A Proofs of Propositions in Section 3

A.1 Proofs of Prop. 1

Suppose cycle $c \notin \mathcal{C}$ is induced by cycles $c_1, \ldots, c_K \in \mathcal{C}$ and intermediate cycles sets $c^{(1)}, \ldots, c^{(K)}$ where $c^{(1)} = c_1$ and $c^{(K)} = c$. $\mathcal{F}$ is a map graph that is consistent for basis $\mathcal{C}$. We will show that $c^{(k)}, 1 \leq k \leq K$ are all consistent over $\mathcal{F}$ inductively.

It is clear $c^{(1)}$ is consistent over F. Consider $c^{(k)} = c^{(k-1)} \oplus c_k$ where $c_{(k-1)}$ is consistent over $\mathcal{F}$ and $c_k \in \mathcal{C}$. Since $c^{(k)}$ is also a simple cycle $c^{(k-1)} \cap c_k$ must be a simple path p. Without loss of generality we can assume that $c^{(k-1)} = p \sim s_1$ and $c_k = p^{-1} \sim s_2$ where $\sim$ is path concatenation operator and p^{-1} means the reversed orientation of path p. Thus by induction of composition over paths we have $f_{s_1} \circ f_p = Id$ and $f_{p^{-1}} \circ f_{s_2} = Id$. Since $f_{p^{-1}} = f_p^{-1}$ we have $f_{s_1} \circ f_{s_2} = Id$. However it is easy to see that $c^{(k)} = s_1 \sim s_2$, which leads to $f_{c^{(k)}}$, or $c^{(k)}$ is consistent over $\mathcal{F}$. The proposition follows immediately by noting that $c^{(K)} = c$.

A.2 Proof of Prop. 2

The cycle-consistent basis is a special case of the binary cycle basis which relaxes the condition that all intermediate products must be simple cycles. It is known that the minimum size of binary cycle basis would be $|\mathcal{E}| - |\mathcal{V}| + 1$ [22].

Next we show the cycle basis $\mathcal{C}_\mathcal{T}$ induced from spanning tree $\mathcal{T}$ is a cycle-consistent basis. Let p_{uv} be the unique path from u to v in spanning tree $\mathcal{T}$. Given simple cycle $c = i_1 i_2 \ldots i_k$, we have

$$f_{(i_j, i_{j+1}) \sim p_{i_{j+1}, i_j}} = Id$$

for all $j = 1, \ldots, k$ since $(i_j, i_{j+1}) \sim p_{i_{j+1}, i_j}$ is in $\mathcal{C}$ by definition. Thus

$$f_{i_j i_{j+1}} = f^{-1}_{p_{i_{j+1}, i_j}} = f_{p_{i_j, i_{j+1}}}.$$

The last equality comes from $f_{uv} = f_{uv}^{-1}$ holds for all $(u, v) \in \mathcal{E}$. As such it is clear that

$$f_{p_{i_k, i_1}} \circ f_{p_{i_{k-1}, i_k}} \circ \cdots \circ f_{p_{i_1, i_2}} = f_{p_{i_1, i_1}} = Id$$

by noticing the uniqueness of paths on spanning tree. Thus we have shown that

$$f_c = f_{i_k i_1} \circ f_{i_{k-1}, i_k} \circ \cdots \circ f_{i_1, i_2} = Id.$$

So the induced basis $\mathcal{C}_\mathcal{T}$ is cycle-consistent.

A.3 Proof of Prop. 3

We provide the following counter example.

Consider the following 6 cycles from the cube graph below:

$$\begin{aligned}
C_1 &: a \to b \to c \to g \to h \to e \to a \\
C_2 &: b \to c \to d \to h \to e \to f \to b \\
C_3 &: c \to d \to a \to e \to f \to g \to c \\
C_4 &: d \to a \to b \to f \to g \to h \to d
\end{aligned}$$

$$\begin{aligned}
C_5 &: a \to b \to c \to d \to a \\
C_6 &: a \to e \to h \to d \to a \\
C_7 &: a \to b \to f \to e \to a \\
C_8 &: e \to f \to g \to h \to e \\
C_9 &: c \to d \to h \to g \to c \\
C_{10} &: b \to c \to g \to f \to b
\end{aligned}$$

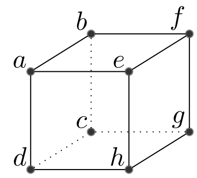

Figure 3: A cube graph

It is easy to check the following equations (all under 2-modulo sense):

$$\begin{array}{lcl} C_1 \oplus C_2 \oplus C_3 & = & C_4 \\ C_5 \oplus C_6 \oplus C_7 \oplus C_8 \oplus C_9 & = & C_{10} \\ C_1 \oplus C_3 & = & C_5 \oplus C_8 \\ C_1 \oplus C_5 & = & C_6 \oplus C_9 \\ C_2 \oplus C_5 & = & C_6 \oplus C_7 \end{array} \tag{12}$$

If we set $\mathcal{B} = \{C_1, C_2, C_3, C_5, C_6\}$, then equation group (12) shows that $C_5, C_6, \ldots, C_{10}$ can be composed from $\mathcal{B}$ and thus furthermore form a cycle basis of the cube graph. However a function network that is consistent on C_1, C_2, C_3, C_5, C_6 is not necessarily consistent on C_4, which means $\mathcal{B}$ is not a cycle-consistency basis even though it is indeed a cycle basis.

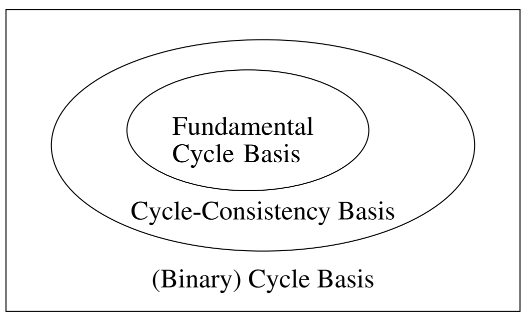

Figure 4: Venn diagram for subsets of cycle bases

B Proof of Theorem. 4.1

We first make a formal statement for Prop. 4.1. Let us begin with a few assumptions about the underlying neural networks with respect to the optimal network parameters $\theta_{ij}^{\star}$:

- **Cycle-consistency is exact.** For any cycle $c \in \overline{\mathcal{C}}$,

$$f_c^{\Theta^\star} = Id \tag{13}$$

  where $\Theta^\star$ collects all network parameters. A consequence of this is that the map f_{ij} from $\mathcal{D}_i$ to $\mathcal{D}_j$ is unique. In the following, we will consider sets of consistent correspondences across the entire map network, i.e, $x_i \in \mathcal{D}_i, 1 \leq i \leq |\mathcal{V}|$, and

$$f_{ij}^{\theta_{ij}^\star}(x_i) = x_j, \quad \forall (i,j) \in \mathcal{E}.$$

- **Bounded distortion.** Let K denote the dimension of the domains $\mathcal{D}_i$. There exists universal constants $0 < c_1 < c_2$, so that for each set of consistent correspondences $x_i \in \mathcal{D}_i, 1 \leq i \leq |\mathcal{V}|$,

$$c_1 \leq \sigma_{\min}\left(\frac{\partial f_{p_{ij}}^{\Theta^\star}}{\partial x}\right) \leq \sigma_{\min}\left(\frac{\partial f_{p_{ij}}^{\Theta^\star}}{\partial x}\right) \leq c_2. \tag{14}$$

- **Bounded network gradients.** There exist universal constants $c_3, c_4 > 0$, so that

$$
\begin{aligned}
c_3 \le \sigma_{\min}(E \frac{\partial f_{ij}^{\theta_{ij}}}{\partial \theta_{ij}}(\theta_{ij}^\star, x))^T \frac{\partial f_{ij}^{\theta_{ij}}}{\partial \theta_{ij}}(\theta_{ij}^\star, x)) \\
\le \sigma_{\max}(E \frac{\partial f_{ij}^{\theta_{ij}}}{\partial \theta_{ij}}(\theta_{ij}^\star, x))^T \frac{\partial f_{ij}^{\theta_{ij}}}{\partial \theta_{ij}}(\theta_{ij}^\star, x)) \le c_4, \quad \forall x \in \mathcal{D}_i,\ (i,j) \in \mathcal{E},
\end{aligned}
$$

- **Bounded Hessian matrices of the loss terms.** There exist universal constants c_5 and c_6 so that the Hessian matrix of each loss term $l_{ij}(\theta_{ij}^\star)$ satisfies

$$
c_5 I \preceq H(l_{ij}(\theta_{ij}^\star)) \preceq c_6 I.
$$

Let $H(\Theta^\star)$ denote the Hessian matrix of joint map optimization. Due to the exactness of the cycle-consistency constraint, it is clear that

$$
H(\Theta^\star) := \sum_{(i,j)\in\mathcal{E}_0} (\boldsymbol{v}_e \otimes I) H(l_{ij}(\theta_{ij}^\star))(\boldsymbol{v}_e^T \otimes I) + \sum_{c=(i_1\cdots i_k i_1)\in\mathcal{C}} w_c E_{x_{i_1}} (\frac{\partial f_c^{\Theta^\star}}{\partial \Theta}(x_{i_1})^T (\frac{\partial f_c^{\Theta^\star}}{\partial \Theta}(x_{i_1}) \tag{15}
$$

where $\frac{\partial f_c^{\Theta^\star}}{\partial \Theta}(x_{i_1})$ is the Jacobi of each f_c with respect to network parameters evaluated at (x_{i_1}). Note that unless other-wise stated, we use I to denote all identity matrices, whose dimension is inferred from the context.

With this set up, we present a formal statement of Theorem. 4.1:

Theorem B.1 *Under the assumptions described above, we have*

$$
\kappa(H(\Theta^\star)) \le \frac{\max(\frac{c_4^2}{c_1^2 c_3^2}, c_2^2) c_2^2 c_4}{\min(\frac{c_3^2}{c_2^2 c_4^2}, c_1^2) c_1^2 c_3} \kappa(H).
$$

B.1 Proof of Theorem. B.1

Since f_c involves the composition of neural networks along cycles, we begin with expanding its Jacobi matrix:

$$
\begin{aligned}
\partial f_c^{\Theta^\star}(x_{i_1}) = & \sum_{l=1}^{k} \frac{\partial f_{i_k i_1} \circ f_{i_{l+1} i_{l+2}}}{\partial x_{i_{k+1}}}(f_{i_l i_{l+1}} \circ f_{i_1 i_2}(x_{i_1})) \cdot \\
& \frac{\partial f_{i_l i_{l+1}}}{\partial \theta_{i_l i_{l+1}}}((f_{i_l i_{l+1}} \circ f_{i_1 i_2}(x))) d\theta_{i_l i_{l+1}}, \quad \forall x_{i_1} \in \mathcal{D}_{i_1}, c \in \mathcal{C}_{\text{sup}}.
\end{aligned} \tag{16}
$$

The key idea of our proof is to define n^2 pairs of matrices $A_{ij} \in \mathbb{R}^{K\times K}, 1 \le i, j \le |\mathcal{V}|$ so that

$$
A_{i_{l+1} i_1} := \frac{\partial f_{i_k i_1} \circ f_{i_{l+1} i_{l+2}}}{\partial x_{i_{k+1}}}(f_{i_l i_{l+1}} \circ f_{i_1 i_2}(x)),
$$

Intuitively, A_{ij} is the Jacobi matrix of f_{ij} at x_i. It is easy to check that this definition is proper (due to the cycle-consistency constraint), and these matrices satisfy the cycle-consistency properties:

Fact 2 *For all triplets of shapes* $1 \le i_1, i_2, i_3 \le n$*, we have*

$$
A_{i_2 i_3} A_{i_1 i_2} = A_{i_1 i_3}, \forall 1 \le i_1, i_2, i_3 \le |\mathcal{V}|
$$

Define $B_e := \frac{\partial f_{ij}^{\theta_{ij}^\star}}{\partial \theta_{ij}}(x_i), \quad \forall e = (i,j) \in \mathcal{E}$. Note that B_e is dependent on x_i but we omit x_i in the expression for brevity. Similarly, introduce matrix $J_A^{i_1} \in \mathbb{R}^{K\times(|\mathcal{E}|K)}$, where the block that corresponds to $e = (i,j)$ is given by A_{ii_1}. We can rewrite $\partial f_c^{\Theta^\star}(x_{i_1})$ as

$$
\begin{aligned}
\partial f_c^{\Theta^\star}(x_{i_1}) &= \sum_{l=1}^{k} A_{i_{l+1} i_1} B_{i_l i_{l+1}} d\theta_{i_l i_{l+1}} \\
&= J_A^{i_1}(\text{diag}(\boldsymbol{v}_c) \otimes I) J_{\Theta^\star} d\Theta
\end{aligned} \tag{17}
$$

where $J_{\Theta^\star}(x) := \mathrm{diag}(B_e)$ collect the Jacobi matrices in the diagonal block, where $\mathrm{diag}(\boldsymbol{v}_c) \in \mathbb{R}^{|\mathcal{E}|\times|\mathcal{E}|}$ is the diagonal matrix whose diagonal elements correspond to $\boldsymbol{v}_c$.

The following inequality characterizes a relation among $J_A^{i_1}$ and J_A^1, which will be used later:

$$c_1^2 {J_A^1}^T J_A^1 \preceq {J_A^{i_1}}^T J_A^{i_1} = {J_A^1}^T A_{i_1 1}^T A_{i_1 1} J_A^1 \preceq c_2^2 {J_A^1}^T J_A^1, \quad \forall 1 \le i_1 \le n. \tag{18}$$

We will also use another formulation of $\partial f_c^{\Theta^\star}(x_{i_1})$. Let $Diag(A^{i_1}) = \mathrm{diag}(J_A^{i_1})$ collect the blocks of $J_A^{i_1}$ in the diagonal block. We also have

$$\partial f_c^{\Theta^\star}(x_{i_1}) = (\boldsymbol{v}_c^T \otimes I)\cdot Diag(A^{i_1}) J_{\Theta^\star} d\Theta \tag{19}$$

We proceed to lower bound $H_{\Theta^\star}$, and obtaining an upper bound can be done in a similar fashion:

$$\begin{aligned} H(\Theta^\star) = & \sum_{e=(i,j)\in\mathcal{E}_0} \big((\boldsymbol{v}_e \otimes I) H(l_{ij}(\theta_{ij}^\star))(\boldsymbol{v}_e \otimes I_m) \\ & + \sum_{c=(i_1\cdots i_k i_1)\in\mathcal{C}} w_c E {J_{\Theta^\star}}^T (\mathrm{diag}(\boldsymbol{v}_c)\otimes I_m) {J_A^{i_1}}^T J_A^{i_1} (\mathrm{diag}(\boldsymbol{v}_c)\otimes I_m) J_{\Theta^\star}. \end{aligned}$$

Using $H(l_{ij}(\theta_{ij}^\star)) \succeq c_5 I$, we have

$$H(\Theta^\star) \succeq c_5 \sum_{e=(i,j)\in\mathcal{E}_0} (\boldsymbol{v}_e \boldsymbol{v}_e^T)\otimes I \tag{20}$$

$$+ \sum_{c=(i_1\cdots i_k i_1)\in\mathcal{C}} w_c E {J_{\Theta^\star}}^T (\mathrm{diag}(\boldsymbol{v}_c)\otimes I_m) {J_A^{i_1}}^T J_A^{i_1} (\mathrm{diag}(\boldsymbol{v}_c)\otimes I_m) J_{\Theta^\star} \tag{21}$$

Combing (21) and (18), we have

$$\begin{aligned} H(\Theta^\star) \succeq\ & c_5^2 \sum_{e=(i,j)\in\mathcal{E}_0} (\boldsymbol{v}_e \boldsymbol{v}_e^T)\otimes I \\ & + c_1^2 \sum_{c=(i_1\cdots i_k i_1)\in\mathcal{C}} w_c E {J_{\Theta^\star}}^T (\mathrm{diag}(\boldsymbol{v}_c)\otimes I) {J_A^1}^T J_A^1 (\mathrm{diag}(\boldsymbol{v}_c)\otimes I) J_{\Theta^\star} \\ =\ & c_5^2 \sum_{e=(i,j)\in\mathcal{E}_0} (\boldsymbol{v}_e \boldsymbol{v}_e^T)\otimes I \\ & + c_1^2 E {J_{\Theta^\star}}^T Diag(A^1)^T \big((\sum_{c=(i_1\cdots i_k i_1)\in\mathcal{C}} w_c \boldsymbol{v}_c \boldsymbol{v}_c^T)\otimes I\big) Diag(A^1) J_{\Theta^\star} \end{aligned}$$

Note that $H = \sum_{e=(i,j)\in\mathcal{E}_0} (\boldsymbol{v}_e \boldsymbol{v}_e^T) + \sum_{c=(i_1\cdots i_k i_1)\in\mathcal{C}} w_c \boldsymbol{v}_c \boldsymbol{v}_c^T)$. It follows that

$$\begin{aligned} H(\Theta^\star) \succeq\ & c_5^2 \sum_{e=(i,j)\in\mathcal{E}_0} (\boldsymbol{v}_e \boldsymbol{v}_e^T)\otimes I \\ & + \min(\frac{c_5^2}{c_2^2 c_4^2}, c_1^2) E {J_{\Theta^\star}}^T Diag(A^{i_1})^T \big((H - \sum_{e=(i,j)\in\mathcal{E}_0} (\boldsymbol{v}_e \boldsymbol{v}_e^T))\otimes I\big) Diag(A^{i_1}) J_{\Theta^\star} \\ \succeq\ & c_5^2 \sum_{e=(i,j)\in\mathcal{E}_0} (\boldsymbol{v}_e \boldsymbol{v}_e^T)\otimes I + \min(\frac{c_5^2}{c_2^2 c_4^2}, c_1^2) E {J_{\Theta^\star}}^T Diag(A^1)^T \big(H\otimes I\big) Diag(A^1) J_{\Theta^\star} \\ & - \min(\frac{c_5^2}{c_2^2 c_4^2}, c_1^2) E {J_{\Theta^\star}}^T Diag(A^1)^T \big(\sum_{e=(i,j)\in\mathcal{E}_0} (\boldsymbol{v}_e \boldsymbol{v}_e^T))\otimes I\big) Diag(A^1) J_{\Theta^\star} \\ \succeq\ & c_5^2 \sum_{e=(i,j)\in\mathcal{E}_0} (\boldsymbol{v}_e \boldsymbol{v}_e^T)\otimes I + \min(\frac{c_5^2}{c_2^2 c_4^2}, c_1^2) \lambda_{\min}(H) E {J_{\Theta^\star}}^T Diag(A^1)^T Diag(A^1) J_{\Theta^\star} \\ & - \min(\frac{c_5^2}{c_2^2 c_4^2}, c_1^2) E {J_{\Theta^\star}}^T Diag(A^1)^T (\sum_{e=(i,j)\in\mathcal{E}_0} \boldsymbol{v}_e \boldsymbol{v}_e^T)\otimes I Diag(A^1) J_{\Theta^\star} \end{aligned}$$

$$\succeq c_5^2 \sum_{e=(i,j)\in\mathcal{E}_0} (\boldsymbol{v}_e\boldsymbol{v}_e^T)\otimes I + \min(\frac{c_5^2}{c_2^2c_4^2}, c_1^2)\lambda_{\min}(H)c_1^2 E J_{\Theta^\star}{}^T J_{\Theta^\star}$$

$$- \min(\frac{c_5^2}{c_2^2c_4^2}, c_1^2)c_2^2 E J_{\Theta^\star}{}^T (\sum_{e=(i,j)\in\mathcal{E}_0} \boldsymbol{v}_e\boldsymbol{v}_e^T)\otimes I J_{\Theta^\star}$$

$$\succeq c_5^2 \sum_{e=(i,j)\in\mathcal{E}_0} (\boldsymbol{v}_e\boldsymbol{v}_e^T)\otimes I + \min(\frac{c_5^2}{c_2^2c_4^2}, c_1^2)\lambda_{\min}(H)c_1^2 c_3$$

$$- \min(\frac{c_5^2}{c_2^2c_4^2}, c_1^2)c_2^2 c_4 (\sum_{e=(i,j)\in\mathcal{E}_0} \boldsymbol{v}_e\boldsymbol{v}_e^T)\otimes I$$

$$\succeq \min(\frac{c_3^2}{c_2^2c_4^2}, c_1^2)c_1^2 c_3 \lambda_{min}(H) I.$$

Like-wise, we can show that the Hessian matrix is upper bounded by

$$H(\Theta) \preceq \max(\frac{c_4^2}{c_1^2c_3^2}, c_2^2)c_2^2 c_4 \lambda_{\max}(H) I.$$

This means the condition number of $H(\Theta^\star)$

$$\kappa(H(\Theta^\star)) \leq \frac{\max(\frac{c_4^2}{c_1^2c_3^2}, c_2^2)c_2^2c_4}{\min(\frac{c_3^2}{c_2^2c_4^2}, c_1^2)c_1^2c_3}\kappa(H),$$

which ends the proof.

C Proof of Theorem 4.2

We first present a formal statement of Theorem 4.2.

Definition 7 *Given a finite point set $P \subseteq S^m$ where $S^m := \{\boldsymbol{x} \in \mathbb{R}^m : |\boldsymbol{x}| = 1\}$ is the unit sphere in $\mathbb{R}^m$, the spherical Voronoi partition of P is defined as a partition $\{P_{\boldsymbol{v}} : \boldsymbol{v} \in P\}$ such that*

$$P_{\boldsymbol{v}} = \{\boldsymbol{u} \in S^m : d(\boldsymbol{u}, \boldsymbol{v}) \leq d(\boldsymbol{u}, \boldsymbol{v}')\ \forall \boldsymbol{v}' \in P\}$$

in which d is the Euclidean distance in $\mathbb{R}^m$. Intuitively, $P_{\boldsymbol{v}}$ is consisting of the neighborhood of $\boldsymbol{v}$ on S^m. Also, $P_{\boldsymbol{v}}$ is known to be connected, so we can define $S(\boldsymbol{v})$ as the generalized area of $P_{\boldsymbol{v}}$ on S^m.

Theorem C.1 *Let $s_1^\star$ and $s_2^\star$ be the optimal solution to (6). Denote $|\mathcal{E}|$ by m.*

1) For all feasible s_1 and s_2 we have

$$s_1 \leq \frac{|\mathcal{E}^0| + \lambda}{m} \leq s_2$$

2) Denote P as $\{\boldsymbol{v}_c/|\boldsymbol{v}_c|\} \cup \{\boldsymbol{v}_e\}$. Let $\{P_{\boldsymbol{v}}\}$ be the Voronoi partition of point set $\overline{P} = P \cup -P$ where $-P := \{-\boldsymbol{v} : \boldsymbol{v} \in P\}$. Define $\epsilon_{\boldsymbol{v}} := \sup_{\boldsymbol{u}\in P_{\boldsymbol{v}}} \|\boldsymbol{v}\boldsymbol{v}^T - \boldsymbol{u}\boldsymbol{u}^T\|$ for $\boldsymbol{v} \in \overline{P}$.

Setting $w_c = \frac{\lambda}{S_m|\boldsymbol{v}_c|^2}(S(\boldsymbol{v}_c) + S(-\boldsymbol{v}_c))$ and $w_e = \frac{\lambda}{S_m}(S(\boldsymbol{v}_e) + S(-\boldsymbol{v}_e))$, we have

$$s_2^\star - s_1^\star \leq \frac{\lambda}{S_m}\sum_{\boldsymbol{v}\in\overline{P}} S(\boldsymbol{v})\epsilon_{\boldsymbol{v}} + \|\sum_{e\in\mathcal{E}^0}(1-w_e)\boldsymbol{v}_e\boldsymbol{v}_e^T\| + \delta\Big\|\sum_{\substack{c\in\mathcal{C}_{\text{sup}}\\ w_c<\delta}} \boldsymbol{v}_c\boldsymbol{v}_c^T\Big\|, \tag{22}$$

in which $S_m = \frac{m\pi^{m/2}}{\Gamma(m/2+1)}$ is the area of unit sphere in $\mathbb{R}^m$.

Observe that whenever $\overline{P}$ is densely distributed over S^m and δ is relatively small, the dominating terms in (22) would be

$$\|\sum_{e\in S^0}(1-w_e)\boldsymbol{v}_e\boldsymbol{v}_e^T\|.$$

Proof 1 *The first part of the proposition is obtained directly by taking trace on (C1). Now focus on the second part.*

Define

$$M = \sum_{e \in \mathcal{E}^0} w_e \boldsymbol{v}_e \boldsymbol{v}_e^T + \sum_{c \in \mathcal{C}_{\text{sup}}} w_c \boldsymbol{v}_c \boldsymbol{v}_c^T \tag{23}$$

$$L = \sum_{e \in \mathcal{E}^0} \boldsymbol{v}_e \boldsymbol{v}_e^T + \sum_{c \in \mathcal{C}_{\text{sup}}} w_c \boldsymbol{v}_c \boldsymbol{v}_c^T \tag{24}$$

$s_1^\star = \lambda_{\min}(L)$ *and* $s_2^\star = \lambda_{\max}(L')$ *together with* $\{w_c\}$ *would be a feasible solution to (6) except constraint (C3) where* $\lambda_{\min}(L)$ *and* $\lambda_{\max}(L)$ *are smallest and largest eigenvalues of L respectively. We postpone (C3) constraint to later discussion.*

Consider the following generalized surface integral on S^m

$$J = \frac{\lambda}{S_m} \int_{\boldsymbol{v} \in S^m} \boldsymbol{v}\boldsymbol{v}^T dS.$$

By symmetry of S^m *we have* $\boldsymbol{u}^T J \boldsymbol{u}$ *is a constant for all unit vector* $\boldsymbol{u}$*. Thus* $J = KI$ *for some constant* K*. Since*

$$Tr[J] = \frac{\lambda}{S_m} \int_{\boldsymbol{v} \in S^m} Tr[\boldsymbol{v}\boldsymbol{v}^T] dS = \lambda m,$$

we have $J = \lambda I$.

Note that

$$\begin{aligned}
&\|M - J\| \\
=&\|\frac{\lambda}{S_m} \sum_{\boldsymbol{v} \in \overline{P}} S(\boldsymbol{v}) \boldsymbol{v}\boldsymbol{v}^T - J\| \\
=&\frac{\lambda}{S_m} \|\sum_{\boldsymbol{v} \in \overline{P}} \int_{\boldsymbol{u} \in P_{\boldsymbol{v}}} (\boldsymbol{v}\boldsymbol{v}^T - \boldsymbol{u}\boldsymbol{u}^T) dS\| \\
\leq&\frac{\lambda}{S_m} \sum_{\boldsymbol{v} \in \overline{P}} \int_{\boldsymbol{u} \in P_{\boldsymbol{v}}} \|\boldsymbol{v}\boldsymbol{v}^T - \boldsymbol{u}\boldsymbol{u}^T\| dS \\
=&\frac{\lambda}{S_m} \sum_{\boldsymbol{v} \in \overline{P}} S(\boldsymbol{v}) \epsilon_{\boldsymbol{v}}.
\end{aligned}$$

and

$$\|M - L\| \leq \|\sum_{e \in \mathcal{E}^0} (1 - w_e) \boldsymbol{v}_e \boldsymbol{v}_e^T\|.$$

We have

$$\|L - J\| \leq \frac{\lambda}{S_m} \sum_{\boldsymbol{v} \in \overline{P}} S(\boldsymbol{v}) \epsilon_{\boldsymbol{v}} + \|\sum_{e \in \mathcal{E}^0} (1 - w_e) \boldsymbol{v}_e \boldsymbol{v}_e^T\|.$$

To make all $w_c \geq \delta$*, we upscale* w_c *to* $\max\{w_c, \delta\}$*, which make* L *increase by*

$$\delta \sum_{\substack{c \in \mathcal{C}_{\text{sup}} \\ w_c < \delta}} \boldsymbol{v}_c \boldsymbol{v}_c^T.$$

This operation may violate the (C2) constraint, so we rescale w_c *to fit (C2). Notice that this rescaling whose amplification is less than 1 will reduce the gap between largest and smallest eigenvalues. Collecting all results above, we have*

$$s_2^* - s_1^* \leq \frac{\lambda}{S_m} \sum_{\boldsymbol{v} \in \overline{P}} S(\boldsymbol{v}) \epsilon_{\boldsymbol{v}} + \|\sum_{e \in \mathcal{E}^0} (1 - w_e) \boldsymbol{v}_e \boldsymbol{v}_e^T\| + \delta \Big\| \sum_{\substack{c \in \mathcal{C}_{\text{sup}} \\ w_c < \delta}} \boldsymbol{v}_c \boldsymbol{v}_c^T \Big\|$$

with a constructive solution of w_c*. Here we used the fact that the eigen-gap of* J *is zero.*

D Solving (6) Using ADMM

In this section, we show how to solve (6) using alternating direction method of multiplers (or ADMM). To this end, we first reformulate the semi-definite program as follows:

$$
\begin{aligned}
\underset{\boldsymbol{w}, s_1, s_2}{\operatorname{argmin}} \quad & \alpha s_2 - s_1 \\
\text{subject to} \quad & X_1 = \sum_{e \in \mathcal{E}_0} \boldsymbol{v}_e \boldsymbol{v}_e^T + \sum_{c \in \mathcal{C}_{\text{sup}}} w_c \boldsymbol{v}_c \boldsymbol{v}_c^T - s_1 I && : Y_1 \\
& X_2 = s_2 I - \sum_{e \in \mathcal{E}_0} \boldsymbol{v}_e \boldsymbol{v}_e^T - \sum_{c \in \mathcal{C}} w_c \boldsymbol{v}_c \boldsymbol{v}_c^T && : Y_2 \\
& X_1 \succeq 0 && : S_1 \succeq 0 \\
& X_2 \succeq 0 && : S_2 \succeq 0 \\
& \boldsymbol{w}_{\min} \geq \delta && : \boldsymbol{y}_{\min} \geq 0 \\
& \boldsymbol{w}_{\text{active}} \geq 0 && : \boldsymbol{y}_{\text{active}} \geq 0
\end{aligned} \tag{25}
$$

where $w_{\min}$ and w_{active} collect weights of the cycles in $\mathcal{C}_{\min}$ and $\mathcal{C}_{\text{active}}$, respectively, and Y_1, Y_2, S_1, S_2, $\boldsymbol{y}_{\min}$ and $\boldsymbol{y}_{\text{active}}$ denote the dual variables. The Lagrangian of (25) is given by

$$
\begin{aligned}
\mathcal{L} := {} & \alpha s_2 - s_1 \\
& - \langle \sum_{e \in \mathcal{E}_0} \boldsymbol{v}_e \boldsymbol{v}_e^T + \sum_{c \in \mathcal{C}_{\text{sup}}} w_c \boldsymbol{v}_c \boldsymbol{v}_c^T - s_1 I - X_1, Y_1 \rangle \\
& - \langle \sum_{e \in \mathcal{E}_0} \boldsymbol{v}_e \boldsymbol{v}_e^T + \sum_{c \in \mathcal{C}_{\text{sup}}} w_c \boldsymbol{v}_c \boldsymbol{v}_c^T - s_2 I + X_2, Y_2 \rangle \\
& - \langle X_1, S_1 \rangle - \langle X_2, S_2 \rangle - \langle \boldsymbol{w}_{\min} - \delta \mathbf{1}, \boldsymbol{y}_{\min} \rangle - \langle \boldsymbol{w}_{\text{active}}, \boldsymbol{y}_{\text{active}} \rangle \\
= {} & \delta \sum_{c \in \mathcal{C}_{\min}} y_c - \langle \sum_{e \in \mathcal{E}_0} \boldsymbol{v}_e \boldsymbol{v}_e^T, Y_1 + Y_2 \rangle - s_1 \big(1 - \langle I, Y_1 \rangle\big) - s_2 \big(-\alpha - \langle I, Y_2 \rangle \big) \\
& - \langle X_1, S_1 - Y_1 \rangle - \langle X_2, S_2 + Y_2 \rangle - \sum_{c \in \mathcal{C}_{\text{sup}}} w_c \big(\langle \boldsymbol{v}_c \boldsymbol{v}_c^T, Y_1 + Y_2 \rangle + y_c \big)
\end{aligned} \tag{26}
$$

The dual problem is given by

$$
\begin{aligned}
\underset{S_1 \succeq 0, S_2 \succeq 0, \boldsymbol{y} \geq 0}{\text{minimize}} \quad & -\delta \sum_{c \in \mathcal{C}_{\min}} y_c + \langle \sum_{e \in \mathcal{E}_0} \boldsymbol{v}_e \boldsymbol{v}_e^T, Y_1 + Y_2 \rangle \\
\text{subject to} \quad & 1 - \langle I, Y_1 \rangle = 0 \\
& -\alpha - \langle I, Y_2 \rangle = 0 \\
& S_1 - Y_1 = 0 \\
& S_2 + Y_2 = 0 \\
& \langle \boldsymbol{v}_c \boldsymbol{v}_c^T, Y_1 + Y_2 \rangle + y_c = 0
\end{aligned} \tag{27}
$$

The augmented Lagrangian is given by

$$
\begin{aligned}
\overline{\mathcal{L}} := {} & -\delta \sum_{c \in \mathcal{C}_{\min}} y_c + \langle \sum_{e \in \mathcal{E}_0} \boldsymbol{v}_e \boldsymbol{v}_e^T, Y_1 + Y_2 \rangle + s_1 \big(1 - \langle I, Y_1 \rangle\big) + s_2 \big(-\alpha - \langle I, Y_2 \rangle \big) \\
& + \langle X_1, S_1 - Y_1 \rangle + \langle X_2, S_2 + Y_2 \rangle + \sum_{c \in \mathcal{C}_{\text{sup}}} w_c \big(\langle \boldsymbol{v}_c \boldsymbol{v}_c^T, Y_1 + Y_2 \rangle + y_c \big) \\
& + \frac{1}{2\mu} \Big(\|S_1 - Y_1\|_{\mathcal{F}}^2 + \|S_2 + Y_2\|_{\mathcal{F}}^2 \\
& + \sum_{c \in \mathcal{C}_{\text{sup}}} (\langle \boldsymbol{v}_c \boldsymbol{v}_c^T, Y_1 + Y_2 \rangle + y_c)^2 + (1 - \langle I, Y_1 \rangle)^2 + (\alpha + \langle I, Y_2 \rangle)^2 \Big).
\end{aligned} \tag{28}
$$

Starting from initial values of the primal variables:

$$
w_c^{(0)} = 0, \forall c \in \mathcal{C}_{\text{sup}}, \quad s_1^{(0)} = s_2^{(0)} = 0, \; X_1^{(0)} = X_2^{(0)} = 0. \tag{29}
$$

At each iteration of the ADMM, we first fix the primal variables to optimize the dual variables in a sequential manner.

Given the current dual variables at iteration k: $Y_1^{(k)}$, $Y_2^{(k)}$, $S_1^{(k)}$, $S_2^{(k)}$, $y_c, \forall c \in \mathcal{C}_{\text{sup}}$, we first fix S_1, S_2 and $y_c, \forall c \in \mathcal{C}_{\text{sup}}$ to optimize Y_1 and Y_2. In this case, optimizing $\mathcal{L}$ leads to two sub-optimization problems with decoupled optimizations of Y_1 and Y_2:

$$\begin{aligned}
\min_{Y_1,Y_2} \quad & \langle \sum_{e\in\mathcal{E}_0} \boldsymbol{v}_e\boldsymbol{v}_e^T, Y_1+Y_2\rangle - s_1^{(k)}\langle I, Y_1\rangle - s_2^{(k)}\langle I, Y_2\rangle \\
& - \langle X_1^{(k)}, Y_1\rangle + \langle X_2^{(k)}, Y_2\rangle + \langle \sum_{c\in\mathcal{C}_{\text{sup}}} w_c\boldsymbol{v}_c\boldsymbol{v}_c^T, Y_1+Y_2\rangle \\
& + \frac{1}{2\mu}\Big(\|S_1 - Y_1\|_{\mathcal{F}}^2 + \|S_2 + Y_2\|_{\mathcal{F}}^2 \\
& + \sum_{c\in\mathcal{C}_{\text{sup}}} (\langle \boldsymbol{v}_c\boldsymbol{v}_c^T, Y_1+Y_2\rangle + y_c)^2 + (1-\langle I, Y_1\rangle)^2 + (\alpha + \langle I, Y_2\rangle)^2\Big)
\end{aligned} \tag{30}$$

We employ conjugate gradient descent for optimzing (30), utilizing the fact that we usually have warm-start for Y_1 and Y_2 from the previous iteration. Note that thanks to the term $\|S_1 - Y_1\|_{\mathcal{F}}^2$ and $\|S_2 + Y_2\|_{\mathcal{F}}^2$. This linear system is usually well-conditioned, resulting fast convergence of the cg solver.

When Y_1, Y_2 and $y_c, c \in \mathcal{C}_{\text{sup}}$ are fixed, S_1 and S_2 may be optimized in isolation as follows:

$$\min_{S_1\succeq 0}\langle X_1, S_1 - Y_1^{(k+1)}\rangle + \frac{1}{2\mu}\|S_1 - Y_1^{(k+1)}\|_{\mathcal{F}}^2$$
$$\min_{S_2\succeq 0}\langle X_1, S_2 + Y_2^{(k+1)}\rangle + \frac{1}{2\mu}\|S_2 + Y_2^{(k+1)}\|_{\mathcal{F}}^2$$

Using the fact that the optimal solution to $\min\limits_{X\succeq 0}\|S - X\|_{\mathcal{F}}^2$ is given by the positive definite component of $\frac{S+S^T}{2}$, we conclude that the optimal value of S_1 and S_2 are given by

$$\begin{aligned}
S_1^{(k+1)} &:= U_1\max(\Lambda_1, 0){U_1}^T, \quad Y_1^{(k+1)} + \mu X_1^{(k)} = U_1\max(\Lambda_1,0){U_1}^T \\
S_2^{(k+1)} &:= U_2\max(\Lambda_2, 0){U_2}^T, \quad -(Y_2^{(k+1)} + \mu X_2^{(k)}) = U_2\Lambda_2{U_2}^T
\end{aligned}$$

Finally when Y_1, Y_2, S_1 and S_2 are fixed, we can optimize each y_c in isolation as

$$\min_{y_c\geq 0}\ \frac{1}{2\mu}\big(y_c + \boldsymbol{v}_c^T(Y_1^{(k)} + Y_2^{(k)})\boldsymbol{v}_c\big)^2 + w_c y_c - \delta\cdot Id[c\in\mathcal{C}_{\min}]\cdot y_c$$

which leads to

$$y_c = \max\big(0,\ \ \mu(\delta\cdot Id[c\in\mathcal{C}_{\min}) - w_c] - \boldsymbol{v}_c^T(Y_1^{(k)} + Y_2^{(k)})\boldsymbol{v}_c\big), \qquad \forall c \in \mathcal{C}_{\text{sup}}. \tag{31}$$

Once we have optimized the dual variables, we proceed to update the primal variables as follows:

$$\begin{aligned}
w_c^{(k+1)} &= w_c^{(k)} + \frac{y_c^{(k+1)} + \boldsymbol{v}_c^T(Y_1+Y_2)\boldsymbol{v}_c}{\mu}, \quad \forall c\in\mathcal{C}_{\text{sup}} \\
s_1^{(k+1)} &= s_1^{(k)} + \frac{1 - \text{Trace}(Y_1^{(k)})}{\mu} \\
s_2^{(k+1)} &= s_2^{(k)} - \frac{\alpha + \text{Trace}(Y_2^{(k)})}{\mu} \\
X_1^{(k+1)} &= X_1^{(k)} + \frac{S_1 - Y_1^{(k)}}{\mu} \\
X_2^{(k+1)} &= X_2^{(k)} + \frac{S_2 + Y_2^{(k)}}{\mu}
\end{aligned} \tag{32}$$

Regarding the hyper-parameters, we set $\mu = 10^{-2}$. At each iteration, we update $\mu = \rho\mu$, where $\rho = 1.01$ in all of our experiments. We run ADMM for 1000 iterations.

E Proofs of Theorems in Section 4

E.1 Concentration Bound for Rank-1 Matrices

Proposition 4 *Given n independent random variable $x_1, \ldots, x_n$ and n fixed rank-1 matrices $B_1, \ldots, B_n$ of the same shape with $\mathbb{E}[x_i] = 0$ and $\Pr[|x_i|\|B_i\| \leq M] = 1$ for $i = 1, \ldots, n$, S_n, the standard deviation of sum*

$$S_n = x_1 B_1 + \cdots + x_n B_n$$

is defined as

$$\sigma = \sqrt{Tr[\mathbb{E}[S_n^T S_n]]} = \sqrt{\sum_{i=1}^{n} \|B_i\|^2 \, \mathbb{E}\, x_i^2}.$$

If $M \leq \sigma$, S_n would be subject to the following concentration inequality:

$$\Pr[\|S_n\| \geq c\sigma] \leq Ae^{-Bc^2}$$

where A, B are universal constants.

Proof 2 *We consider controlling the m-th moment of S_n where m is an positive even number. Since*

$$Tr[\mathbb{E}\, S_n^m] = \sum_{i=1}^{d} \lambda_d^m(S_n) \geq \|S_n\|^m,$$

by Markov's inequality we have

$$\Pr[\|S_n\| \geq c\sigma] \leq c^{-m} \frac{Tr[\mathbb{E}\, S_n^m]}{\sigma^m}. \tag{33}$$

The terms in the expansion of S_n^m have the form

$$x_{i_1} B_{i_1} x_{i_2} B_{i_2} \ldots x_{i_m} B_{i_m}$$

where $i_j \in \{1, \ldots, n\}$ for $j \in \{1, \ldots, m\}$. The expectation of such term does not vanish only if no subscript appears exactly one time, otherwise $\mathbb{E}[x_i] = 0$ and the independence of x_i makes the term vanish.

Next we divide the non-vanishing terms into a collection of sets based on the equality relations of subscripts. To illustrate this idea, consider a special case of $m = 4$. There are four different types of terms:

$$\begin{array}{l|l}
x_{i_1}^4 B_{i_1}^4 & i_1 = i_2 = i_3 = i_4 \\
x_{i_1}^2 x_{i_2}^2 B_{i_1} B_{i_2} B_{i_1} B_{i_2} & i_1 = i_3, i_2 = i_4, i_1 \neq i_2 \\
x_{i_1}^2 x_{i_2}^2 B_{i_1} B_{i_2}^2 B_{i_1} & i_1 = i_4, i_2 = i_4, i_1 \neq i_2 \\
x_{i_1}^2 x_{i_3}^2 B_{i_1}^2 B_{i_3}^2 & i_1 = i_2, i_3 = i_4, i_1 \neq i_3
\end{array}$$

The sum of all terms of the first type would be bounded by

$$Tr[\mathbb{E}[\sum_{i_1=1}^{n} x_{i_1}^4 B_{i_1}^4]] \leq M^2 \sum_{i_1=1}^{n} Tr[\mathbb{E}[x_{i_1}^2 B_{i_1}^2]] \leq M^2 \sigma^2.$$

For the second type we have

$$\begin{aligned}
&Tr[\mathbb{E}[\sum_{i_1=1}^{n} \sum_{i_2 \neq i_1} x_{i_1}^2 x_{i_2}^2 B_{i_1} B_{i_2} B_{i_1} B_{i_2}]] \\
\leq &\sum_{i_1=1}^{n} \sum_{i_2=1}^{n} \mathbb{E}[x_{i_1}^2 x_{i_2}^2] |Tr[B_{i_1} B_{i_2} B_{i_1} B_{i_2}]| \\
\leq &\sum_{i_1=1}^{n} \sum_{i_2=1}^{n} \mathbb{E}[x_{i_1}^2 x_{i_2}^2] \|B_{i_1}\|^2 \|B_{i_2}\|^2 \\
\leq &\sum_{i_1=1}^{n} \|B_{i_1}\|^2 \, \mathbb{E}[x_{i_1}^2] \sum_{i_2=1}^{n} \|B_{i_2}\|^2 \, \mathbb{E}[x_{i_2}^2] = \sigma^4.
\end{aligned}$$

The second inequality comes from the fact that $B_{i_1}B_{i_2}B_{i_1}B_{i_2}$ is a rank-matrix whose trace is equal to the operator norm. Similar argument works for the last two types of terms. Summing them up we have

$$Tr[\mathbb{E}\, S_n^4] \leq M^2\sigma^2 + 3\sigma^4.$$

Now we consider general even m. Given some type that has k free subscripts, the sum of terms of such type could be bounded by

$$\sigma^{2k}M^{m-2k}.$$

Moreover, we count the number of types having k free subscripts as follows. In fact, each variable $x_{i_j}, j \in \{1, \dots, m\}$ can bind to one of k free subscripts, which means there are at most k^m types having k free subscripts. Hence we obtain

$$Tr[\mathbb{E}\, S_n^m] \leq \sum_{k=1}^{m/2} k^m \sigma^{2k} M^{m-2k}.$$

Whenever $M \leq \sigma$ the formula reduces to

$$Tr[\mathbb{E}\, S_n^m] \leq (m/2)^{m/2+1}\sigma^m < m^{m/2}\sigma^m$$

where the last inequality is from $2^{m/2+1} > m$ for $m \in \mathbb{N}^+$. In this way, inequality (33) turns to be

$$\Pr[\|S_n\| \geq c\sigma] < \left(\frac{\sqrt{m}}{c}\right)^m.$$

Setting $m = \lfloor \frac{c^2}{e^2} \rfloor$ the desired bound follows immediately.

E.2 Proof of Lemma. 4.3

Since rank-1 matrices include scalars as a special case, we just apply Prop. 4 to $\sum_c x_c$ and $\sum_c x_c \overline{w}_c \boldsymbol{v}_c \boldsymbol{v}_c^T$ respectively, and then lemma follows immediately by setting $c = \Theta(\log n)$ in Prop. 4.