Supplementary Material for ‘Reduction Techniques for Graph-based Convex Clustering’

Proofs

The Lagrangian Dual of TCC

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

\[
L(Y, B; \lambda \Theta) = \frac{1}{2} \| Y \|_F^2 + \lambda \| WB \|_{1,q} + \langle \lambda \Theta, \bar{X} - \bar{D}B - Y \rangle
\]

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

\[
\min_Y f_1(Y) = \frac{1}{2} \| Y \|_F^2 - \langle \lambda \Theta, Y \rangle,
\]

\[
\min_B f_2(B) = -\langle \lambda \Theta, DB \rangle + \lambda \| WB \|_{1,q}.
\]

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

\[
Y^* = \lambda \Theta.
\]

Therefore,

\[
\min_Y f_1(Y) = f_1(Y^*) = -\frac{\lambda^2}{2} \| \Theta - \frac{X}{\lambda} \|_F^2 + \| X \|_F^2.
\]

Consider \( f_2(B) \). Let \( \beta_r \) be the \( r \)th row of \( B \). The optimality condition is

\[
0 \in \partial f_2(\beta_r) = -\lambda [\bar{D}^T \Theta]_r + \lambda w_r \partial \| \beta_r \|_q,
\]

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

\[
\bar{d}_r \Theta = w_r v_r, \quad v_r \in \partial \| \beta_r \|_q.
\]

By noting that

\[
\langle v_r, \beta_r \rangle = \| \beta_r \|_q,
\]

we have

\[
\langle \bar{d}_r \Theta, \beta_r \rangle = w_r \| \beta_r \|_q.
\]

Thus we can see that

\[
0 = \min_{\beta_r} f_2(\beta_r).
\]

Eq. (21) implies that

\[
\bar{d}_r \Theta \in w_r B_q,
\]

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

\[
\min_{\Theta} \left\{ g(\Theta) = \frac{\lambda^2}{2} \| \Theta - \frac{X}{\lambda} \|_F^2 - \| X \|_F^2 : \| \bar{d}_r \Theta \|_q \leq w_r, \ i = 1, \cdots, n - 1 \right\}.
\]

The Lagrangian Dual of CGCC

Let \( Z = CA \in \mathbb{R}^{m \times p} \), then the Lagrangian is

\[
L(Z, A; \lambda \Phi) = \frac{1}{2} \| A - X \|_F^2 + \lambda \| \tilde{W}Z \|_{1,q} + \langle \lambda \Phi, \tilde{C}A - Z \rangle
\]

\[
= \frac{1}{2} \| A - X \|_F^2 + \langle \lambda \Phi, \tilde{C}A \rangle - \langle \lambda \Phi, Z \rangle + \lambda \| \tilde{W}Z \|_{1,q}.
\]

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

\[
\min_A f_1(A) = \frac{1}{2} \| A - X \|_F^2 + \langle \lambda \Phi, \tilde{C}A \rangle,
\]

\[
\min_Z f_2(Z) = -\langle \lambda \Phi, Z \rangle + \lambda \| \tilde{W}Z \|_{1,q}.
\]

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

\[
A^* = X - \tilde{C}^T \Phi.
\]

Therefore,

\[
\min_A f_1(A) = \frac{\lambda^2}{2} \| \tilde{C}^T \Phi \|_F^2 + \langle \lambda \Phi, \tilde{C}(X - \tilde{C}^T \Phi) \rangle = \lambda \Phi, \tilde{C}X - \frac{\lambda^2}{2} \| \tilde{C}^T \Phi \|_F^2.
\]

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

\[
\min_{z_r} f_2(z_r) = -\langle \lambda \phi_r, z_r \rangle + \lambda \bar{w}_r \| z_r \|_q, \ r = 1, \cdots, m.
\]

Now, the optimality condition is

\[
0 \in \partial f_2(z_r) = -\lambda \phi_r + \lambda \bar{w}_r \| z_r \|_q,
\]

which leads to

\[
\phi_r = v_r, \quad v_r \in \bar{w}_r \partial \| z_r \|_q.
\]

By noting that

\[
\langle v_r, z_r \rangle = \bar{w}_r \| z_r \|_q,
\]

we have

\[
\langle \phi_r, z_r \rangle = \bar{w}_r \| z_r \|_q.
\]

Thus we can see that

\[
0 = \min_{z_r} f_2(z_r).
\]

Moreover, Eq. (23) implies that

\[
\phi_r \in \bar{w}_r B_q.
\]

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

\[
\min_{\Phi} \left\{ g(\Phi) = \frac{\lambda^2}{2} \| \tilde{C}^T \Phi - \frac{X}{\lambda} \|_F^2 - \| X \|_F^2 : \| \phi_r \|_q \leq 1, \ r = 1, \cdots, m \right\}.
\]

Proof of Lemma 1

We prove the statement by contradiction. Since each row of \( C \) denotes one edge in \( E_T \), for the \( r \)th row \( c_r \), according to the edge \((i, j)\) in \( E_T \), we represent \( r \leftrightarrow (i, j) \). Assume that \( C \) is rank-deficient, then there exist at least one row \( c_{rh} \) such that \( c_{rh} \) can be linearly represented by the residual \( n - 2 \) rows, i.e., \( c_{rh} = \sum_{k \neq k'} a_{rk} c_{rk'} \), where \( a_{rk}'s \) 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 \( i \)th and \( j \)th position for one row \( c_r \). 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 \( T \), which contradicts the fact that \( T \) is a tree. □
Proof of Lemma 2
From the definition of $C$, it is easy to verify that $1_n$ is orthogonal to all the rows in $C$, therefore \( \text{rank}(D) = n \) and $D$ is invertible. \( \square \)

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

Useful Lemmas

Lemma 4. (Ruszczyński 2006; Bauschke and Combettes 2011) For a closed convex set $S \in \mathbb{R}^{n \times p}$ and a point $u \in S$, the normal cone to $S$ at $u$ is defined by
\[
\mathcal{N}_S(u) = \{ v : \langle v, u' - u \rangle \leq 0, \forall u' \in S \}. \tag{24}\]
Denote by
\[
\mathcal{P}_S(u) = \arg \min_{u' \in S} \| u - u' \|_F.
\]
Then, the following statements hold: (i) $\mathcal{N}_S(u) = \{ v : \mathcal{P}_S(u + v) = u \}$; (ii) $\mathcal{P}_S(u + v) = u$, $\forall v \in \mathcal{N}_S(u)$; (iii) For $\pi \notin S$, $u = \mathcal{P}_S(\pi) \iff \pi - u \in \mathcal{N}_S(u)$.

Lemma 5. (Nesterov 2004) For any convex constrained optimization problem:
\[
\min_{X \in S} f(X), \tag{25}
\]
where $S$ is closed and set and $f(\cdot)$ is convex and differentiable. $X^* \in S$ is an optimal solution to Eq. (25) if and only if
\[
f'(X^*) \in \mathcal{N}_S(X^*). \tag{26}
\]

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

Proof: We prove the case when $q = \tilde{q} = 2$, and other cases can be proved in a similar way. We first discuss the condition that $\lambda' < \lambda_{\text{max}}$. When $\lambda' < \lambda_{\text{max}}$, from Theorem 1 we know $\frac{X}{\lambda'} \notin \mathcal{F}$. Therefore,
\[
\frac{\tilde{X}}{\lambda'} - \mathcal{P}_F \left( \frac{\tilde{X}}{\lambda'} \right) = \frac{\tilde{X}}{\lambda'} - \Theta^*(\lambda') \neq 0.
\]
From condition (iii) in Lemma 4, we have
\[
\frac{\tilde{X}}{\lambda'} - \Theta^*(\lambda') \in \mathcal{N}_F(\Theta^*(\lambda')).
\]
Next, we consider $\lambda' = \lambda_{\text{max}}$. From Theorem 1, we have $\Theta^*(\lambda') = \frac{\tilde{X}}{\lambda'} \in \mathcal{F}$. Now we have to show
\[
\langle d^*_\lambda \frac{\tilde{X}}{\lambda_{\text{max}}}, \Theta - \frac{\tilde{X}}{\lambda_{\text{max}}} \rangle \leq 0, \forall \Theta \in \mathcal{F},
\]
which is equivalent to
\[
\langle d^*_\lambda \frac{\tilde{X}}{\lambda_{\text{max}}}, \Theta - \frac{\tilde{X}}{\lambda_{\text{max}}} \rangle \leq 0, \forall \Theta \in \mathcal{F}.
\]
From the definition of $d_\lambda$, we have
\[
\left\| d^*_\lambda \frac{\tilde{X}}{\lambda_{\text{max}}} \right\|_2 = w_* = \max \{ w_r \}_{r=1}^{n-1}.
\]
Recall the definition of $\mathcal{F}$, where
\[
\mathcal{F}_r = \left\{ \Theta : \left\| d^*_\lambda \Theta \right\|_q \leq w_r \right\}, \ r = 1, \cdots, n - 1.
\]
Then we have
\[
\langle d^*_\lambda \frac{\tilde{X}}{\lambda_{\text{max}}}, \Theta - \frac{\tilde{X}}{\lambda_{\text{max}}} \rangle \leq \left\| d^*_\lambda \frac{\tilde{X}}{\lambda_{\text{max}}} \right\|_2 \left\| d^*_\lambda \Theta \right\|_2 - w_*^2 \leq 0,
\]
which completes the proof. \( \square \)

Proof of Theorem 2
We prove the case when $q = \tilde{q} = 2$, and other cases can be proved in a similar way. The statement in Eq. (9) is equivalent to
\[
\left\| \Theta^*(\lambda) - \Theta^*(\lambda') \right\|_F^2 \leq \left\langle \Theta^*(\lambda) - \Theta^*(\lambda'), v^+(\lambda, \lambda') \right\rangle. \tag{27}
\]
We will show Eq. (27) holds. From Lemma 4 and Lemma 6, we have
\[
\langle n(\lambda'), \Theta - \Theta^*(\lambda') \rangle \leq 0, \forall \Theta \in \mathcal{F}. \tag{28}
\]
By letting $\Theta = \Theta^*(\lambda)$ and $\Theta = 0$, we can obtain the following results respectively:
\[
\langle n(\lambda'), \Theta^*(\lambda) - \Theta^*(\lambda') \rangle \leq 0, \tag{29}
\]
\[
\left\{ \begin{array}{l}
\langle n(\lambda'), \tilde{X} \rangle \geq 0, \quad \text{if } \lambda' = \lambda_{\text{max}}, \\
\left\| \tilde{X} \right\|_F \geq \left\| \Theta^*(\lambda') \right\|_F, \quad \text{if } \lambda' < \lambda_{\text{max}}.
\end{array} \right. \tag{30}
\]
Moreover, from Lemma 6, we also have
\[
\frac{\tilde{X}}{\lambda} - \Theta^*(\lambda') \in \mathcal{N}_F(\Theta^*(\lambda')).
\]
Then, we have
\[
\left\langle \tilde{X} - \Theta^*(\lambda') - \Theta^*(\lambda) \right\rangle \leq 0. \tag{31}
\]
Eq. (31) is equivalent to
\[
\left\| \Theta^*(\lambda) - \Theta^*(\lambda') \right\|_F^2 \leq \left\langle \Theta^*(\lambda) - \Theta^*(\lambda'), v(\lambda, \lambda') \right\rangle. \tag{32}
\]
Comparing Eq. (32) with Eq. (27), we consider Eq. (27) again:
\[
\left\langle \Theta^*(\lambda) - \Theta^*(\lambda'), v^+(\lambda, \lambda') \right\rangle = \left\langle \Theta^*(\lambda) - \Theta^*(\lambda'), v(\lambda, \lambda') \right\rangle - \left\langle \Theta^*(\lambda) - \Theta^*(\lambda'), v(\lambda, \lambda') - v^+(\lambda, \lambda') \right\rangle \tag{33}
\]
\[
= \left\langle \Theta^*(\lambda) - \Theta^*(\lambda'), v(\lambda, \lambda') \right\rangle - \left\langle \Theta^*(\lambda) - \Theta^*(\lambda'), \langle v(\lambda, \lambda'), n(\lambda') \rangle \right\rangle.
\]
Based on Eq. (33), recall Eq. (29) and we know that if \( \langle v(\lambda, \lambda'), n(\lambda') \rangle \geq 0 \), we can prove the theorem. Actually, we have
\[
\langle v(\lambda, \lambda'), n(\lambda') \rangle = \left( \frac{1}{\lambda} - \frac{1}{\lambda'} \right) \langle \tilde{X}, n(\lambda') \rangle + \left( \frac{1}{\lambda} - \Theta^*(\lambda') \right) n(\lambda')
\]

If \( \lambda' = \lambda_{\text{max}} \), recall the first statement in Eq. (30) and \( \lambda < \lambda' \), it is easy to see that
\[
\left( \frac{1}{\lambda} - \frac{1}{\lambda'} \right) \langle \tilde{X}, n(\lambda') \rangle \geq 0.
\]

If \( \lambda' < \lambda_{\text{max}} \), from the second statement in Eq. (30), we also have
\[
\langle \tilde{X}, n(\lambda') \rangle = \left\{ \begin{array}{ll}
0, & \text{if } \lambda' = \lambda_{\text{max}}, \\
\|n(\lambda')\|_F^2, & \text{if } \lambda' < \lambda_{\text{max}}.
\end{array} \right.
\]

Now the last thing is to show that
\[
\langle \tilde{X}, n(\lambda') \rangle \geq 0. \tag{34}
\]

Eq. (34) is obvious, since
\[
\left\langle \tilde{X}, n(\lambda') \right\rangle = \left\{ \begin{array}{ll}
0, & \text{if } \lambda' = \lambda_{\text{max}}, \\
\|n(\lambda')\|_F^2, & \text{if } \lambda' < \lambda_{\text{max}}.
\end{array} \right.
\]

Finally, we reach the conclusion.

**Proof of Theorem 3**

From the feasible region \( \mathcal{O} \) in Eq. (10), for any \( \Theta \in \mathcal{O} \) we can write
\[
\Theta = o(\lambda, \lambda') + \Upsilon, \| \Upsilon \|_F \leq R(\lambda, \lambda').
\]

Therefore, if \( q = 1 \), i.e. \( q = \infty \), we have
\[
\sup_{\Theta \in \mathcal{O}} \left\| \tilde{d}, \Theta \right\|_1 = \sup_{\| \Upsilon \|_F \leq R(\lambda, \lambda')} \left\| \tilde{d}, o(\lambda, \lambda') + \tilde{d}, \Upsilon \right\|_1 = \| \tilde{d}, o(\lambda, \lambda') \|_1 + \| \tilde{d}, \Upsilon \|_2.
\]

else if \( q = 2 \), we have
\[
\sup_{\Theta \in \mathcal{O}} \left\| \tilde{d}, \Theta \right\|_2 = \sup_{\| \Upsilon \|_F \leq R(\lambda, \lambda')} \left\| \tilde{d}, o(\lambda, \lambda') + \tilde{d}, \Upsilon \right\|_2 = \| \tilde{d}, o(\lambda, \lambda') \|_2 + \| \tilde{d}, \Upsilon \|_2.
\]

else if \( q = \infty \), we have
\[
\sup_{\Theta \in \mathcal{O}} \left\| \tilde{d}, \Theta^* \right\|_\infty = \sup_{\| \Upsilon \|_F \leq R(\lambda, \lambda')} \left\| \tilde{d}, o(\lambda, \lambda') + \tilde{d}, \Upsilon \right\|_\infty = \| \tilde{d}, o(\lambda, \lambda') \|_\infty + \| \tilde{d}, \Upsilon \|_2.
\]

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

**Proof of Lemma 3**

Because the graph \( G \) in CGCC model is a cyclic graph, there exists at least one ring in \( E_C \). Recall that we require \( G \) to be connected graph. Therefore, if there exist rings in \( E_C \), we must have that the cardinality \( |E_C| = m \geq n > n - 1 \).

Moreover, from the proof of Lemma 1, we know that if a ring exists in \( E_C \), there must exist one row of \( \bar{C} \) that can be linearly represented by some other rows. For any \( m \geq n \), there exists at least one ring in \( E_C \), therefore, \( \bar{C} < n \) is rank-deficient. \( \square \)

**Proof of Theorem 4**

The proofs can be completed by following those of Theorem 1. \( \square \)

**Proof of Theorem 5**

We prove the case when \( q = \hat{q} = 2 \), and other cases can be proved in a similar way. When \( \lambda' < \lambda_{\text{max}} \), Eq. (19) is equivalent to
\[
\| \Lambda \hat{F}^*(\lambda) - \Lambda \tilde{F}^*(\lambda') \|_F^2 \leq \langle \Lambda \hat{F}^*(\lambda) - \Lambda \tilde{F}^*(\lambda'), \nabla^+(\lambda, \lambda') \rangle.
\]

(35)

We will show Eq. (35) holds. Note that when \( \lambda' < \lambda_{\text{max}} \), \( \Lambda \tilde{F}(\lambda' = -\tilde{h}(\Phi^*) \). From Lemma 5, we have
\[
\langle \tilde{h}(\Phi^*) - \Lambda \tilde{F}^*(\lambda'), \nabla^+(\lambda, \lambda') \rangle \leq 0, \forall \tilde{h} \in \mathcal{H}.
\]

By letting \( \tilde{f} = \tilde{F}^*(\lambda) \) and \( \tilde{h} = 0 \), we can obtain the following results respectively:
\[
\langle \tilde{h}(\Phi^*) - \Lambda \tilde{F}^*(\lambda'), \nabla^+(\lambda, \lambda') \rangle \leq 0, \tag{37}
\]
\[
\| \Lambda^{-1} \tilde{h} \|_F \leq \| \Lambda \tilde{F}^*(\lambda') \|_F, \tag{38}
\]

Moreover, from Lemma 5, we also have
\[
-\tilde{h}(\Phi^*) = \frac{\sqrt{\lambda}}{\lambda} \frac{\| \tilde{F}^*(\lambda') \|_F}{\| \nabla^+(\lambda, \lambda') \|_F}.
\]

Then, we have
\[
\langle \tilde{h}(\Phi^*) - \Lambda \tilde{F}^*(\lambda'), \nabla^+(\lambda, \lambda') \rangle \leq 0. \tag{39}
\]

Eq. (39) is equivalent to
\[
\| \Lambda \tilde{F}^*(\lambda) - \Lambda \tilde{F}^*(\lambda') \|_F^2 \leq \langle \Lambda \tilde{F}^*(\lambda) - \Lambda \tilde{F}^*(\lambda'), \nabla^+(\lambda, \lambda') \rangle. \tag{40}
\]

Comparing Eq. (40) with Eq. (35), we consider Eq. (35) again:
\[
\langle \Lambda \tilde{F}^*(\lambda) - \Lambda \tilde{F}^*(\lambda'), \nabla^+(\lambda, \lambda') \rangle = \langle \Lambda \tilde{F}^*(\lambda) - \Lambda \tilde{F}^*(\lambda'), \nabla^+(\lambda, \lambda') \rangle.
\]

(41)
Based on Eq. (41), recall Eq. (37) and we know that if \( \langle \nabla \lambda, \nabla \lambda^* \rangle \geq 0 \), we can prove the theorem. Actually, we have
\[
\langle \nabla \lambda, \nabla \lambda^* \rangle = \left\langle \frac{\Lambda^{-1} \nabla \lambda}{\lambda} - \Lambda \Phi^*(\lambda^*), \nabla \lambda^* \right\rangle
\]
\[
= \left( \frac{1}{\lambda} - \frac{1}{\lambda^*} \right) \langle \Lambda^{-1} \nabla \lambda, \nabla \lambda^* \rangle + \left\langle \frac{\Lambda^{-1} \nabla \lambda}{\lambda} - \Lambda \Phi^*(\lambda^*), \nabla \lambda^* \right\rangle.
\]
From Eq. (38), we have
\[
\langle \Lambda^{-1} \nabla \lambda, \nabla \lambda^* \rangle \geq \|\Lambda^{-1} \nabla \lambda\|_{\lambda}^{2} - \|\Lambda^{-1} \nabla \lambda\|_{\lambda} \|\Lambda \Phi^*(\lambda^*)\|_{\lambda}^{2} \geq 0.
\]
Now the last thing is to show that
\[
\left\langle \frac{\Lambda^{-1} \nabla \lambda}{\lambda} - \Lambda \Phi^*(\lambda^*), \nabla \lambda^* \right\rangle \geq 0.
\]
Eq. (42) is obvious, since
\[
\left\langle \frac{\Lambda^{-1} \nabla \lambda}{\lambda} - \Lambda \Phi^*(\lambda^*), \nabla \lambda^* \right\rangle = \|\nabla \lambda^*\|_{\lambda}^{2} \geq 0.
\]
Now we complete the proof. \( \square \)

**Details for Definition 1**

Let \( \mathcal{O} \) be the feasible region constructed from Eqs. (19):
\[
\mathcal{O}(\lambda, \lambda^*) = \{ \Phi(\lambda) : \|\Lambda \Phi(\lambda) - \nabla (\lambda, \lambda^*)\|_{\lambda} \leq \nabla(\lambda, \lambda^*) \}.
\]
We have to estimate the following supreme values:
\[
\sup_{\Phi} \{ \|\nabla \Phi_s\|_{\lambda} : \Phi \in \mathcal{O}, \; r = 1, \cdots, m \}.
\]
Let \( \Xi = \Lambda \Phi^* \). The problem becomes
\[
\sup_{\Xi} \{ \|\zeta_r\|_{\lambda} : \Lambda^{-1} \Xi \in \mathcal{O}, \; r = 1, \cdots, m \},
\]
where \( \zeta_r \) is the \( r \)th row of \( \Lambda^{-1} \). The supreme values in Eq. (45) can be obtained exactly from the proof of Theorem 3. \( \square \)

**Analysis for \( \delta \)**

The parameter \( \delta \) in problem (18) plays an important role, since \( \lambda_{\max} \) depends on the value of \( \delta \). When \( \delta \) is very small, computing \( \lambda_{\max} \) may be numerically unstable but when \( \delta \) is large, the relaxed dual form will have a large deviation from the original one, leading to an inaccurate estimation for \( \lambda_{\max} \). 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 \( \delta \) such that the induced \( \lambda_{\max}^* \) satisfies that \( \lambda_{\max}^* \) is close to \( \lambda_{\max}^* \) but \( \lambda_{\max}^* < \lambda_{\max} \), which makes the condition in Theorem 1 satisfied. Empirically, we can assign a relatively large initial value to \( \delta \) which will induce a small \( \lambda_{\max}^* \) and then decrease \( \delta \) gradually until \( \lambda_{\max}^* < \lambda_{\max} \) is satisfied.

---

**Efficient Ways for Computing The Matrices Used in The Cigar Rule**

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

\[
\mathbf{D}^{-1} = \left( \mathbf{C} \mathbf{C}^T + \delta \mathbf{I} \right)^{-1} = \frac{1}{\delta} \left( \mathbf{C} \mathbf{C}^T + \delta \mathbf{I} \right)^{-1} = \frac{1}{\delta} \left( \mathbf{I} + \frac{\mathbf{C} \mathbf{C}^T}{\sqrt{\delta}} \right)^{-1},
\]

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

---

**Additional Experimental Results**

Fig. 5 shows the \( \ell_2 \) 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 \( n = 1000 \).

---

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

<table>
<thead>
<tr>
<th>Data</th>
<th>Solver</th>
<th>Eater</th>
<th>E+S</th>
<th>Speedup</th>
</tr>
</thead>
<tbody>
<tr>
<td>halfmoon</td>
<td>1495.6</td>
<td>4.1</td>
<td>288.7</td>
<td>5.2</td>
</tr>
<tr>
<td>spiral</td>
<td>3563.5</td>
<td>4.3</td>
<td>804.1</td>
<td>4.4</td>
</tr>
</tbody>
</table>

---

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

<table>
<thead>
<tr>
<th>Data</th>
<th>Model</th>
<th>Solver</th>
<th>Cigar</th>
<th>C+S</th>
<th>Speedup</th>
</tr>
</thead>
<tbody>
<tr>
<td>Halfmoon</td>
<td>CGCC-1</td>
<td>3867.8</td>
<td>149.7</td>
<td>722.7</td>
<td>5.4</td>
</tr>
<tr>
<td>(n=1000)</td>
<td>CGCC-2</td>
<td>4709.7</td>
<td>301.1</td>
<td>1028.4</td>
<td>4.6</td>
</tr>
<tr>
<td>Spiral</td>
<td>CGCC-1</td>
<td>10314.8</td>
<td>164.3</td>
<td>2349.8</td>
<td>4.4</td>
</tr>
<tr>
<td>(n=1000)</td>
<td>CGCC-2</td>
<td>12735.2</td>
<td>338.4</td>
<td>3178.4</td>
<td>4.0</td>
</tr>
</tbody>
</table>
Figure 5: $\ell_2$ clusterpath generated by the GCC models on the synthetic datasets. (a)-(c): halfmoon data (n=200); (d)-(f) spiral data (n=200).

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.