A Proof of Theorem 4.3

To prove the theorem, we need to prove the two conditions are satisfied as required in definition 4.1 of cluster recovery property (CRP), i.e., the optimal solution is *nonzero* and has support indices of columns only *from the same cluster* (SRP - the self-reconstruction property). At the core of our proof lies the primal-dual witness method [32] widely used in lasso analysis, which involves the explicit construction of an optimal primal-dual problem and a dual certificate. We organize the proof as follows:

(1) We prove the *SRP* by duality: first we establish a set of conditions on the optimal dual variable, which corresponds to the case where all the primal solutions satisfy SRP; we then construct a dual variable to certify the proof. This step includes §A.1, §A.2 and §A.3.

(2) We prove the optimal solution is *nonzero* by showing that there exists a value smaller than the value of (A.2) when the solution is zero. This step corresponds to §A.4.

(3) We derive the range of λ for successfully recovering the clustering structure: the proof of step (1) imposes an upper bound on λ while step (2) reduces to a lower bound of λ. It is essential to require the lower bound to be smaller than the upper bound, and based on this we will derive a sufficient condition that guarantees the existence of a valid λ. This step will be explained in §A.5.

A.1 Optimal Primal-Dual Problem

We propose a regularized Dantzig selector and define our primal convex optimization problem as the following general form:

$$\mathbf{P}_0: \ \arg\min_{\boldsymbol{\beta}_i} \ \lambda\|\hat{\boldsymbol{\Sigma}}\boldsymbol{\beta}_i - \hat{\boldsymbol{\gamma}}_i\|_{\infty,\infty} + \|\boldsymbol{\beta}_i\|_1, \tag{A.1}$$

where $\hat{\boldsymbol{\Sigma}} = \mathbf{X}_{\mathcal{S}}^\top \mathbf{X}_{\mathcal{S}}/(T-1)$ and $\hat{\boldsymbol{\gamma}}_i = \mathbf{X}_{\mathcal{S}}^\top (\mathbf{X}_{\mathcal{T}})_{*i}/(T-1)$. Then the dual problem of (A.1) is:

$$\mathbf{D}_0: \ \arg\max_{\boldsymbol{\nu}} \ \langle \boldsymbol{\nu}, \hat{\boldsymbol{\gamma}}_i \rangle \quad \text{subject to} \quad \|\boldsymbol{\nu}\|_1 = \lambda, \ \|\hat{\boldsymbol{\Sigma}}\boldsymbol{\nu}\|_{\infty,\infty} \le 1.$$

To avoid cluttered notations, we omit some subscriptions and use $\hat{\boldsymbol{\gamma}}$ to refer to $\hat{\boldsymbol{\gamma}}_i$ unless stated otherwise, and likewise use $\hat{\boldsymbol{\gamma}}_{\mathcal{S}_l}$ for $(\hat{\boldsymbol{\gamma}}_{\mathcal{S}_l})_i$ in the rest of the proof. We state the following technical lemma on the support of the optimal solution, which extends the Lemma 7.1 in [30]:

Lemma A.1. *Consider a vector $\boldsymbol{\gamma} \in \mathbb{R}^d$ and a matrix $\boldsymbol{\Sigma} \in \mathbb{R}^{d\times d}$. If there exists a pair $(\boldsymbol{\beta}, \boldsymbol{\nu})$ with $\boldsymbol{\beta}$'s support $S \subseteq T$, and a dual certificate vector $\boldsymbol{\nu}$ satisfying:*

$$\boldsymbol{\Sigma}_{S,\eta}\boldsymbol{\nu}_\eta + sgn(\boldsymbol{\beta}_S) = 0, \quad \|\boldsymbol{\nu}\|_1 = \lambda, \quad \|\boldsymbol{\Sigma}_{T\cap S^c,\eta}\boldsymbol{\nu}_\eta\|_{\infty,\infty} \le 1, \quad \|\boldsymbol{\Sigma}_{T^c,\eta}\boldsymbol{\nu}_\eta\|_{\infty,\infty} < 1,$$

where η is the set of indices of entry i such that $|(\boldsymbol{\Sigma}\boldsymbol{\beta} - \boldsymbol{\gamma})_i| = \|\boldsymbol{\Sigma}\boldsymbol{\beta} - \boldsymbol{\gamma}\|_{\infty,\infty}$, then all the optimal solutions $\boldsymbol{\beta}^*$ to (A.1) *obey $\boldsymbol{\beta}^*_{T^c} = 0$.*

The detailed proof of this lemma is deferred to Section B.

A.2 Construction of the Dual Certificate

We define $\hat{\boldsymbol{\Sigma}}_{\mathcal{S}_l,\mathcal{S}_l} = (\mathbf{X}_{\mathcal{S}}^l)^\top \mathbf{X}_{\mathcal{S}}^l \in \mathbb{R}^{d_l\times d_l}$ and $\hat{\boldsymbol{\gamma}}_{\mathcal{S}_l} = (\mathbf{X}_{\mathcal{S}}^l)^\top (\mathbf{X}_{\mathcal{T}}^l)_{*i} \in \mathbb{R}^{d_l}$, and we next apply the above lemma to $\hat{\boldsymbol{\Sigma}}_{\mathcal{S}_l,\mathcal{S}_l}$ and $\hat{\boldsymbol{\gamma}}_{\mathcal{S}_l}$ for a particular cluster $\mathcal{C}_l$, and prove the existence of a pair $(\hat{\boldsymbol{\beta}}, \boldsymbol{\nu})$ such that the optimal solution $\hat{\boldsymbol{\beta}}$ satisfies the self-reconstruction property and the certificate $\boldsymbol{\nu}$ satisfies all the conditions in Lemma A.1. Thereafter, the SRP holds for any optimal solution to (A.1).

To construct the dual certificate, we consider the following problem:

$$\mathbf{P}_1: \ \hat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta}} \ \lambda\|\hat{\boldsymbol{\Sigma}}_{\mathcal{S}_l,\mathcal{S}_l}\boldsymbol{\beta} - \hat{\boldsymbol{\gamma}}_{\mathcal{S}_l}\|_{\infty,\infty} + \|\boldsymbol{\beta}\|_1, \tag{A.2}$$

and its dual form is:

$$\mathbf{D}_1: \ \arg\max_{\boldsymbol{\nu}} \ \langle \boldsymbol{\nu}, \hat{\boldsymbol{\gamma}}_{\mathcal{S}_l} \rangle \quad \text{subject to} \quad \|\boldsymbol{\nu}\|_1 = \lambda, \ \|\hat{\boldsymbol{\Sigma}}_{\mathcal{S}_l,\mathcal{S}_l}\boldsymbol{\nu}\|_{\infty,\infty} \le 1.$$

Define $\hat{\boldsymbol{\beta}}$ to be the optimal solution to the primal problem $\mathbf{P}_1$, the dual problem $\mathbf{D}_1$ is then also feasible by strong duality, which implies that for each optimal solution $\hat{\boldsymbol{\beta}}$ to (A.2) with its support on $\hat{S}$, there exists a dual certificate $\hat{\boldsymbol{\nu}}$ that by definition satisfies:

$$\hat{\boldsymbol{\Sigma}}_{\hat{S},\hat{\eta}}\hat{\boldsymbol{\nu}}_{\hat{\eta}} + sgn(\hat{\boldsymbol{\beta}}_{\hat{S}}) = 0, \quad \|\hat{\boldsymbol{\nu}}\|_1 = \lambda, \quad \|\hat{\boldsymbol{\Sigma}}_{\hat{S}^c,\hat{\eta}}\hat{\boldsymbol{\nu}}_{\hat{\eta}}\|_{\infty,\infty} \leq 1, \tag{A.3}$$

where $\hat{\eta}$ is the set of indices of entry i such that $\left|(\hat{\boldsymbol{\Sigma}}_{\mathcal{S}_l,\mathcal{S}_l}\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\gamma}}_{\mathcal{S}_l})_i\right| = \|\hat{\boldsymbol{\Sigma}}_{\mathcal{S}_l,\mathcal{S}_l}\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\gamma}}_{\mathcal{S}_l}\|_{\infty,\infty}$. If we define:

$$\boldsymbol{\beta} = \left[\mathbf{0}, ..., \mathbf{0}, \hat{\boldsymbol{\beta}}, \mathbf{0}, ..., \mathbf{0}\right] \text{ and } \boldsymbol{\nu} = [\mathbf{0}, ..., \mathbf{0}, \hat{\boldsymbol{\nu}}, \mathbf{0}, ..., \mathbf{0}]\,. \tag{A.4}$$

Note that if we use $\hat{T}$ to index the columns of $\mathbf{X}_{\mathcal{S}}$ from cluster $\mathcal{C}_l$, it is easy to see that the $\boldsymbol{\beta}$ defined above obeys $(\boldsymbol{\beta})_{\hat{T}^c} = \mathbf{0}$ (i.e., the SRP holds), and we have $\hat{T} \cap \hat{S}^c = \hat{S}^c$, which makes the last condition in (A.3) equivalent to the third condition as required in Lemma A.1. Therefore, it remains to check the following condition according to the lemma:

$$\|\hat{\boldsymbol{\Sigma}}_{\mathcal{S}_l^c,\hat{\eta}}\hat{\boldsymbol{\nu}}_{\hat{\eta}}\|_{\infty,\infty} < 1. \tag{A.5}$$

If the condition is satisfied, we can then conclude that $\boldsymbol{\nu}$ as constructed in (A.4) is a dual certificate as required in Lemma A.1, and the SRP holds for all the optimal solutions.

A.3 Upper Bound of λ

In this section, we show that (A.5) can be separated into two terms and bounded individually, we then further impose inequality on the bound to obtain the desired condition, which boils down to an upper bound of λ.

Following (A.5), we have:

$$\begin{aligned}\|\hat{\boldsymbol{\Sigma}}_{\mathcal{S}_l^c,\hat{\eta}}\hat{\boldsymbol{\nu}}_{\hat{\eta}}\|_{\infty,\infty} &= \|(\hat{\boldsymbol{\Sigma}}_{\mathcal{S}_l^c,\hat{\eta}} - \boldsymbol{\Sigma}_{\mathcal{S}_l^c,\hat{\eta}} + \boldsymbol{\Sigma}_{\mathcal{S}_l^c,\hat{\eta}})\hat{\boldsymbol{\nu}}_{\hat{\eta}}\|_{\infty,\infty} \\ &\leq \|(\hat{\boldsymbol{\Sigma}}_{\mathcal{S}_l^c,\hat{\eta}} - \boldsymbol{\Sigma}_{\mathcal{S}_l^c,\hat{\eta}})\hat{\boldsymbol{\nu}}_{\hat{\eta}}\|_{\infty,\infty} + \|\boldsymbol{\Sigma}_{\mathcal{S}_l^c,\hat{\eta}}\hat{\boldsymbol{\nu}}_{\hat{\eta}}\|_{\infty,\infty},\end{aligned} \tag{A.6}$$

where we applied the triangle inequality for the last inequality. To bound the first term, we introduce the following lemma from [15]:

Lemma A.2. *Let $\hat{\boldsymbol{\Sigma}}$ be the sample covariance matrix and $\boldsymbol{\Sigma}$ be the population covariance matrix, when $T \geq max(6\log d, 1)$ for the sample data matrix $\mathbf{X} \in \mathbb{R}^{T\times d}$, we have the following condition hold with probability at least $1 - 6d^{-1}$:*

$$\|\hat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}\|_{\infty,\infty} \leq \frac{16\|\boldsymbol{\Sigma}\|_2 \max_j(\Sigma_{jj})}{\min_j(\Sigma_{jj})(1 - \|\mathbf{A}\|_2)}\sqrt{\frac{6\log d + 4}{T}}. \tag{A.7}$$

Particularly, for the ease of reference, we write the RHS of (A.7) as ρ. Therefore, we have:

$$\|(\hat{\boldsymbol{\Sigma}}_{\mathcal{S}_l^c,\hat{\eta}} - \boldsymbol{\Sigma}_{\mathcal{S}_l^c,\hat{\eta}})\hat{\boldsymbol{\nu}}_{\hat{\eta}}\|_{\infty,\infty} \leq \|\hat{\boldsymbol{\Sigma}}_{\mathcal{S}_l^c,\hat{\eta}} - \boldsymbol{\Sigma}_{\mathcal{S}_l^c,\hat{\eta}}\|_{\infty,\infty} \cdot \|\hat{\boldsymbol{\nu}}_{\hat{\eta}}\|_1 \leq \rho\|\hat{\boldsymbol{\nu}}_{\hat{\eta}}\|_1 \leq \rho\|\hat{\boldsymbol{\nu}}\|_1 \leq \rho\lambda, \tag{A.8}$$

where we applied Lemma A.2 for the second inequality and the second condition in (A.3) for the last inequality. Next we consider the second term in (A.6):

$$\|\boldsymbol{\Sigma}_{\mathcal{S}_l^c,\hat{\eta}}\hat{\boldsymbol{\nu}}_{\hat{\eta}}\|_{\infty,\infty} \leq \|\boldsymbol{\Sigma}_{\mathcal{S}_l^c,\hat{\eta}}\|_{\infty,\infty} \cdot \|\hat{\boldsymbol{\nu}}_{\hat{\eta}}\|_1 \leq \lambda\|\boldsymbol{\Sigma}_{\mathcal{S}_l^c,\hat{\eta}}\|_{\infty,\infty}, \tag{A.9}$$

where the last inequality again follows from the second condition in (A.3). Putting together (A.6), (A.8) and (A.9), we get a sufficient condition for (A.5):

$$\rho\lambda + \lambda\|\boldsymbol{\Sigma}_{\mathcal{S}_l^c,\hat{\eta}}\|_{\infty,\infty} < 1,$$

which implies:

$$\lambda < \frac{1}{\rho + \|\boldsymbol{\Sigma}_{\mathcal{S}_l^c,\hat{\eta}}\|_{\infty,\infty}}.$$

To generalize the above condition to all the clusters $l = 1, 2, \ldots, k$, we further need:

$$\lambda < \min_l \frac{1}{\rho + \|\boldsymbol{\Sigma}_{\mathcal{S}_l^c,\hat{\eta}}\|_{\infty,\infty}}. \tag{A.10}$$

A.4 Lower Bound of λ

In this section, we set to prove the nontrivialness of the optimal solution to (A.2), i.e., the solution is nonzero. We will show that this step reduces to a lower bound of λ, and we state that if λ satisfies the following condition, the optimal solution to (A.2) can not be trivial, i.e., $\boldsymbol{\beta} = 0$ can never be optimal:

$$\lambda > \frac{1}{r(\mathcal{P}((\boldsymbol{\Sigma}_0)_{\mathcal{S}_l,\mathcal{S}_l}))r(\mathcal{P}((\boldsymbol{\Sigma}_1)_{\mathcal{S}_l,\mathcal{S}_l}))(\|\gamma\|_{\infty,\infty} - 2\rho) - \rho}.$$

On the one hand, if $\boldsymbol{\beta} = \mathbf{0}$ is the optimal solution to (A.2), then the optimal value will be $\lambda\|\hat{\gamma}_{\mathcal{S}_l}\|_{\infty,\infty}$; on the other hand, suppose we can choose a particular $\boldsymbol{\beta}^*$ such that:

$$\lambda\|\hat{\boldsymbol{\Sigma}}_{\mathcal{S}_l,\mathcal{S}_l}\boldsymbol{\beta}^* - \hat{\gamma}_{\mathcal{S}_l}\|_{\infty,\infty} + \|\boldsymbol{\beta}^*\|_1 < \lambda\|\hat{\gamma}_{\mathcal{S}_l}\|_{\infty,\infty}. \tag{A.11}$$

If we can prove the existence of such a $\boldsymbol{\beta}^*$, we are then equivalently proving $\boldsymbol{\beta} = \mathbf{0}$ is not optimal. In particular, we let $\boldsymbol{\beta}^*$ be the optimal solution to the following fictitious problem:

$$\boldsymbol{\beta}^* = \arg\min_{\boldsymbol{\beta}} \|\boldsymbol{\beta}\|_1 \quad \text{subject to} \quad \boldsymbol{\Sigma}_{\mathcal{S}_l,\mathcal{S}_l}\boldsymbol{\beta} = \boldsymbol{\gamma}_{\mathcal{S}_l}, \tag{A.12}$$

and its dual is:

$$\arg\max_{\boldsymbol{\nu}} \langle \boldsymbol{\nu}, \boldsymbol{\gamma}_{\mathcal{S}_l}\rangle \quad \text{subject to} \quad \|\boldsymbol{\Sigma}_{\mathcal{S}_l,\mathcal{S}_l}\boldsymbol{\nu}\|_{\infty,\infty} \leq 1. \tag{A.13}$$

Here $\boldsymbol{\Sigma}_{\mathcal{S}_l,\mathcal{S}_l} \in \mathbb{R}^{d_l \times d_l}$ and $\boldsymbol{\gamma}_{\mathcal{S}_l} \in \mathbb{R}^{d_l}$ is a column of $\boldsymbol{\Sigma}_1$. Based on the condition $\boldsymbol{\Sigma}\mathbf{A}^\top = \boldsymbol{\Sigma}_1$, it is valid to assume there exists such a $\boldsymbol{\beta}^*$.

Next, we first deal with the LHS of (A.11):

$$\begin{aligned}
&\lambda\|\hat{\boldsymbol{\Sigma}}_{\mathcal{S}_l,\mathcal{S}_l}\boldsymbol{\beta}^* - \hat{\gamma}_{\mathcal{S}_l}\|_{\infty,\infty} + \|\boldsymbol{\beta}^*\|_1 \\
&= \|\boldsymbol{\beta}^*\|_1 + \lambda\|(\hat{\boldsymbol{\Sigma}}_{\mathcal{S}_l,\mathcal{S}_l} - \boldsymbol{\Sigma}_{\mathcal{S}_l,\mathcal{S}_l} + \boldsymbol{\Sigma}_{\mathcal{S}_l,\mathcal{S}_l})\boldsymbol{\beta}^* - (\hat{\gamma}_{\mathcal{S}_l} - \boldsymbol{\gamma}_{\mathcal{S}_l} + \boldsymbol{\gamma}_{\mathcal{S}_l})\|_{\infty,\infty} \\
&= \|\boldsymbol{\beta}^*\|_1 + \lambda\|(\hat{\boldsymbol{\Sigma}}_{\mathcal{S}_l,\mathcal{S}_l} - \boldsymbol{\Sigma}_{\mathcal{S}_l,\mathcal{S}_l})\boldsymbol{\beta}^* - (\hat{\gamma}_{\mathcal{S}_l} - \boldsymbol{\gamma}_{\mathcal{S}_l})\|_{\infty,\infty} \\
&\leq \|\boldsymbol{\beta}^*\|_1 + \lambda\|(\hat{\boldsymbol{\Sigma}}_{\mathcal{S}_l,\mathcal{S}_l} - \boldsymbol{\Sigma}_{\mathcal{S}_l,\mathcal{S}_l})\boldsymbol{\beta}^*\|_{\infty,\infty} + \lambda\|(\hat{\gamma}_{\mathcal{S}_l} - \boldsymbol{\gamma}_{\mathcal{S}_l})\|_{\infty,\infty} \\
&\leq \|\boldsymbol{\beta}^*\|_1 + \lambda\rho\|\boldsymbol{\beta}^*\|_1 + \lambda\rho \\
&= \|\boldsymbol{\beta}^*\|_1(1 + \rho\lambda) + \rho\lambda,
\end{aligned} \tag{A.14}$$

where the first inequality follows again from the triangle inequality while the second inequality is based on the bound as given in Lemma A.2. To obtain an upper bound of $\boldsymbol{\beta}^*$, we rely on the use of polar set, inradius and circumradius as defined in Section C. Based on the condition in (A.13) and according to definition C.1, we have:

$$\|\boldsymbol{\nu}\|_2 \leq R([\mathcal{P}(\boldsymbol{\Sigma}_{\mathcal{S}_l,\mathcal{S}_l})]^\circ) = \frac{1}{r(\mathcal{P}(\boldsymbol{\Sigma}_{\mathcal{S}_l,\mathcal{S}_l}))}, \tag{A.15}$$

where we applied Lemma C.3 for the equality. Note that, $\boldsymbol{\gamma}_{\mathcal{S}_l}$ is a column of $(\boldsymbol{\Sigma}_1)_{\mathcal{S}_l,\mathcal{S}_l}$ and according to the definition of circumradius, we have:

$$\|\boldsymbol{\gamma}_{\mathcal{S}_l}\|_2 \leq R([\mathcal{P}((\boldsymbol{\Sigma}_1)_{\mathcal{S}_l,\mathcal{S}_l})]) = R([\mathcal{P}((\boldsymbol{\Sigma}_1)_{\mathcal{S}_l,\mathcal{S}_l})]^\circ) = \frac{1}{r(\mathcal{P}((\boldsymbol{\Sigma}_1)_{\mathcal{S}_l,\mathcal{S}_l}))}. \tag{A.16}$$

The first equality holds because $\mathcal{P}(\boldsymbol{\Sigma}_{\mathcal{S}_l,\mathcal{S}_l})$ is convex itself in the first place, and again we applied Lemma C.3 for the last equality. Since (A.12) is a linear program, strong duality holds:

$$\|\boldsymbol{\beta}^*\|_1 = \langle \boldsymbol{\nu}, \boldsymbol{\gamma}_{\mathcal{S}_l}\rangle \leq \|\boldsymbol{\nu}\|_2 \cdot \|\boldsymbol{\gamma}_{\mathcal{S}_l}\|_2 \leq \frac{1}{r(\mathcal{P}(\boldsymbol{\Sigma}_{\mathcal{S}_l,\mathcal{S}_l}))r(\mathcal{P}((\boldsymbol{\Sigma}_1)_{\mathcal{S}_l,\mathcal{S}_l}))}. \tag{A.17}$$

For the last inequality, we applied (A.15) and (A.16). Plugging (A.17) into (A.14), we get:

$$\lambda\|\hat{\boldsymbol{\Sigma}}_{\mathcal{S}_l,\mathcal{S}_l}\boldsymbol{\beta}^* - \hat{\gamma}_{\mathcal{S}_l}\|_{\infty,\infty} + \|\boldsymbol{\beta}^*\|_1 \leq \frac{1 + \rho\lambda}{r(\mathcal{P}(\boldsymbol{\Sigma}_{\mathcal{S}_l,\mathcal{S}_l}))r(\mathcal{P}((\boldsymbol{\Sigma}_1)_{\mathcal{S}_l,\mathcal{S}_l}))} + \rho\lambda. \tag{A.18}$$

Next, we consider the RHS of (A.11):

$$\begin{aligned}\lambda\|\hat{\boldsymbol{\gamma}}_{\mathcal{S}_l}\|_{\infty,\infty} &= \lambda\|\boldsymbol{\gamma}_{\mathcal{S}_l}+\hat{\boldsymbol{\gamma}}_{\mathcal{S}_l}-\boldsymbol{\gamma}_{\mathcal{S}_l}\|_{\infty,\infty}\\ &\geq \lambda\|\boldsymbol{\gamma}_{\mathcal{S}_l}\|_{\infty,\infty}-\lambda\|\hat{\boldsymbol{\gamma}}_{\mathcal{S}_l}-\boldsymbol{\gamma}_{\mathcal{S}_l}\|_{\infty,\infty}\\ &\geq \lambda\|\boldsymbol{\gamma}_{\mathcal{S}_l}\|_{\infty,\infty}-\rho\lambda, \end{aligned} \tag{A.19}$$

where it follows from the triangle inequality and the bound in Lemma A.2 for the first and second inequality, respectively. Therefore, to guarantee there exists a $\boldsymbol{\beta}^*$, by putting together (A.18) and (A.19), a sufficient condition for (A.11) to hold is:

$$\frac{1+\rho\lambda}{r(\mathcal{P}(\boldsymbol{\Sigma}_{\mathcal{S}_l,\mathcal{S}_l}))r(\mathcal{P}((\boldsymbol{\Sigma}_1)_{\mathcal{S}_l,\mathcal{S}_l}))}+\rho\lambda<\lambda\|\boldsymbol{\gamma}_{\mathcal{S}_l}\|_{\infty,\infty}-\rho\lambda,$$

which implies:

$$\lambda>\frac{1}{r(\mathcal{P}(\boldsymbol{\Sigma}_{\mathcal{S}_l,\mathcal{S}_l}))r(\mathcal{P}((\boldsymbol{\Sigma}_1)_{\mathcal{S}_l,\mathcal{S}_l}))(\|\boldsymbol{\gamma}_{\mathcal{S}_l}\|_{\infty,\infty}-2\rho)-\rho}. \tag{A.20}$$

A.5 Existence of λ

We compactly denote $\mathcal{P}_0^l=\mathcal{P}(\boldsymbol{\Sigma}_{\mathcal{S}_l,\mathcal{S}_l})=\mathcal{P}(\mathbb{E}[\boldsymbol{X}_t^l(\boldsymbol{X}_t^l)^\top])$, $\mathcal{P}_1^l=\mathcal{P}((\boldsymbol{\Sigma}_1)_{\mathcal{S}_l,\mathcal{S}_l})=\mathcal{P}(\mathbb{E}[\boldsymbol{X}_t^l(\boldsymbol{X}_{t+1}^l)^\top])$, $r_0^l=r(\mathcal{P}_0^l)$, $r_1^l=r(\mathcal{P}_1^l)$, and $r_0r_1=\min_l r_0^l r_1^l$. To guarantee there exists a λ satisfying both (A.20) and (A.10), the following condition needs to hold:

$$\max_l\frac{1}{r_0^l r_1^l(\|\boldsymbol{\gamma}_{\mathcal{S}_l}\|_{\infty,\infty}-2\rho)-\rho}<\lambda<\min_l\frac{1}{\rho+\|\boldsymbol{\Sigma}_{\mathcal{S}_l^c,\hat{\eta}}\|_{\infty,\infty}}.$$

Also note that:

$$\max_l\frac{1}{r_0^l r_1^l(\|\boldsymbol{\gamma}_{\mathcal{S}_l}\|_{\infty,\infty}-2\rho)-\rho}=\frac{1}{\min_l r_0^l r_1^l(\|\boldsymbol{\gamma}_{\mathcal{S}_l}\|_{\infty,\infty}-2\rho)-\rho}.$$

Therefore, it suffices to require λ obey:

$$\frac{1}{r_0r_1(\|\boldsymbol{\gamma}_{\mathcal{S}_l}\|_{\infty,\infty}-2\rho)-\rho}<\lambda<\frac{1}{\rho+\|\boldsymbol{\Sigma}_{\mathcal{S}_l^c,\mathcal{S}_l}\|_{\infty,\infty}}. \tag{A.21}$$

Note that we further relaxed $\|\boldsymbol{\Sigma}_{\mathcal{S}_l^c,\hat{\eta}}\|_{\infty,\infty}$ to $\|\boldsymbol{\Sigma}_{\mathcal{S}_l^c,\mathcal{S}_l}\|_{\infty,\infty}$ on the RHS, and (A.21) is equivalent to:

$$\rho+\|\boldsymbol{\Sigma}_{\mathcal{S}_l^c,\mathcal{S}_l}\|_{\infty,\infty}<r_0r_1(\|\boldsymbol{\gamma}_{\mathcal{S}_l}\|_{\infty,\infty}-2\rho)-\rho,$$

which implies that:

$$r_0r_1>\frac{\|\boldsymbol{\Sigma}_{\mathcal{S}_l^c,\mathcal{S}_l}\|_{\infty,\infty}+2\rho}{\|\boldsymbol{\gamma}_{\mathcal{S}_l}\|_{\infty,\infty}-2\rho}, \tag{A.22}$$

with

$$\rho=\frac{16\|\boldsymbol{\Sigma}\|_2\max_j\Sigma_{jj}}{\min_j\Sigma_{jj}(1-\|\mathbf{A}\|_2)}\sqrt{\frac{6\log d+4}{T}}.$$

Here (A.22) is a sufficient condition to guarantee that the range given in (A.21) is non-empty. This concludes the proof of Theorem 4.3.

B Proof of Lemma A.1

Proof. Observe that for any optimal solution $\boldsymbol{\beta}^*$ to problem (A.1), we have:

$$\begin{aligned}
&\|\boldsymbol{\beta}^*\|_1 + \lambda\|\boldsymbol{\Sigma}\boldsymbol{\beta}^* - \boldsymbol{\gamma}\|_{\infty,\infty} - \|\boldsymbol{\beta}\|_1 - \lambda\|\boldsymbol{\Sigma}\boldsymbol{\beta} - \boldsymbol{\gamma}\|_{\infty,\infty} \\
&= \|\boldsymbol{\beta}^*_S\|_1 + \|\boldsymbol{\beta}^*_{T\cap S^c}\|_1 + \|\boldsymbol{\beta}^*_{T^c}\|_1 + \lambda\|\boldsymbol{\Sigma}\boldsymbol{\beta}^* - \boldsymbol{\gamma}\|_{\infty,\infty} - \|\boldsymbol{\beta}\|_1 - \lambda\|\boldsymbol{\Sigma}\boldsymbol{\beta} - \boldsymbol{\gamma}\|_{\infty,\infty} \\
&= \langle sgn(\boldsymbol{\beta}_S), \boldsymbol{\beta}^*_S - \boldsymbol{\beta}_S\rangle + \|\boldsymbol{\beta}^*_{T\cap S^c}\|_1 + \|\boldsymbol{\beta}^*_{T^c}\|_1 + \lambda\|\boldsymbol{\Sigma}\boldsymbol{\beta}^* - \boldsymbol{\gamma}\|_{\infty,\infty} - \lambda\|\boldsymbol{\Sigma}\boldsymbol{\beta} - \boldsymbol{\gamma}\|_{\infty,\infty} \\
&\geq \langle sgn(\boldsymbol{\beta}_S), \boldsymbol{\beta}^*_S - \boldsymbol{\beta}_S\rangle + \|\boldsymbol{\beta}^*_{T\cap S^c}\|_1 + \|\boldsymbol{\beta}^*_{T^c}\|_1 + \langle \boldsymbol{\Sigma}_{*,\eta}\boldsymbol{\nu}_\eta, \boldsymbol{\beta}^* - \boldsymbol{\beta}\rangle \\
&= \langle -\boldsymbol{\Sigma}_{S,\eta}\boldsymbol{\nu}_\eta, \boldsymbol{\beta}^*_S - \boldsymbol{\beta}_S\rangle + \|\boldsymbol{\beta}^*_{T\cap S^c}\|_1 + \|\boldsymbol{\beta}^*_{T^c}\|_1 + \langle \boldsymbol{\Sigma}_{*,\eta}\boldsymbol{\nu}_\eta, \boldsymbol{\beta}^* - \boldsymbol{\beta}\rangle \\
&= \|\boldsymbol{\beta}^*_{T\cap S^c}\|_1 + \|\boldsymbol{\beta}^*_{T^c}\|_1 + \langle \boldsymbol{\Sigma}_{*,\eta}\boldsymbol{\nu}_\eta, \boldsymbol{\beta}^*\rangle - \langle \boldsymbol{\Sigma}_{S,\eta}\boldsymbol{\nu}_\eta, \boldsymbol{\beta}^*_S\rangle + \langle \boldsymbol{\Sigma}_{S,\eta}\boldsymbol{\nu}_\eta, \boldsymbol{\beta}_S\rangle - \langle \boldsymbol{\Sigma}_{*,\eta}\boldsymbol{\nu}_\eta, \boldsymbol{\beta}\rangle \\
&= \|\boldsymbol{\beta}^*_{T\cap S^c}\|_1 + \|\boldsymbol{\beta}^*_{T^c}\|_1 + \left\langle \boldsymbol{\nu}_\eta, \boldsymbol{\Sigma}^\top_{*,\eta}\boldsymbol{\beta}^*\right\rangle - \left\langle \boldsymbol{\nu}_\eta, \boldsymbol{\Sigma}^\top_{S,\eta}\boldsymbol{\beta}^*_S\right\rangle + \left\langle \boldsymbol{\nu}_\eta, \boldsymbol{\Sigma}^\top_{S,\eta}\boldsymbol{\beta}_S\right\rangle - \left\langle \boldsymbol{\nu}_\eta, \boldsymbol{\Sigma}^\top_{*,\eta}\boldsymbol{\beta}\right\rangle \\
&= \|\boldsymbol{\beta}^*_{T\cap S^c}\|_1 + \left\langle \boldsymbol{\nu}_\eta, \boldsymbol{\Sigma}^\top_{T\cap S^c,\eta}\boldsymbol{\beta}^*_{T\cap S^c}\right\rangle + \|\boldsymbol{\beta}^*_{T^c}\|_1 + \left\langle \boldsymbol{\nu}_\eta, \boldsymbol{\Sigma}^\top_{T^c,\eta}\boldsymbol{\beta}^*_{T^c}\right\rangle.
\end{aligned}$$

The first inequality holds because of the convexity of the original optimization function. The last equality holds because :

$$\boldsymbol{\Sigma}\boldsymbol{\beta}^* = \boldsymbol{\Sigma}^\top_{S,*}\boldsymbol{\beta}^*_S + \boldsymbol{\Sigma}^\top_{T\cap S^c,*}\boldsymbol{\beta}^*_{T\cap S^c} + \boldsymbol{\Sigma}^\top_{T^c,*}\boldsymbol{\beta}^*_{T^c}, \quad \text{and} \quad \boldsymbol{\Sigma}\boldsymbol{\beta} = \boldsymbol{\Sigma}^\top_{S,*}\boldsymbol{\beta}_S.$$

With the third inequality condition stated in Lemma A.1, we have:

$$\left\langle \boldsymbol{\nu}_\eta, \boldsymbol{\Sigma}^\top_{T\cap S^c,\eta}\boldsymbol{\beta}^*_{T\cap S^c}\right\rangle = \langle \boldsymbol{\Sigma}_{T\cap S^c,\eta}\boldsymbol{\nu}_\eta, \boldsymbol{\beta}^*_{T\cap S^c}\rangle \leq \|\boldsymbol{\Sigma}_{T\cap S^c,\eta}\boldsymbol{\nu}_\eta\|_{\infty,\infty} \cdot \|\boldsymbol{\beta}^*_{T\cap S^c}\|_1 \leq \|\boldsymbol{\beta}^*_{T\cap S^c}\|_1,$$

which implies:

$$\|\boldsymbol{\beta}^*_{T\cap S^c}\|_1(1 - \|\boldsymbol{\Sigma}_{T\cap S^c,\eta}\boldsymbol{\nu}_\eta\|_{\infty,\infty}) \geq 0. \tag{B.1}$$

In a similar manner, we have $\left\langle \boldsymbol{\nu}_\eta, \boldsymbol{\Sigma}^\top_{T^c,\eta}\boldsymbol{\beta}^*_{T^c}\right\rangle \leq \|\boldsymbol{\Sigma}_{T^c,\eta}\boldsymbol{\nu}_\eta\|_{\infty,\infty} \cdot \|\boldsymbol{\beta}^*_{T^c}\|_1$. We also have $\|\boldsymbol{\beta}_S\|_1 = \|\boldsymbol{\beta}\|_1$ since $\boldsymbol{\beta}$ is supported on S, and therefore we get:

$$\begin{aligned}
&\|\boldsymbol{\beta}^*\|_1 + \lambda\|\boldsymbol{\Sigma}\boldsymbol{\beta}^* - \boldsymbol{\gamma}\|_{\infty,\infty} - \|\boldsymbol{\beta}\|_1 - \lambda\|\boldsymbol{\Sigma}\boldsymbol{\beta} - \boldsymbol{\gamma}\|_{\infty,\infty} \\
&\geq \|\boldsymbol{\beta}^*_{T\cap S^c}\|_1 + \left\langle \boldsymbol{\nu}_\eta, \boldsymbol{\Sigma}^\top_{T\cap S^c,\eta}\boldsymbol{\beta}^*_{T\cap S^c}\right\rangle + \|\boldsymbol{\beta}^*_{T^c}\|_1 + \left\langle \boldsymbol{\nu}_\eta, \boldsymbol{\Sigma}^\top_{T^c,\eta}\boldsymbol{\beta}^*_{T^c}\right\rangle \\
&\geq \|\boldsymbol{\beta}^*_{T\cap S^c}\|_1 - \left\langle \boldsymbol{\nu}_\eta, \boldsymbol{\Sigma}^\top_{T\cap S^c,\eta}\boldsymbol{\beta}^*_{T\cap S^c}\right\rangle + \|\boldsymbol{\beta}^*_{T^c}\|_1 - \left\langle \boldsymbol{\nu}_\eta, \boldsymbol{\Sigma}^\top_{T^c,\eta}\boldsymbol{\beta}^*_{T^c}\right\rangle \\
&= \|\boldsymbol{\beta}^*_{T\cap S^c}\|_1(1 - \|\boldsymbol{\Sigma}_{T\cap S^c,\eta}\boldsymbol{\nu}_\eta\|_{\infty,\infty}) + \|\boldsymbol{\beta}^*_{T^c}\|_1(1 - \|\boldsymbol{\Sigma}_{T^c,\eta}\boldsymbol{\nu}_\eta\|_{\infty,\infty}) \\
&\geq \|\boldsymbol{\beta}^*_{T^c}\|_1(1 - \|\boldsymbol{\Sigma}_{T^c,\eta}\boldsymbol{\nu}_\eta\|_{\infty,\infty}),
\end{aligned}$$

where the last inequality follows from the inequality in (B.1). Furthermore, based on the last inequality condition stated in Lemma A.1, it is easy to see that $(1 - \|\boldsymbol{\Sigma}_{T^c,\eta}\boldsymbol{\nu}_\eta\|_{\infty,\infty})$ is strictly greater than 0. On the other side, using the fact that $\boldsymbol{\beta}^*$ is the optimal solution, $\|\boldsymbol{\beta}^*\|_1 + \lambda\|\boldsymbol{\Sigma}\boldsymbol{\beta}^* - \boldsymbol{\gamma}\|_{\infty,\infty} - \|\boldsymbol{\beta}\|_1 - \lambda\|\boldsymbol{\Sigma}\boldsymbol{\beta} - \boldsymbol{\gamma}\|_{\infty,\infty} \leq 0$. Therefore it follows that $\boldsymbol{\beta}^*_{T^c} = 0$. We conclude the proof. □

C Auxiliary Results

We state some necessary definitions and lemmas in this section.

Definition C.1 (Polar Set)**.** *The polar set $\mathcal{K}^\circ$ of set $\mathcal{K} \in \mathbb{R}^d$ is defined as:*

$$\mathcal{K}^\circ = \left\{\boldsymbol{y} \in \mathbb{R}^d \mid \langle \boldsymbol{x}, \boldsymbol{y}\rangle \leq 1 \textit{ for all } \boldsymbol{x} \in \mathcal{K}\right\}.$$

Definition C.2 (Circumradius [30])**.** *The circumradius of a convex body $\mathcal{P}$, denoted by $R(\mathcal{P})$, is defined as the radius of the smallest ball containing $\mathcal{P}$.*

We will also leverage the following lemma:

Lemma C.3 ([5], page 448)**.** *For a symmetric convex body $\mathcal{P}$, i.e., $\mathcal{P} = -\mathcal{P}$, the following relationship between the inradius of $\mathcal{P}$ and the circumradius of its polar $\mathcal{P}^\circ$ holds:*

$$r(\mathcal{P})R(\mathcal{P}^\circ) = 1.$$