A. Proof for Sparse Principal Submatrix Estimation

Before we establish the proof of Theorem 1, we present a corollary of Theorem 2.2 of [8]. Let $\overline{\beta}$ be a quantity that scales with s^* and d. It establishes the sufficient conditions under which distinguishing $\beta^* = 0$ and $\beta^* = \overline{\beta}$ is impossible. Recall $\mathcal{P}(s^*, d)$ denotes the distribution family specified in §2.1.

Corollary 8. We consider testing $H_0 : \beta_0^* = 0$ against $H_1 : \beta_1^* = \overline{\beta}$. For any test $\phi : \mathbb{R}^{d\times d} \to \{0, 1\}$ based on $\mathbf{X}$, if $\overline{\beta}^2 (s^*)^4/d^2 = o(1)$ and $\limsup \overline{\beta}^2 s^*/\log(d/s^*) < C$, there exist $\mathbb{P}_0, \mathbb{P}_1 \in \mathcal{P}(s^*, d)$, which correspond to H_0 and H_1, such that

$$\inf_{\phi} \max\bigl\{\mathbb{P}_0(\phi = 1),\ \mathbb{P}_1(\phi = 0)\bigr\} \geq 1/4.$$

Here $C > 0$ is an absolute constant.

Proof. Theorem 2.2 of [8] gives a similar result for $\mathbf{X}$ with Gaussian entries. Therefore, their $\mathbb{P}_0$ and $\mathbb{P}_1$ fall within $\mathcal{P}(s^*, d)$ specified in §2.1 up to rescaling of variance. Besides, it is worth noting that [8] do not assume $\mathbf{X}$ is symmetric. Nevertheless, the proof for symmetric $\mathbf{X}$ follows similarly from their proof. □

Equipped with Corollary 8, we are now ready to prove Theorem 1.

Proof of Theorem 1. We consider testing hypotheses $H_0 : \beta_0^* = 0$ and $H_1 : \beta_1^* = \overline{\beta}$ with

$$\overline{\beta} = C\sqrt{1/s^* \cdot \log(d/s^*)}, \tag{A.1}$$

where C is an absolute constant that is sufficiently small. By Corollary 8, there exist $\mathbb{P}_0, \mathbb{P}_1 \in \mathcal{P}(s^*, d)$ corresponding to H_0 and H_1, such that for any test $\phi : \mathbb{R}^{d\times d} \to \{0, 1\}$,

$$\inf_{\phi} \max\bigl\{\mathbb{P}_0(\phi = 1),\ \mathbb{P}_1(\phi = 0)\bigr\} \geq 1/4, \tag{A.2}$$

$$\text{for } \overline{\beta}(s^*)^2/d = o(1).$$

We consider a specific test $\overline{\phi}(\widehat{\beta})$ based on $\widehat{\beta}$, which is defined as $\overline{\phi}(\widehat{\beta}) = \mathbb{1}(\widehat{\beta} > \overline{\beta}/2)$. From (A.2) we have

$$\begin{aligned}
&\inf_{\widehat{\beta}} \max\Bigl\{\mathbb{P}_0\bigl(|\widehat{\beta} - \beta_0^*| \geq \overline{\beta}/2\bigr),\ \mathbb{P}_1\bigl(|\widehat{\beta} - \beta_1^*| \geq \overline{\beta}/2\bigr)\Bigr\}\\
&= \inf_{\widehat{\beta}} \max\Bigl\{\mathbb{P}_0\bigl(|\widehat{\beta}| \geq \overline{\beta}/2\bigr),\ \mathbb{P}_1\bigl(|\widehat{\beta} - \overline{\beta}| \geq \overline{\beta}/2\bigr)\Bigr\}\\
&\geq \inf_{\widehat{\beta}} \max\Bigl\{\mathbb{P}_0\bigl[\overline{\phi}(\widehat{\beta}) = 1\bigr],\ \mathbb{P}_1\bigl[\overline{\phi}(\widehat{\beta}) = 0\bigr]\Bigr\}\\
&\geq \inf_{\phi} \max\bigl\{\mathbb{P}_0(\phi = 1),\ \mathbb{P}_1(\phi = 0)\bigr\} \geq 1/4.
\end{aligned} \tag{A.3}$$

Here the first inequality holds because under H_0, $\overline{\phi}(\widehat{\beta}) = 1$ implies $|\widehat{\beta} - \beta_0^*| \geq \overline{\beta}/2$ by definition and under H_1, $\overline{\phi}(\widehat{\beta}) = 0$ implies $|\widehat{\beta} - \beta_1^*| \geq \overline{\beta}/2$. Here the second last inequality holds because $\overline{\phi}(\widehat{\beta})$ is a specific class of tests. Consequently, we have

$$\begin{aligned}
&\inf_{\widehat{\beta}} \sup_{\mathbb{P}\in\mathcal{P}(s^*,d)} \mathbb{E}_{\mathbb{P}}|\widehat{\beta} - \beta^*|\\
&\geq \inf_{\widehat{\beta}} \max\Bigl\{\mathbb{E}_{\mathbb{P}_0}|\widehat{\beta} - \beta_0^*|,\ \mathbb{E}_{\mathbb{P}_1}|\widehat{\beta} - \beta_1^*|\Bigr\}\\
&\geq \overline{\beta}/2 \cdot \inf_{\widehat{\beta}} \max\Bigl\{\mathbb{P}_0\bigl(|\widehat{\beta} - \beta_0^*| \geq \overline{\beta}/2\bigr),\\
&\qquad \mathbb{P}_1\bigl(|\widehat{\beta} - \beta_1^*| \geq \overline{\beta}/2\bigr)\Bigr\} \geq \overline{\beta}/8,
\end{aligned} \tag{A.4}$$

where the second inequality is from Markov's inequality and the last is from (A.3). By plugging (A.1) into (A.4), we reach the conclusion. □

In the sequel, we prove the upper bound in Proposition 2.

Proof of Proposition 2. For integer $s > 0$, we denote by $\mathcal{V}_s$ the set of $\mathbf{v} \in \mathbb{R}^d$ with exactly s entries being one and the others being zero. By definition, in (2.1) we have

$$\sup_{\substack{\mathcal{S}\subseteq[d]\\ |\mathcal{S}|=s^*}} \sum_{(i,j)\in\mathcal{S}\times\mathcal{S}} X_{i,j} = \sup_{\mathbf{v}\in\mathcal{V}_{s^*}} \mathbf{v}^\top \mathbf{X}\mathbf{v}/2. \tag{A.5}$$

Recall that by our definition we have $X_{i,i} = 0$ for all $i \in [d]$ and $\mathbb{E}\mathbf{X} = \mathbf{\Theta}$. Note that

$$\Bigl|\sup_{\mathbf{v}\in\mathcal{V}_{s^*}} \mathbf{v}^\top \mathbf{X}\mathbf{v} - \sup_{\mathbf{v}\in\mathcal{V}_{s^*}} \mathbf{v}^\top \mathbf{\Theta}\mathbf{v}\Bigr| \leq \sup_{\mathbf{v}\in\mathcal{V}_{s^*}} \bigl|\mathbf{v}^\top(\mathbf{X} - \mathbf{\Theta})\mathbf{v}\bigr|. \tag{A.6}$$

Since $\mathbf{X} \sim \mathbb{P} \in \mathcal{P}(s^*, d)$, for any fixed $\mathbf{v} \in \mathcal{V}_{s^*}$, $\mathbf{v}^\top(\mathbf{X} - \mathbf{\Theta})\mathbf{v}$ is twice the summation of $s^*(s^* - 1)/2$ independent sub-Gaussian random variables that have mean zero and ψ_2-norm at most one. Hence, for any fixed $\mathbf{v} \in \mathcal{V}_{s^*}$ we have

$$\mathbb{P}\bigl[\bigl|\mathbf{v}^\top(\mathbf{X} - \mathbf{\Theta})\mathbf{v}\bigr| > t\bigr] < \exp\bigl\{1 - Ct^2/[s^*(s^* - 1)]\bigr\}.$$

Then by union bound, we have

$$\begin{aligned}
&\mathbb{P}\Bigl[\sup_{\mathbf{v}\in\mathcal{V}_{s^*}} \bigl|\mathbf{v}^\top(\mathbf{X} - \mathbf{\Theta})\mathbf{v}\bigr| > t\Bigr]\\
&\leq \binom{d}{s^*} \exp\bigl\{1 - Ct^2/[s^*(s^* - 1)]\bigr\}\\
&\leq \exp\bigl\{1 - Ct^2/[s^*(s^* - 1)] + s^* \log(d/s^*)\bigr\}.
\end{aligned}$$

Setting the right-hand side to be δ, we obtain

$$t = C\sqrt{\log(e/\delta) + s^* \log(d/s^*)} \cdot \sqrt{s^*(s^* - 1)}. \tag{A.7}$$

Plugging (A.7) into (A.6), we have that with probability at least $1 - \delta$,

$$\begin{aligned}
&\Bigl|\sup_{\mathbf{v}\in\mathcal{V}_{s^*}} \mathbf{v}^\top \mathbf{X}\mathbf{v} - \sup_{\mathbf{v}\in\mathcal{V}_{s^*}} \mathbf{v}^\top \mathbf{\Theta}\mathbf{v}\Bigr|\\
&\leq C\sqrt{\log(e/\delta) + s^* \log(d/s^*)} \cdot \sqrt{s^*(s^* - 1)}.
\end{aligned}$$

Note that $\sup_{\mathbf{v}\in\mathcal{V}_{s^*}} \mathbf{v}^\top \mathbf{\Theta}\mathbf{v} = s^*(s^* - 1) \cdot \beta^*$. Then by (2.1) and (A.5) we obtain that

$$\bigl|\widehat{\beta}^{\text{scan}} - \beta^*\bigr| \leq C\sqrt{\log(e/\delta) + s^* \log(d/s^*)}\big/\sqrt{s^*(s^* - 1)}$$

holds with probability at least $1 - \delta$. Setting $\delta = 1/d$, we reach the conclusion. $\square$

In the following we prove Proposition 4.

Proof of Proposition 4. We have

$$\begin{aligned}
&\mathbb{P}\big(\big|\widehat{\beta}^{\max} - \beta^*\big| \geq t\big) \qquad \text{(A.8)}\\
&\leq \mathbb{P}\Big(\Big|\sup_{i,j\in[d]} X_{i,j} - \sup_{i,j\in[d]} \Theta_{i,j}\Big| \geq t\Big)\\
&\leq \mathbb{P}\Big(\sup_{i,j\in[d]} |X_{i,j} - \Theta_{i,j}| \geq t\Big)\\
&\leq d^2 \cdot \mathbb{P}(|X_{i,j} - \Theta_{i,j}| \geq t),
\end{aligned}$$

where the last inequality follows from union bound. Since $\mathbb{E}X_{i,j} = \Theta_{i,j}$, we have $\mathbb{E}(X_{i,j} - \Theta_{i,j}) = 0$. Moreover, we know that $\|X_{i,j} - \Theta_{i,j}\|_{\psi_2} \leq 1$. By the definition of sub-Gaussian random variable, we have

$$\mathbb{P}(|X_{i,j} - \Theta_{i,j}| \geq t) \leq \exp\big(1 - Ct^2\big). \qquad \text{(A.9)}$$

Substituting (A.9) into (A.8), we obtain

$$\begin{aligned}
\mathbb{P}\big(\big|\widehat{\beta}^{\max} - \beta^*\big| \geq t\big) &\leq d^2 \exp\big(1 - Ct^2\big) \qquad \text{(A.10)}\\
&= \exp\big(1 - Ct^2 + 2\log d\big).
\end{aligned}$$

Setting the right hand side of (A.10) to be $1/d$, and solving for t, we obtain with probability at least $1 - 1/d$ that

$$\big|\widehat{\beta}^{\max} - \beta^*\big| \leq C\sqrt{\log d}.$$

This completes the proof. $\square$

B. Proof for Stochastic Block Model

In this section, we present the detailed proofs of the main results for edge probability estimation in stochastic block model. We need the following lemma from [1], which provides the sufficient conditions under which the hypotheses $H_0 : \beta_0^* = p_0$ and $H_1 : \beta_1^* = p_1$ are not distinguishable. Recall $\mathbf{A}$ denotes the adjacency matrix and $\mathcal{P}(s^*, d)$ denotes the distribution family specified in §2.2.

Lemma 9. We consider testing $H_0 : \beta_0^* = p_0$ against $H_1 : \beta_1^* = p_1$. For any test $\phi : \mathbb{R}^{d\times d} \to \{0, 1\}$ based on $\mathbf{A}$, assuming $(s^*)^2(p_1 - p_0)/(\sqrt{p_0}d) = o(1)$, $\limsup(p_1 - p_0)^2 s^*/[4p_0(1 - p_0)\log(d/s^*)] < 1$ and $\log(d/s^*)/(s^*p_0) = o(1)$, we have

$$\inf_{\phi} \max\big\{\mathbb{P}_0(\phi = 1),\ \mathbb{P}_1(\phi = 0)\big\} \geq 1/4,$$

where $\mathbb{P}_0, \mathbb{P}_1 \in \mathcal{P}(s^*, d)$ are distributions corresponding to H_0 and H_1.

Now we are ready to lay out the proof of Theorem 5.

Proof of Theorem 5. The proof strategy is similar to Theorem 1. In the sequel, we assume $\widetilde{\beta}^*$ is known, since the obtained lower bound implies the lower bound for unknown $\widetilde{\beta}^*$. We invoke Lemma 9 with $p_0 = \widetilde{\beta}^*$ and $p_1 = \widetilde{\beta}^* + \overline{\beta}$, where

$$\overline{\beta} = C\sqrt{1/s^* \cdot \log(d/s^*)}. \qquad \text{(B.1)}$$

Then we have that for any test $\phi : \mathbb{R}^{d\times d} \to \{0, 1\}$ based on the adjacency matrix $\mathbf{A}$, it holds that

$$\inf_{\phi} \max\big\{\mathbb{P}_0(\phi = 1),\ \mathbb{P}_1(\phi = 0)\big\} \geq 1/4, \qquad \text{(B.2)}$$

for $(s^*)^2\overline{\beta}/d = o(1)$ and $\log(d/s^*)/\big(s^*\widetilde{\beta}^*\big) = o(1)$.

It is easy to verify the conditions in (B.2) are implied by the conditions of Theorem 5 and (B.1). Following the derivation of (A.4) in the proof of Theorem 1, we consider a specific test $\overline{\phi}\big(\widehat{\beta}\big)$ based on $\widehat{\beta}$, which is defined as $\overline{\phi}\big(\widehat{\beta}\big) = \mathbb{1}\big(\widehat{\beta} > \widetilde{\beta}^* + \overline{\beta}/2\big)$. We have

$$\begin{aligned}
&\inf_{\widehat{\beta}} \max\Big\{\mathbb{P}_0\big(\big|\widehat{\beta} - \beta_0^*\big| \geq \overline{\beta}/2\big),\ \mathbb{P}_1\big(\big|\widehat{\beta} - \beta_1^*\big| \geq \overline{\beta}/2\big)\Big\}\\
&= \inf_{\widehat{\beta}} \max\Big\{\mathbb{P}_0\big(\big|\widehat{\beta} - \widetilde{\beta}^*\big| \geq \overline{\beta}/2\big), \mathbb{P}_1\big(\big|\widehat{\beta} - \widetilde{\beta}^* - \overline{\beta}\big| \geq \overline{\beta}/2\big)\Big\}\\
&\geq \inf_{\widehat{\beta}} \max\Big\{\mathbb{P}_0\big[\overline{\phi}\big(\widehat{\beta}\big) = 1\big],\ \mathbb{P}_1\big[\overline{\phi}\big(\widehat{\beta}\big) = 0\big]\Big\}\\
&\geq \inf_{\phi} \max\big\{\mathbb{P}_0(\phi = 1),\ \mathbb{P}_1(\phi = 0)\big\} \geq 1/4, \qquad \text{(B.3)}
\end{aligned}$$

where the equality is obtained by plugging β_0^* and β_1^*. The first inequality holds becasuse $\overline{\phi}\big(\widehat{\beta}\big) = 1$ implies $\big|\widehat{\beta} - \widetilde{\beta}^*\big| \geq \overline{\beta}/2$, and $\overline{\phi}\big(\widehat{\beta}\big) = 0$ implies $\big|\widehat{\beta} - \widetilde{\beta}^* - \overline{\beta}\big| \geq \overline{\beta}/2$. From (B.3) we obtain

$$\begin{aligned}
&\inf_{\widehat{\beta}} \sup_{\mathbb{P}\in\mathcal{P}(s^*,d)} \mathbb{E}_{\mathbb{P}}\big|\widehat{\beta} - \beta^*\big|\\
&\geq \inf \max\Big\{\mathbb{E}_{\mathbb{P}_0}\big|\widehat{\beta} - \beta_0^*\big|,\ \mathbb{E}_{\mathbb{P}_1}\big|\widehat{\beta} - \beta_1^*\big|\Big\}\\
&\geq \overline{\beta}/2 \cdot \inf_{\widehat{\beta}} \max\Big\{\mathbb{P}_0\big(\big|\widehat{\beta} - \beta_0^*\big| \geq \overline{\beta}/2\big),\\
&\qquad \mathbb{P}_1\big(\big|\widehat{\beta} - \beta_1^*\big| \geq \overline{\beta}/2\big)\Big\} \geq \overline{\beta}/8,
\end{aligned}$$

where $\overline{\beta}$ is defined in (B.1), the second inequality follows from Markov's inequality. This concludes the proof. $\square$

In the following, we prove Proposition 6.

Proof of Proposition 6. The proof is similar to Proposition 2. We only need to note that $\mathbf{A} - \mathbb{E}[\mathbf{A}]$ is a symmetric matrix, whose entires within the upper-right triangle are independently sub-Gaussian and satisfy

$$\|A_{i,j} - \mathbb{E}A_{i,j}\|_{\psi_2} \leq 1, \quad \text{for all } i < j,$$

since $A_{i,j}$ is Bernoulli and $|A_{i,j} - \mathbb{E}A_{i,j}| \leq 1$. Then replacing $\mathbf{X}$ with $\mathbf{A}$ in the proof of Proposition 2, we reach the conclusion. $\square$

In the following, we lay out the proof of Theorem 7.

Proof of Theorem 7. We consider a specific distribution in $\mathcal{P}(s^*, d)$ under which the edge probability $\beta^* = \widetilde{\beta}^* = 1/2$. Let $\overline{\mathbf{A}} = \mathbf{A} + \mathbf{I}_d$. Under such a distribution, we construct a matrix $\mathbf{\Pi}^{(\ell)} \in \mathbb{R}^{d^{(\ell)}\times d^{(\ell)}}$, which is a feasible solution to the ℓ-th level SoS optimization problem with high probability. Then we prove the objective value corresponding to $\mathbf{\Pi}^{(\ell)}$ is one, which implies that the maximum of the respective SoS program is at least one with high probability.

Different from the proof of Theorem 3, we define the expansivity $\eta\big(\mathcal{S}, \overline{\mathbf{A}}\big)$ of $\mathcal{S} \subseteq [d]$ as the number of sets $\mathcal{S}' \subseteq [d]$ satisfying $|\mathcal{S}'| = 2\ell$, $\mathcal{S} \subseteq \mathcal{S}'$ and $\overline{\mathbf{A}}_{\mathcal{S}',\mathcal{S}'} = \mathbf{1}_{2\ell,2\ell}$. Note $\eta\big(\mathcal{S}, \overline{\mathbf{A}}\big)$ is nonzero only if $\overline{\mathbf{A}}_{\mathcal{S},\mathcal{S}} = \mathbf{1}_{|\mathcal{S}|,|\mathcal{S}|}$. Therefore, $\eta\big(\mathcal{S}, \overline{\mathbf{A}}\big)$ gives the number of $\overline{\mathbf{A}}$'s submatrices which are extended from $\overline{\mathbf{A}}_{\mathcal{S},\mathcal{S}}$ and have size $2\ell \times 2\ell$ with all entries being one. Recall that each entry $\Pi^{(\ell)}_{\mathcal{C}_1,\mathcal{C}_2}$ of $\mathbf{\Pi}^{(\ell)}$ are indexed by collections $\mathcal{C}_1$ and $\mathcal{C}_2$, and $M(\mathcal{C}_1)$ and $M(\mathcal{C}_1)$ are the corresponding sets, which have distinct elements. Similar to (5.1), we construct each entry of $\mathbf{\Pi}^{(\ell)}$ as

$$\Pi^{(\ell)}_{\mathcal{C}_1,\mathcal{C}_2} = \frac{\eta\big[M(\mathcal{C}_1) \cup M(\mathcal{C}_2), \overline{\mathbf{A}}\big]}{\eta\big(\emptyset, \overline{\mathbf{A}}\big)} \cdot \frac{s^*!/[s^* - |M(\mathcal{C}_1) \cup M(\mathcal{C}_2)|]!}{(2\ell)!/[2\ell - |M(\mathcal{C}_1) \cup M(\mathcal{C}_2)|]!}. \tag{B.4}$$

Note that the construction of $\Pi^{(\ell)}_{\mathcal{C}_1,\mathcal{C}_2}$ is exactly the same as (5.1), except that we replace $\overline{\mathbf{X}}$ with $\overline{\mathbf{A}}$. Also, by the same calculation as in the proof of Theorem 3, we can verify $\mathbf{\Pi}^{(\ell)}$ defined in (B.4) satisfies the constraints of ℓ-th level SoS optimization problem.

Next we calculate the value of objective function corresponding to $\mathbf{\Pi}^{(\ell)}$. Note that

$$\sum_{i,j=1}^{d} \overline{A}_{i,j} \Pi^{(\ell)}_{\{i\},\{j\}} = \sum_{i,j=1}^{d} \mathbb{1}\big(\overline{A}_{i,j} = 1\big) \cdot \Pi^{(\ell)}_{\{i\},\{j\}} = \sum_{i,j=1}^{d} \Pi^{(\ell)}_{\{i\},\{j\}}.$$

Here both equalities hold because according to the definition of $\eta\big(\cdot, \overline{\mathbf{A}}\big)$, it holds that $\eta\big(\{i,j\}, \overline{\mathbf{A}}\big) = 0$ for $\overline{A}_{i,j} \neq 1$, which implies $\Pi^{(\ell)}_{\{i\},\{j\}} = 0$ correspondingly. Moreover, we have

$$\sum_{i,j=1}^{d} \Pi^{(\ell)}_{\{i\},\{j\}} = \sum_{j=1}^{d}\sum_{i=1}^{d} \Pi^{(\ell)}_{\{i\},\{j\}} = \sum_{j=1}^{d} s^* \Pi^{(\ell)}_{\emptyset,\{j\}} = s^* \sum_{j=1}^{d} \Pi^{(\ell)}_{\{j\},\emptyset} = s^* \cdot s^* \Pi^{(\ell)}_{\emptyset,\emptyset} = (s^*)^2,$$

where the third and second last equalities are from the constraint $\sum_{i=1}^{d} \Pi^{(\ell)}_{\mathcal{C}_1+\{i\},\mathcal{C}_2} = s^* \Pi^{(\ell)}_{\mathcal{C}_1,\mathcal{C}_2}$, while the last is from $\Pi^{(\ell)}_{\emptyset,\emptyset} = 1$. Recall that the objective function is equivalent to

$$\frac{1}{s^*(s^*-1)} \sum_{i,j=1}^{d} A_{i,j} \Pi^{(\ell)}_{\{i\},\{j\}} = \frac{1}{s^*(s^*-1)} \sum_{i,j=1}^{d} \overline{A}_{i,j} \Pi^{(\ell)}_{\{i\},\{j\}} - \frac{1}{s^*(s^*-1)} \sum_{i=1}^{d} \Pi^{(\ell)}_{\{i\},\{i\}} = \frac{(s^*)^2 - s^*}{s^*(s^*-1)} = 1.$$

Here the last equality holds because we have

$$\sum_{i=1}^{d} \Pi^{(\ell)}_{\{i\},\{i\}} = \sum_{i=1}^{d} \Pi^{(\ell)}_{\{i\},\emptyset} = s^* \Pi^{(\ell)}_{\emptyset,\emptyset} = s^*,$$

where the first equality follows from the constraints $\Pi^{(\ell)}_{\mathcal{C}_1+\{i,i\},\mathcal{C}_2} = \Pi^{(\ell)}_{\mathcal{C}_1+\{i\},\mathcal{C}_2}$ and $\Pi^{(\ell)}_{\mathcal{C}_1,\mathcal{C}_2} = \Pi^{(\ell)}_{\mathcal{C}_1',\mathcal{C}_2'}$ for $\mathcal{C}_1 + \mathcal{C}_2 = \mathcal{C}_1' + \mathcal{C}_2'$, and the second is from $\sum_{i=1}^{d} \Pi^{(\ell)}_{\mathcal{C}_1+\{i\},\mathcal{C}_2} = s^* \Pi^{(\ell)}_{\mathcal{C}_1,\mathcal{C}_2}$. Therefore, the objective value corresponding to $\mathbf{\Pi}^{(\ell)}$ is one. Because $\widehat{\beta} \in \mathcal{H}^{(\ell)}$ is the maximum of the ℓ-th level SoS program or its relaxed versions, so far we obtain

$$\mathbb{P}\big(\widehat{\beta} \geq 1 \mid \mathbf{\Pi}^{(\ell)} \succeq 0\big) = 1. \tag{B.5}$$

According to the same proof of Theorem 3, we have $\mathbf{\Pi}^{(\ell)} \succeq 0$ holds with probability at least $1/2$ for $s^* = o\big\{\big[d/(\log d)^2\big]^{1/2\ell}\big\}$. Also, according to (B.5) and our setting that $\beta^* = 1/2$, from Markov's inequality we have

$$\begin{aligned} \mathbb{E}\big|\widehat{\beta} - \beta^*\big| &\geq 1/2 \cdot \mathbb{P}\big(\big|\widehat{\beta} - \beta^*\big| \geq 1/2\big) \\ &\geq 1/2 \cdot \mathbb{P}\big(\widehat{\beta} \geq 1 \mid \mathbf{\Pi}^{(\ell)} \succeq \mathbf{0}\big) \cdot \mathbb{P}\big(\mathbf{\Pi}^{(\ell)} \succeq \mathbf{0}\big) \geq 1/4, \end{aligned} \tag{B.6}$$

for any $\widehat{\beta} \in \mathcal{H}^{(\ell)}$ and $s^* = o\big\{\big[d/(\log d)^2\big]^{1/2\ell}\big\}$. Recall that our construction of distribution is within $\mathcal{P}(s^*, d)$. Hence we conclude the proof. □