_r$ be the rth row of $\mathbf{B}$. The optimality condition is

$$\mathbf{0} \in \partial f_2(\boldsymbol{\beta}_r) = -\lambda\left[\widetilde{\mathbf{D}}^T\boldsymbol{\Theta}\right]_r + \lambda w_r \partial\|\boldsymbol{\beta}_r\|_q,$$
$$= -\lambda\widetilde{\mathbf{d}}_r\boldsymbol{\Theta} + \lambda w_r\partial\|\boldsymbol{\beta}_r\|_q,$$

where $r \leftrightarrow (i,j)$, i.e. the rth row of $\mathbf{B}$ corresponds to the data pair (i,j), and $\widetilde{\mathbf{d}}_r$ is the rth row of $\widetilde{\mathbf{D}}^T$. The above subgradient leads to

$$\widetilde{\mathbf{d}}_r\boldsymbol{\Theta} = w_r\mathbf{v}_r,\ \mathbf{v}_r \in \partial\|\boldsymbol{\beta}_r\|_q. \quad (21)$$

By noting that

$$\langle \mathbf{v}_r, \boldsymbol{\beta}_r\rangle = \|\boldsymbol{\beta}_r\|_q,$$

we have

$$\left\langle \widetilde{\mathbf{d}}_r\boldsymbol{\Theta}, \boldsymbol{\beta}_r\right\rangle = w_r\|\boldsymbol{\beta}_r\|_q.$$

Thus we can see that

$$0 = \min_{\boldsymbol{\beta}_r} f_2(\boldsymbol{\beta}_r).$$

Eq. (21) implies that

$$\widetilde{\mathbf{d}}_r\boldsymbol{\Theta} \in w_r\mathcal{B}_{\bar{q}},$$

where $\bar{q} = \frac{q}{q-1}$ and $\mathcal{B}_{\bar{q}}$ is the unit $\ell_{\bar{q}}$ ball in $\mathbb{R}^p$. Combining all the derivations above, the dual of problem (5) is

$$\min_{\boldsymbol{\Theta}}\left\{ g(\boldsymbol{\Theta}) = \frac{\lambda^2}{2}\left\|\boldsymbol{\Theta} - \frac{\widetilde{\mathbf{X}}}{\lambda}\right\|_F^2 - \|\widetilde{\mathbf{X}}\|_F^2 : \left\|\widetilde{\mathbf{d}}_r\boldsymbol{\Theta}\right\|_{\bar{q}} \le w_r,\ i = 1, \cdots, n-1 \right\}.$$

The Lagrangian Dual of CGCC

Let $\mathbf{Z} = \mathbf{C}\mathbf{A} \in \mathbb{R}^{m\times p}$, then the Lagrangian is

$$L(\mathbf{Z}, \mathbf{A}; \lambda\boldsymbol{\Phi}) = \frac{1}{2}\|\mathbf{A} - \mathbf{X}\|_F^2 + \lambda\|\widetilde{\mathbf{W}}\mathbf{Z}\|_{1,q} + \langle \lambda\boldsymbol{\Phi}, \widetilde{\mathbf{C}}\mathbf{A} - \mathbf{Z}\rangle$$
$$= \frac{1}{2}\|\mathbf{A} - \mathbf{X}\|_F^2 + \langle \lambda\boldsymbol{\Phi}, \widetilde{\mathbf{C}}\boldsymbol{\alpha}\rangle - \langle \lambda\boldsymbol{\Phi}, \mathbf{Z}\rangle + \lambda\|\widetilde{\mathbf{W}}\mathbf{Z}\|_{1,q}, \quad (22)$$

where $\lambda\boldsymbol{\Phi} \in \mathbb{R}^{m\times p}$ is the Lagrangian multiplier. We solve the following subproblems:

$$\min_{\mathbf{A}} f_1(\mathbf{A}) = \frac{1}{2}\|\mathbf{A} - \mathbf{X}\|_F^2 + \langle \lambda\boldsymbol{\Phi}, \widetilde{\mathbf{C}}\mathbf{A}\rangle,$$
$$\min_{\mathbf{Z}} f_2(\mathbf{Z}) = -\langle \lambda\boldsymbol{\Phi}, \mathbf{Z}\rangle + \lambda\|\widetilde{\mathbf{W}}\mathbf{Z}\|_{1,q}.$$

By setting $\frac{\partial f_1(\mathbf{A})}{\partial \mathbf{A}} = 0$, we obtain

$$\mathbf{A}^* = \mathbf{X} - \lambda\widetilde{\mathbf{C}}^T\boldsymbol{\Phi}.$$

Therefore,

$$\min_{\mathbf{A}} f_1(\mathbf{A}) = \frac{\lambda^2}{2}\|\widetilde{\mathbf{C}}^T\boldsymbol{\Phi}\|_F^2 + \langle \lambda\boldsymbol{\Phi}, \widetilde{\mathbf{C}}(\mathbf{X} - \lambda\widetilde{\mathbf{C}}^T\boldsymbol{\Phi})\rangle$$
$$= \lambda\langle \boldsymbol{\Phi}, \widetilde{\mathbf{C}}\mathbf{X}\rangle - \frac{\lambda^2}{2}\|\widetilde{\mathbf{C}}^T\boldsymbol{\Phi}\|_F^2.$$

Consider $f_2(\mathbf{Z})$. Note that $f_2(\mathbf{Z})$ can be decomposed into m subproblems corresponding to the rows of $\mathbf{Z}$:

$$\min_{\mathbf{z}_r} f_2(\mathbf{z}_r) = -\langle \lambda\boldsymbol{\phi}_r, \mathbf{z}_r\rangle + \lambda\widetilde{w}_r\|\mathbf{z}_r\|_q,\ r = 1, \cdots, m.$$

Now, the optimality condition is

$$\mathbf{0} \in \partial f_2(\mathbf{z}_r) = -\lambda\boldsymbol{\phi}_r + \lambda\widetilde{w}_r\partial\|\mathbf{z}_r\|_q,$$

which leads to

$$\boldsymbol{\phi}_r = \mathbf{v}_r,\ \mathbf{v}_r \in \widetilde{w}_r\partial\|\mathbf{z}_r\|_q. \quad (23)$$

By noting that

$$\langle \mathbf{v}_r, \mathbf{z}_r\rangle = \widetilde{w}_r\|\mathbf{z}_r\|_q,$$

we have

$$\langle \boldsymbol{\phi}_r, \mathbf{z}_r\rangle = \widetilde{w}_r\partial\|\mathbf{z}_r\|_q.$$

Thus we can see that

$$0 = \min_{\mathbf{z}_r} f_2(\mathbf{z}_r).$$

Moreover, Eq. (23) implies that

$$\boldsymbol{\phi}_r \in \widetilde{w}_r\mathcal{B}_{\bar{q}},$$

Combining all the derivations above, the dual problem of the CGCC problem in (1) is

$$\min_{\boldsymbol{\Phi}}\left\{ g(\boldsymbol{\Phi}) = \frac{\lambda^2}{2}\left\|\widetilde{\mathbf{C}}^T\boldsymbol{\Phi} - \frac{\mathbf{X}}{\lambda}\right\|_F^2 - \|\mathbf{X}\|_F^2 : \|\boldsymbol{\phi}_r\|_{\bar{q}} \le 1,\ r = 1, \cdots, m \right\}.$$

Proof of Lemma 1

We prove the statement by contradiction. Since each row of $\mathbf{C}$ denotes one edge in $\mathbf{E}_{\mathcal{T}}$, for the rth row $\mathbf{c}_r$ according to the edge (i,j) in $\mathbf{E}_{\mathcal{T}}$, we represent $r \leftrightarrow (i,j)$. Assume that $\mathbf{C}$ is rank-deficient, then there exist at least one row $\mathbf{c}_{r_k}$ such that $\mathbf{c}_{r_k}$ can be linearly represented by the residual $n-2$ rows, i.e., $\mathbf{c}_{r_k} = \sum_{k'\neq k} a_{k'}\mathbf{c}_{r_{k'}}$, where $a_{k'}$'s are scalars. Then we must have $i_k \in \{i_{k'}\}_{k'\neq k} \cup \{j_{k'}\}_{k'\neq k}$ and $j_k \in \{i_{k'}\}_{k'\neq k} \cup \{j_{k'}\}_{k'\neq k}$, where $r_k \leftrightarrow (i_k, j_k)$, because there are only two non-zero elements at the ith and jth position for one row $\mathbf{c}_r$. In other words, node i_k and node j_k are connected via another path instead of (i_k, j_k), therefore, there exists a ring in $\boldsymbol{\mathcal{T}}$, which contradicts the fact that $\boldsymbol{\mathcal{T}}$ is a tree. □

Proof of Lemma 2

From the definition of $\mathbf{C}$, it is easy to verify that $\mathbf{1}_n$ is orthogonal to all the rows in $\mathbf{C}$, therefore $rank(\mathbf{D}) = n$ and $\mathbf{D}$ is invertible. $\square$

Proof of Theorem 1

As mentioned previously, (i)⇔(ii) is obvious. Then we show (ii)⇔(iii). We first prove (ii)⇒(iii). Assume (ii) is satisfied, then from the KKT condition in Eq. (7), we have $\widetilde{\mathbf{D}}\mathbf{B}^* = \mathbf{0}$. Assume $\mathbf{B}^* \neq 0$, it must be that the value of the objective function in problem (5) satisfies $h(\mathbf{0}) > h(\mathbf{B}^*)$, i.e. $\frac{1}{2}\|\widetilde{\mathbf{X}}\|_F^2 > \frac{1}{2}\|\widetilde{\mathbf{X}}\|_F^2 + \lambda\|\mathbf{W}\mathbf{B}^*\|_{1,q}$, which is impossible. Therefore, we have $\mathbf{B}^* = \mathbf{0}$. The converse (iii)⇒(ii) can be easily verified. $\square$

Useful Lemmas

Lemma 4. *(Ruszczyński 2006; Bauschke and Combettes 2011) For a closed convex set $\boldsymbol{\mathcal{S}} \in \mathbb{R}^{n\times p}$ and a point $\mathbf{u} \in \boldsymbol{\mathcal{S}}$, the normal cone to $\boldsymbol{\mathcal{S}}$ at $\mathbf{u}$ is defined by*

$$\boldsymbol{\mathcal{N}}_{\boldsymbol{\mathcal{S}}}(\mathbf{u}) = \{\mathbf{v} : \langle \mathbf{v}, \mathbf{u}' - \mathbf{u}\rangle \leq 0, \forall \mathbf{u}' \in \boldsymbol{\mathcal{S}}\}. \tag{24}$$

Denote by

$$\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}}(\mathbf{u}) = \arg\min_{\mathbf{u}'\in\boldsymbol{\mathcal{S}}} \|\mathbf{u} - \mathbf{u}'\|_F.$$

Then, the following statements hold: (i) $\boldsymbol{\mathcal{N}}_{\boldsymbol{\mathcal{S}}}(\mathbf{u}) = \{\mathbf{v} : \boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}}(\mathbf{u}+\mathbf{v}) = \mathbf{u}\}$; (ii) $\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}}(\mathbf{u}+\mathbf{v}) = \mathbf{u}$, $\forall \mathbf{v} \in \boldsymbol{\mathcal{N}}_{\boldsymbol{\mathcal{S}}}(\mathbf{u})$; (iii) For $\overline{\mathbf{u}} \notin \boldsymbol{\mathcal{S}}$, $\mathbf{u} = \boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}}(\overline{\mathbf{u}}) \Leftrightarrow \overline{\mathbf{u}} - \mathbf{u} \in \boldsymbol{\mathcal{N}}_{\boldsymbol{\mathcal{S}}}(\mathbf{u})$.

Lemma 5. *(Nesterov 2004) For any convex constrained optimization problem:*

$$\min_{\mathbf{X}\in\boldsymbol{\mathcal{S}}} f(\mathbf{X}), \tag{25}$$

where $\boldsymbol{\mathcal{S}}$ is convex and closed set and $f(\cdot)$ is convex and differentiable. $\mathbf{X}^* \in \boldsymbol{\mathcal{S}}$ is an optimal solution of Eq. (25) if and only if

$$-f'(\mathbf{X}^*) \in \boldsymbol{\mathcal{N}}_{\boldsymbol{\mathcal{S}}}(\mathbf{X}^*). \tag{26}$$

Lemma 6. *Let $\mathbf{n}(\lambda')$ be defined in Theorem 2 for any $\lambda' < \lambda_{\max}$ and $\bar{q} \in \{1, 2, \infty\}$, we have $\mathbf{n}(\lambda') \in \boldsymbol{\mathcal{N}}_{\boldsymbol{\mathcal{F}}}(\boldsymbol{\Theta}^*(\lambda'))$.*

Proof: We prove the case when $q = \bar{q} = 2$, and other cases can be proved in a similar way. We first discuss the condition that $\lambda' < \lambda_{max}$. When $\lambda' < \lambda_{max}$, from Theorem 1 we know $\frac{\widetilde{\mathbf{X}}}{\lambda'} \notin \boldsymbol{\mathcal{F}}$. Therefore,

$$\frac{\widetilde{\mathbf{X}}}{\lambda'} - \boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{F}}}\left(\frac{\widetilde{\mathbf{X}}}{\lambda'}\right) = \frac{\widetilde{\mathbf{X}}}{\lambda'} - \boldsymbol{\Theta}^*(\lambda') \neq \mathbf{0}.$$

From condition (iii) in Lemma 4, we have

$$\frac{\widetilde{\mathbf{X}}}{\lambda'} - \boldsymbol{\Theta}^*(\lambda') \in \boldsymbol{\mathcal{N}}_{\boldsymbol{\mathcal{F}}}(\boldsymbol{\Theta}^*(\lambda')).$$

Next, we consider $\lambda' = \lambda_{\max}$. From Theorem 1, we have $\boldsymbol{\Theta}^*(\lambda') = \frac{\widetilde{\mathbf{X}}}{\lambda'} \in \boldsymbol{\mathcal{F}}$. Now we have to show

$$\left\langle \mathbf{d}_*^T\mathbf{d}_*\frac{\widetilde{\mathbf{X}}}{\lambda_{\max}}, \boldsymbol{\Theta} - \frac{\widetilde{\mathbf{X}}}{\lambda_{\max}}\right\rangle \leq 0,\ \forall \boldsymbol{\Theta} \in \boldsymbol{\mathcal{F}},$$

which is equivalent to

$$\left\langle \frac{\mathbf{d}_*\widetilde{\mathbf{X}}}{\lambda_{\max}}, \mathbf{d}_*\boldsymbol{\Theta} - \frac{\mathbf{d}_*\widetilde{\mathbf{X}}}{\lambda_{\max}}\right\rangle \leq 0,\ \forall \boldsymbol{\Theta} \in \boldsymbol{\mathcal{F}}.$$

From the definition of $\mathbf{d}_*$, we have

$$\left\|\frac{\mathbf{d}_*\widetilde{\mathbf{X}}}{\lambda_{\max}}\right\|_2 = w_* = \max\{w_r\}_{r=1}^{n-1}.$$

Recall the definition of $\boldsymbol{\mathcal{F}}$, where

$$\boldsymbol{\mathcal{F}}_r = \left\{\boldsymbol{\Theta} : \left\|\widetilde{\mathbf{d}}_r\boldsymbol{\Theta}\right\|_{\bar{q}} \leq w_r\right\},\ r = 1, \cdots, n-1.$$

Then we have

$$\begin{aligned}\left\langle \frac{\mathbf{d}_*\widetilde{\mathbf{X}}}{\lambda_{\max}}, \mathbf{d}_*\boldsymbol{\Theta} - \frac{\mathbf{d}_*\widetilde{\mathbf{X}}}{\lambda_{\max}}\right\rangle &= \left\langle \frac{\mathbf{d}_*\widetilde{\mathbf{X}}}{\lambda_{\max}}, \mathbf{d}_*\boldsymbol{\Theta}\right\rangle - w_*^2 \\ &\leq \left\|\frac{\mathbf{d}_*\widetilde{\mathbf{X}}}{\lambda_{\max}}\right\|_2 \|\mathbf{d}_*\boldsymbol{\Theta}\|_2 - w_*^2 \\ &\leq 0,\end{aligned}$$

which completes the proof. $\square$

Proof of Theorem 2

We prove the case when $q = \bar{q} = 2$, and other cases can be proved in a similar way. The statement in Eq. (9) is equivalent to

$$\left\|\boldsymbol{\Theta}^*(\lambda) - \boldsymbol{\Theta}^*(\lambda')\right\|_F^2 \leq \left\langle \boldsymbol{\Theta}^*(\lambda) - \boldsymbol{\Theta}^*(\lambda'), \mathbf{v}^\perp(\lambda, \lambda')\right\rangle. \tag{27}$$

We will show Eq. (27) holds. From Lemma 4 and Lemma 6, we have

$$\langle \mathbf{n}(\lambda'), \boldsymbol{\Theta} - \boldsymbol{\Theta}^*(\lambda')\rangle \leq 0,\ \forall \boldsymbol{\Theta} \in \boldsymbol{\mathcal{F}}. \tag{28}$$

By letting $\boldsymbol{\Theta} = \boldsymbol{\Theta}^*(\lambda)$ and $\boldsymbol{\Theta} = \mathbf{0}$, we can obtain the following results respectively:

$$\langle \mathbf{n}(\lambda'), \boldsymbol{\Theta}^*(\lambda) - \boldsymbol{\Theta}^*(\lambda')\rangle \leq 0, \tag{29}$$

$$\begin{cases} \left\langle \mathbf{n}(\lambda'), \widetilde{\mathbf{X}}\right\rangle \geq 0, & \text{if } \lambda' = \lambda_{\max}, \\ \frac{\|\widetilde{\mathbf{X}}\|_F}{\lambda'} \geq \|\boldsymbol{\Theta}^*(\lambda')\|_F, & \text{if } \lambda' < \lambda_{\max}. \end{cases} \tag{30}$$

Moreover, from Lemma 6, we also have

$$\frac{\widetilde{\mathbf{X}}}{\lambda} - \boldsymbol{\Theta}^*(\lambda) \in \boldsymbol{\mathcal{N}}_{\boldsymbol{\mathcal{F}}}(\boldsymbol{\Theta}^*(\lambda)).$$

Then, we have

$$\left\langle \frac{\widetilde{\mathbf{X}}}{\lambda} - \boldsymbol{\Theta}^*(\lambda), \boldsymbol{\Theta}^*(\lambda') - \boldsymbol{\Theta}^*(\lambda)\right\rangle \leq 0. \tag{31}$$

Eq. (31) is equivalent to

$$\left\|\boldsymbol{\Theta}^*(\lambda) - \boldsymbol{\Theta}^*(\lambda')\right\|_F^2 \leq \langle \boldsymbol{\Theta}^*(\lambda) - \boldsymbol{\Theta}^*(\lambda'), \mathbf{v}(\lambda, \lambda')\rangle. \tag{32}$$

Comparing Eq. (32) with Eq. (27), we consider Eq. (27) again:

$$\begin{aligned}&\left\langle \boldsymbol{\Theta}^*(\lambda) - \boldsymbol{\Theta}^*(\lambda'), \mathbf{v}^\perp(\lambda, \lambda')\right\rangle \\ &= \langle \boldsymbol{\Theta}^*(\lambda) - \boldsymbol{\Theta}^*(\lambda'), \mathbf{v}(\lambda, \lambda')\rangle \\ &- \left\langle \boldsymbol{\Theta}^*(\lambda) - \boldsymbol{\Theta}^*(\lambda'), \mathbf{v}(\lambda, \lambda') - \mathbf{v}^\perp(\lambda, \lambda')\right\rangle \\ &= \langle \boldsymbol{\Theta}^*(\lambda) - \boldsymbol{\Theta}^*(\lambda'), \mathbf{v}(\lambda, \lambda')\rangle \\ &- \left\langle \boldsymbol{\Theta}^*(\lambda) - \boldsymbol{\Theta}^*(\lambda'), \frac{\langle \mathbf{v}(\lambda, \lambda'), \mathbf{n}(\lambda')\rangle}{\|\mathbf{n}(\lambda')\|_F^2}\mathbf{n}(\lambda')\right\rangle\end{aligned} \tag{33}$$

Based on Eq. (33), recall Eq. (29) and we know that if $\langle \mathbf{v}(\lambda,\lambda'), \mathbf{n}(\lambda')\rangle \geq 0$, we can prove the theorem. Actually, we have

$$\begin{aligned}\langle \mathbf{v}(\lambda,\lambda'), \mathbf{n}(\lambda')\rangle &= \left\langle \frac{\widetilde{\mathbf{X}}}{\lambda} - \mathbf{\Theta}^*(\lambda'), \mathbf{n}(\lambda') \right\rangle \\ &= \left(\frac{1}{\lambda} - \frac{1}{\lambda'}\right)\left\langle \widetilde{\mathbf{X}}, \mathbf{n}(\lambda')\right\rangle + \left\langle \frac{\widetilde{\mathbf{X}}}{\lambda'} - \mathbf{\Theta}^*(\lambda'), \mathbf{n}(\lambda')\right\rangle.\end{aligned}$$

If $\lambda' = \lambda_{\max}$, recall the first statement in Eq. (30) and $\lambda < \lambda'$, it is easy to see that

$$\left(\frac{1}{\lambda} - \frac{1}{\lambda'}\right)\left\langle \widetilde{\mathbf{X}}, \mathbf{n}(\lambda')\right\rangle \geq 0.$$

If $\lambda' < \lambda_{\max}$, from the second statement in Eq. (30), we also have

$$\begin{aligned}\left\langle \widetilde{\mathbf{X}}, \mathbf{n}(\lambda')\right\rangle &= \left\langle \widetilde{\mathbf{X}}, \frac{\widetilde{\mathbf{X}}}{\lambda'} - \mathbf{\Theta}^*(\lambda')\right\rangle \\ &\geq \frac{\|\widetilde{\mathbf{X}}\|_F^2}{\lambda'} - \|\widetilde{\mathbf{X}}\|_F \|\mathbf{\Theta}^*(\lambda')\|_F \\ &\geq 0.\end{aligned}$$

Now the last thing is to show that

$$\left\langle \frac{\widetilde{\mathbf{X}}}{\lambda'} - \mathbf{\Theta}^*(\lambda'), \mathbf{n}(\lambda')\right\rangle \geq 0. \tag{34}$$

Eq. (34) is obvious, since

$$\left\langle \frac{\widetilde{\mathbf{X}}}{\lambda'} - \mathbf{\Theta}^*(\lambda'), \mathbf{n}(\lambda')\right\rangle = \begin{cases} 0, & \text{if } \lambda' = \lambda_{\max}, \\ \|\mathbf{n}(\lambda')\|_F^2, & \text{if } \lambda' < \lambda_{\max}. \end{cases}$$

Finally, we reach the conclusion. □

Proof of Theorem 3

From the feasible region $\mathcal{O}$ in Eq. (10), for any $\mathbf{\Theta} \in \mathcal{O}$ we can write

$$\mathbf{\Theta} = \mathbf{o}(\lambda,\lambda') + \mathbf{\Upsilon},\ \|\mathbf{\Upsilon}\|_F \leq R(\lambda,\lambda').$$

Therefore, if $\bar{q} = 1$, i.e. $q = \infty$, we have

$$\begin{aligned}\sup_{\mathbf{\Theta}\in\mathcal{O}} \left\|\widetilde{\mathbf{d}}_r \mathbf{\Theta}^*\right\|_1 &= \sup_{\|\mathbf{\Upsilon}\|_F \leq R(\lambda,\lambda')} \left\|\widetilde{\mathbf{d}}_r \mathbf{o}(\lambda,\lambda') + \widetilde{\mathbf{d}}_r \mathbf{\Upsilon}\right\|_1 \\ &= \left\|\widetilde{\mathbf{d}}_r \mathbf{o}(\lambda,\lambda')\right\|_1 + R(\lambda,\lambda')\|\widetilde{\mathbf{d}}_r\|_2,\end{aligned}$$

else if $\bar{q} = 2$, i.e. $q = 2$, we have

$$\begin{aligned}\sup_{\mathbf{\Theta}\in\mathcal{O}} \left\|\widetilde{\mathbf{d}}_r \mathbf{\Theta}^*\right\|_2 &= \sup_{\|\mathbf{\Upsilon}\|_F \leq R(\lambda,\lambda')} \left\|\widetilde{\mathbf{d}}_r \mathbf{o}(\lambda,\lambda') + \widetilde{\mathbf{d}}_r \mathbf{\Upsilon}\right\|_2 \\ &= \left\|\widetilde{\mathbf{d}}_r \mathbf{o}(\lambda,\lambda')\right\|_2 + R(\lambda,\lambda')\|\widetilde{\mathbf{d}}_r\|_2,\end{aligned}$$

else if $\bar{q} = \infty$, i.e. $q = 1$, we have

$$\begin{aligned}\sup_{\mathbf{\Theta}\in\mathcal{O}} \left\|\widetilde{\mathbf{d}}_r \mathbf{\Theta}^*\right\|_\infty &= \sup_{\|\mathbf{\Upsilon}\|_F \leq R(\lambda,\lambda')} \left\|\widetilde{\mathbf{d}}_r \mathbf{o}(\lambda,\lambda') + \widetilde{\mathbf{d}}_r \mathbf{\Upsilon}\right\|_\infty \\ &= \left\|\widetilde{\mathbf{d}}_r \mathbf{o}(\lambda,\lambda')\right\|_\infty + R(\lambda,\lambda')\|\widetilde{\mathbf{d}}_r\|_2,\end{aligned}$$

where the supreme values can be obtained directly by Cauchy inequality and norm inequalities. From these supreme values, we can directly reach the conclusion. □

Proof of Lemma 3

Because the graph $\mathbf{G}$ in CGCC model is a cyclic graph, there exists at least one ring in $\mathbf{E}_\mathcal{C}$. Recall that we require $\mathbf{G}$ to be connected graph. Therefore, if there exist rings in $\mathbf{E}_\mathcal{C}$, we must have that the cardinality $|\mathbf{E}_\mathcal{C}| = m \geq n > n-1$.

Moreover, from the proof of Lemma 1, we know that if a ring exists in $\mathbf{E}_\mathcal{C}$, there must exist one row of $\overline{\mathbf{C}}$ that can be linearly represented by some other rows. For any $m \geq n$, there exists at least one ring in $\mathbf{E}_\mathcal{C}$, therefore, $\overline{\mathbf{C}} < n$ is rank-deficient. □

Proof of Theorem 4

The proofs can be completed by following those of Theorem 1. □

Proof of Theorem 5

We prove the case when $q = \bar{q} = 2$, and other cases can be proved in a similar way. When $\lambda' < \overline{\lambda}_{\max}$, Eq. (19) is equivalent to

$$\left\|\mathbf{\Lambda}\overline{\mathbf{\Phi}}^*(\lambda) - \mathbf{\Lambda}\overline{\mathbf{\Phi}}^*(\lambda')\right\|_F^2 \leq \left\langle \mathbf{\Lambda}\overline{\mathbf{\Phi}}^*(\lambda) - \mathbf{\Lambda}\overline{\mathbf{\Phi}}^*(\lambda'), \overline{\mathbf{v}}^\perp(\lambda,\lambda')\right\rangle. \tag{35}$$

We will show Eq. (35) holds. Note that when $\lambda' < \overline{\lambda}_{\max}$, $\mathbf{\Lambda}\overline{\mathbf{n}}(\lambda') = -\overline{h}'(\mathbf{\Phi}^*)$. From Lemma 5, we have

$$\left\langle \overline{\mathbf{n}}(\lambda'), \mathbf{\Lambda}\overline{\mathbf{\Phi}} - \mathbf{\Lambda}\overline{\mathbf{\Phi}}^*(\lambda')\right\rangle \leq 0,\ \forall \overline{\mathbf{\Phi}} \in \overline{\mathcal{F}}. \tag{36}$$

By letting $\overline{\mathbf{\Phi}} = \overline{\mathbf{\Phi}}^*(\lambda)$ and $\overline{\mathbf{\Phi}} = \mathbf{0}$, we can obtain the following results respectively:

$$\left\langle \overline{\mathbf{n}}(\lambda'), \mathbf{\Lambda}\overline{\mathbf{\Phi}}^*(\lambda) - \mathbf{\Lambda}\overline{\mathbf{\Phi}}^*(\lambda')\right\rangle \leq 0, \tag{37}$$

$$\frac{\left\|\mathbf{\Lambda}^{-1}\overline{\mathbf{Y}}\right\|_F}{\lambda'} \geq \|\mathbf{\Lambda}\overline{\mathbf{\Phi}}^*(\lambda')\|_F, \tag{38}$$

Moreover, from Lemma 5, we also have

$$-\overline{h}'(\overline{\mathbf{\Phi}}^*(\lambda)) = \frac{\overline{\mathbf{Y}}}{\lambda} - \overline{\mathbf{D}}\overline{\mathbf{\Phi}}^*(\lambda) \in \mathcal{N}_{\overline{\mathcal{F}}}(\overline{\mathbf{\Phi}}^*(\lambda)).$$

Then, we have

$$\left\langle \frac{\overline{\mathbf{Y}}}{\lambda} - \overline{\mathbf{D}}\overline{\mathbf{\Phi}}^*(\lambda), \overline{\mathbf{\Phi}}^*(\lambda') - \overline{\mathbf{\Phi}}^*(\lambda)\right\rangle \leq 0. \tag{39}$$

Eq. (39) is equivalent to

$$\left\|\mathbf{\Lambda}\overline{\mathbf{\Phi}}^*(\lambda) - \mathbf{\Lambda}\overline{\mathbf{\Phi}}^*(\lambda')\right\|_F^2 \leq \left\langle \mathbf{\Lambda}\overline{\mathbf{\Phi}}^*(\lambda) - \mathbf{\Lambda}\overline{\mathbf{\Phi}}^*(\lambda'), \overline{\mathbf{v}}(\lambda,\lambda')\right\rangle. \tag{40}$$

Comparing Eq. (40) with Eq. (35), we consider Eq. (35) again:

$$\begin{aligned}&\left\langle \mathbf{\Lambda}\overline{\mathbf{\Phi}}^*(\lambda) - \mathbf{\Lambda}\overline{\mathbf{\Phi}}^*(\lambda'), \overline{\mathbf{v}}^\perp(\lambda,\lambda')\right\rangle \\ &= \left\langle \mathbf{\Lambda}\overline{\mathbf{\Phi}}^*(\lambda) - \mathbf{\Lambda}\overline{\mathbf{\Phi}}^*(\lambda'), \overline{\mathbf{v}}(\lambda,\lambda')\right\rangle \\ &- \left\langle \mathbf{\Lambda}\overline{\mathbf{\Phi}}^*(\lambda) - \mathbf{\Lambda}\overline{\mathbf{\Phi}}^*(\lambda'), \overline{\mathbf{v}}(\lambda,\lambda') - \overline{\mathbf{v}}^\perp(\lambda,\lambda')\right\rangle \\ &= \left\langle \mathbf{\Lambda}\overline{\mathbf{\Phi}}^*(\lambda) - \mathbf{\Lambda}\overline{\mathbf{\Phi}}^*(\lambda'), \overline{\mathbf{v}}(\lambda,\lambda')\right\rangle \\ &- \left\langle \mathbf{\Lambda}\overline{\mathbf{\Phi}}^*(\lambda) - \mathbf{\Lambda}\overline{\mathbf{\Phi}}^*(\lambda'), \frac{\langle \overline{\mathbf{v}}(\lambda,\lambda'), \overline{\mathbf{n}}(\lambda')\rangle}{\|\overline{\mathbf{n}}(\lambda')\|_F^2}\overline{\mathbf{n}}(\lambda')\right\rangle\end{aligned} \tag{41}$$

Based on Eq. (41), recall Eq. (37) and we know that if $\langle \overline{\mathbf{v}}(\lambda,\lambda'), \overline{\mathbf{n}}(\lambda')\rangle \geq 0$, we can prove the theorem. Actually, we have

$$\langle \overline{\mathbf{v}}(\lambda,\lambda'), \overline{\mathbf{n}}(\lambda')\rangle = \left\langle \frac{\mathbf{\Lambda}^{-1}\overline{\mathbf{Y}}}{\lambda} - \mathbf{\Lambda}\overline{\mathbf{\Phi}}^*(\lambda'), \overline{\mathbf{n}}(\lambda') \right\rangle$$
$$= \left(\frac{1}{\lambda} - \frac{1}{\lambda'}\right)\langle \mathbf{\Lambda}^{-1}\overline{\mathbf{Y}}, \overline{\mathbf{n}}(\lambda')\rangle + \left\langle \frac{\mathbf{\Lambda}^{-1}\overline{\mathbf{Y}}}{\lambda'} - \mathbf{\Lambda}\overline{\mathbf{\Phi}}^*(\lambda'), \overline{\mathbf{n}}(\lambda') \right\rangle.$$

From Eq. (38), we have

$$\begin{aligned}\langle \mathbf{\Lambda}^{-1}\overline{\mathbf{Y}}, \overline{\mathbf{n}}(\lambda')\rangle &= \left\langle \mathbf{\Lambda}^{-1}\overline{\mathbf{Y}}, \frac{\mathbf{\Lambda}^{-1}\overline{\mathbf{Y}}}{\lambda'} - \mathbf{\Lambda}\overline{\mathbf{\Phi}}^*(\lambda') \right\rangle \\ &\geq \frac{\|\mathbf{\Lambda}^{-1}\overline{\mathbf{Y}}\|_F^2}{\lambda'} - \|\mathbf{\Lambda}^{-1}\overline{\mathbf{Y}}\|_F \|\mathbf{\Lambda}\overline{\mathbf{\Phi}}^*(\lambda')\|_F \\ &\geq 0.\end{aligned}$$

Now the last thing is to show that

$$\left\langle \frac{\mathbf{\Lambda}^{-1}\overline{\mathbf{Y}}}{\lambda'} - \mathbf{\Lambda}\overline{\mathbf{\Phi}}^*(\lambda'), \overline{\mathbf{n}}(\lambda') \right\rangle \geq 0. \quad (42)$$

Eq. (42) is obvious, since

$$\left\langle \frac{\mathbf{\Lambda}^{-1}\overline{\mathbf{Y}}}{\lambda'} - \mathbf{\Lambda}\overline{\mathbf{\Phi}}^*(\lambda'), \overline{\mathbf{n}}(\lambda') \right\rangle = \|\overline{\mathbf{n}}(\lambda')\|_F^2 \geq 0.$$

Now we complete the proof. □

Details for Definition 1

Let $\overline{\mathcal{O}}$ be the feasible region constructed from Eqs. (19):

$$\overline{\mathcal{O}}(\lambda,\lambda') = \left\{\overline{\mathbf{\Phi}}(\lambda) : \|\mathbf{\Lambda}\overline{\mathbf{\Phi}}(\lambda) - \overline{\mathbf{o}}(\lambda,\lambda')\|_F \leq \overline{R}(\lambda,\lambda')\right\}. \quad (43)$$

We have to estimate the following supreme values:

$$\sup_{\overline{\mathbf{\Phi}}} \left\{\|\overline{\phi}_r\|_{\bar{q}} : \overline{\mathbf{\Phi}} \in \overline{\mathcal{O}},\ r = 1,\cdots,m\right\}, \quad (44)$$

Let $\mathbf{\Xi} = \mathbf{\Lambda}\overline{\mathbf{\Phi}}$. The problem becomes

$$\sup_{\mathbf{\Xi}} \left\{\|\boldsymbol{\zeta}_r \mathbf{\Xi}\|_{\bar{q}} : \mathbf{\Lambda}^{-1}\mathbf{\Xi} \in \overline{\mathcal{O}},\ r = 1,\cdots,m\right\}, \quad (45)$$

where $\boldsymbol{\zeta}_r$ is the rth row of $\mathbf{\Lambda}^{-1}$. The supreme values in Eq. (45) can be obtained exactly from the proof of Theorem 3. □

Analysis for δ

The parameter δ in problem (18) plays an important role, since $\overline{\lambda}_{\max}$ depends on the value of δ. When δ is very small, computing $\overline{\lambda}_{\max}$ may be numerically unstable but when δ is large, the relaxed dual form will have a large deviation from the original one, leading to an inaccurate estimation for $\overline{\lambda}_{\max}$. However, since the CGCC problem is convex problem, there exist a certain value of $\lambda^*_{\max}$ that will guarantee all the data points are clustered. Therefore, In the implementation of the Cigar rules, we propose to empirically find the maximum value $\lambda^*_{\max}$ first, and then choose an appropriate δ such that the induced $\overline{\lambda}_{\max}$ satisfes that $\overline{\lambda}_{\max}$ is close to $\lambda^*_{\max}$ but $\lambda^*_{\max} < \overline{\lambda}_{\max}$, which makes the condition in Theorem 1 satisfied. Empirically, we can assign a relatively large initial value to δ which will induce a small $\overline{\lambda}_{\max}$ and then decrease δ gradually until $\lambda^*_{\max} < \overline{\lambda}_{\max}$ is satisfied.

Efficient Ways for Computing The Matrices Used in The Cigar Rule

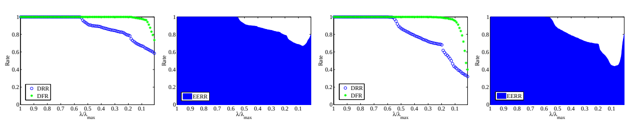
When the number of rows in $\overline{\mathbf{C}}$ is large, e.g. the CC problem with fully connected graph $m = \frac{n(n-1)}{2}$, calculating the inverse of $\overline{\mathbf{D}} \in \mathbb{R}^{m\times m}$ directly is infeasible. However, from the definition of $\overline{\mathbf{D}}$, we have the following efficient way to compute $\overline{\mathbf{D}}^{-1}$:

$$\begin{aligned}\overline{\mathbf{D}}^{-1} &= \left(\overline{\mathbf{C}}\,\overline{\mathbf{C}}^T + \delta\mathbf{I}\right)^{-1} = \frac{1}{\delta}\left(\frac{\overline{\mathbf{C}}}{\sqrt{\delta}}\frac{\overline{\mathbf{C}}^T}{\sqrt{\delta}} + \mathbf{I}\right)^{-1} \\ &= \frac{1}{\delta}\left(\mathbf{I} - \frac{\overline{\mathbf{C}}}{\sqrt{\delta}}\left(\mathbf{I} + \frac{\overline{\mathbf{C}}^T}{\sqrt{\delta}}\frac{\overline{\mathbf{C}}}{\sqrt{\delta}}\right)^{-1}\frac{\overline{\mathbf{C}}^T}{\sqrt{\delta}}\right)\end{aligned}$$

where $\overline{\mathbf{C}}^T\overline{\mathbf{C}} \in \mathbb{R}^{n\times n}$ and the matrix inverse can be completed efficiently. The square root matrices $\mathbf{\Lambda}$ and $\mathbf{\Lambda}^{-1}$ can be obtained by the eigen-decomposition technique from matrices $\overline{\mathbf{D}}$ or $\overline{\mathbf{D}}^{-1}$.

Additional Experimental Results

Fig. 5 shows the ℓ_2 clusterpath generated from the TCC and CGCC models on the two synthetic datasets when $n = 200$. According to the results, we can see that all the models can correctly detect the cluster. Figs. 6 and 7, Tables 4 and 5 provide the additional experimental results on the synthetic data where $n = 1000$.



(a) Halfmoon (b) Halfmoon (c) Spiral (d) Spiral

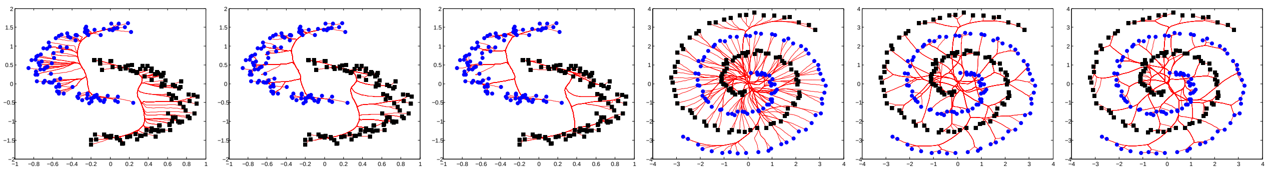
Figure 6: The performance of the Eater rule on synthetic data (n=1000).

Table 4: Running Time (seconds). E+S denotes the total time cost of using the Eater rule and the solver.

Data	Solver	Eater	E+S	Speedup
halfmoon (n=1000)	1495.6	4.1	288.7	**5.2**
spiral (n=1000)	3563.5	4.3	804.1	**4.4**

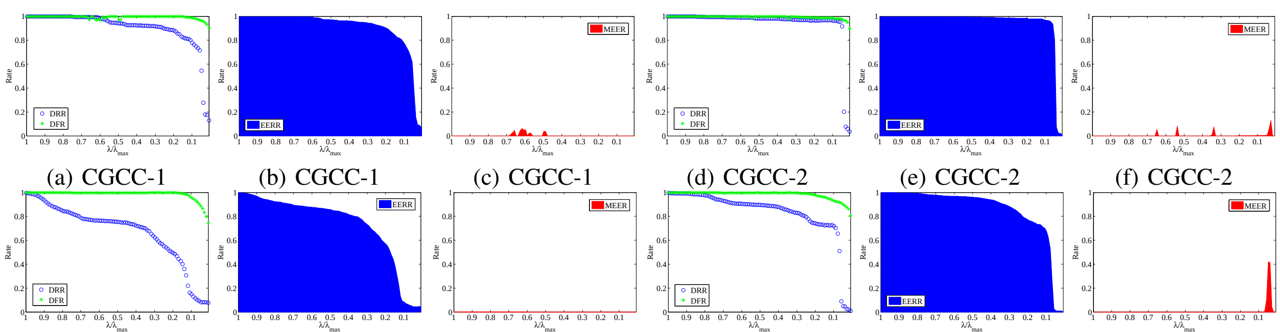
Table 5: Running Time (seconds). C+S means the total time cost of using the Cigar rule and the solver.

Data	Model	Solver	Cigar	C+S	Speedup
Halfmoon	CGCC-1	3867.8	149.7	722.7	**5.4**
(n=1000)	CGCC-2	4709.7	301.1	1028.4	**4.6**
Spiral	CGCC-1	10314.8	164.3	2349.8	**4.4**
(n=1000)	CGCC-2	12735.2	338.4	3178.4	**4.0**



(a) TCC (b) CGCC-1 (c) CGCC-2 (d) TCC (e) CGCC-1 (f) CGCC-2

Figure 5: ℓ_2 clusterpath generated by the GCC models on the synthetic datasets. (a)-(c): halfmoon data (n=200); (d)-(f) spiral data (n=200).



(a) CGCC-1 (b) CGCC-1 (c) CGCC-1 (d) CGCC-2 (e) CGCC-2 (f) CGCC-2

(g) CGCC-1 (h) CGCC-1 (i) CGCC-1 (j) CGCC-2 (k) CGCC-2 (l) CGCC-2

Figure 7: The performance of the Cigar rule on synthetic data (n=1000). The first and second rows denote the results on the halfmoon and spiral dataset respectively.