A. Appendix

In this section, we provide the proofs of several technical results that are claimed or used in our main paper.

A.1. Proof of Proposition 4.1

The proof follows via a reduction from the so-called SUBSETSUM problem, which is known to be NP-hard (Garey & Johnson, 1979). Recall that the SUBSETSUM decision problem is as follows: given \( n \) numbers, \( a_1, \ldots, a_n \) in \( \mathbb{R} \), decide if there exists a partition \( S \subseteq \{n\} \) such that

\[
\sum_{i \in S} a_i = \sum_{j \in S^c} a_j.
\]

We show that if we can solve the mixed linear equations problem in polynomial time, then we can solve the SUBSETSUM problem, which would thus imply that \( P = NP \).

Given \( \{a_1, \ldots, a_n\} \), we must design a matrix \( X \), and output variable \( y \), such that if we could solve the mixed linear equation problem specified by \((y, X)\), then we could decide the subset sum problem on \( \{a_1, \ldots, a_n\} \). To this end, we define:

\[
X = \begin{bmatrix}
I_n & I_n & \cdots & I_n \\
0_{n \times 1} & 0_{n \times 1} & \cdots & 0_{n \times 1} \\
1 & 1 & \cdots & 1 \\
1 & 1 & \cdots & 1
\end{bmatrix}, \quad y = \begin{pmatrix}
1_n \times 1 \\
0_{n \times 1} \\
\sum_i a_i / 2
\end{pmatrix}.
\]

Here, \( I_n \) denotes the \( n \times n \) identity matrix, \( 1_{n \times 1} \) the \( n \times 1 \) vector of 1’s, and similarly, \( 0_{n \times 1} \) the \( n \times 1 \) vector of 0’s. Finding a solution to the mixed linear equations problem amounts to finding a subset \( S \subseteq \{2n + 1\} \) of the \( 2n + 1 \) constraints, and vectors \( \beta^{(1)}, \beta^{(2)} \in \mathbb{R}^n \), so that \( \beta^{(1)} \) satisfies the equalities \( X_S \beta^{(1)} = y_S \) and \( \beta^{(2)} \) the equalities \( X_S \beta_2 = y_S \). Note that \( S \) cannot contain \( i \) and \( n + i \), since these equalities are mutually exclusive. The consequence is that we have \( \beta^{(1)}_i = 1 - \beta^{(2)}_i \). Thus if the first \( 2n \) constraints are satisfied, the final constraint, therefore, can only be satisfied if we have

\[
\sum_{i \in S} a_i = \sum_i a_i \beta^{(1)}_i = \sum_j a_j \beta^{(2)}_j = \sum_{j \in S^c} a_j,
\]

thus proving the result.

A.2. Proof of Proposition 4.2

To show that our SVD initialization produces a good initial solution, requires two steps. Recall that Algorithm 5 finds the two dimensional subspace spanned by the top two eigenvectors of the matrix \( M = \frac{1}{\delta^2} \sum_{i \in S^c} y_i^2 x_i \otimes x_i \), and then searches on a discretization of the circle in that subspace for two vectors that minimize the loss function, \( \mathcal{L}_+ \) evaluated on the samples in \( S_+ \).

We first show that the top eigenspace of \( M \) is indeed close to the top eigenspace of its expectation, \( p_1 \beta^*_1 \otimes \beta^*_1 + p_2 \beta^*_2 \otimes \beta^*_2 + I \), i.e., it is close to \( \text{span}\{\beta^*_1, \beta^*_2\} \), and that some pair of elements of the discretization are close to \( (\beta^*_1, \beta^*_2) \). This is the content of lemma A.1. We then show that our loss function \( \mathcal{L}_+ \) is able to select good points from the discretization.

Our algorithm then uses the loss function \( \mathcal{L}_+ \) (evaluated on new samples in \( S_+ \)) to select good points from the grid \( G \). Lemma A.2 shows that as long as the number \( S_+ \) of these new samples is large enough, we can upper and lower bound, with high probability, the empirically evaluated loss \( \mathcal{L}_+(\beta_1, \beta_2) \) of any candidate pair \( \beta_1, \beta_2 \) by the true error err of that candidate pair. This provides the critical result allowing us to do the correct selection in the 1-d search phase.

Now we are ready to prove the result. Suppose the conditions of lemma A.1 hold. Then we are guaranteed the existence of \( (\hat{\beta}_1, \hat{\beta}_2) \) in the grid \( G \) with \( \delta \)-resolution, such that \( \max_i \| \hat{\beta}_i - \beta^*_i \| < \delta \). Next, let \( (\beta_1^{(0)}, \beta_2^{(0)}) \) be the output of our SVD initialization, and let \( \text{err} \) denote their distance from \( (\beta^*_1, \beta^*_2) \). By definition, the vectors \( (\beta_1^{(0)}, \beta_2^{(0)}) \) minimize the loss function \( \mathcal{L}_+ \) taken on inputs \( S_+ \), and hence \( \mathcal{L}_+(\beta_1^{(0)}, \beta_2^{(0)}) \leq \mathcal{L}_+(\hat{\beta}_1, \hat{\beta}_2) \). Using the lower bound from lemma A.2, applied to \( (\beta_1^{(0)}, \beta_2^{(0)}) \) we have:

\[
\frac{1}{\delta} \sqrt{\text{min}\{p_1, p_2\} \text{err}} \leq \sqrt{\frac{\mathcal{L}_+(\beta_1^{(0)}, \beta_2^{(0)})}{|S_+|}}.
\]
Alternating Minimization for Mixed Linear Regression

From the upper bound applied to \((\bar{\beta}_1, \bar{\beta}_2)\), we have
\[
\sqrt{\frac{\mathcal{L}_+ (\bar{\beta}_1, \bar{\beta}_2)}{|S_+|}} \leq 1.1\delta.
\]

Recalling that \(\mathcal{L}_+ (\beta_1^{(0)}, \beta_2^{(0)}) \leq \mathcal{L}_+ (\bar{\beta}_1, \bar{\beta}_2)\), and taking
\[
\delta \leq \frac{2}{11} \tilde{c} \|\beta_1^* - \beta_2^*\|_2 \sqrt{\min \{p_1, p_2\}^3},
\]
we combine to finally obtain:
\[
\text{err} \leq \frac{11}{2} \delta \sqrt{\min \{p_1, p_2\}} \leq \tilde{c} \min \{p_1, p_2\} \|\beta_1^* - \beta_2^*\|_2.
\]

where \(\tilde{c}\) is as in the statement of proposition 4.2.

A.3. Proof of Proposition 4.3

Using standard concentration results, in lemma A.1, we have shown if
\[
|S_*| > c (1/\bar{\delta})^2 k \log^2 k,
\]
with probability at least \(1 - \frac{1}{k^2}\),
\[
\|M - \mathbb{E}(M)\| < 3\bar{\delta}
\]
Hence, we have
\[
||\lambda_1^* - \lambda_2^*| - |\lambda_1 - \lambda_2|| \leq 6\bar{\delta}.
\]
The approximate error of \(\Delta_b^*\) can be bounded as:
\[
2p_b|\Delta_b^* - \Delta_b| \leq 6\bar{\delta} + (p_b^2 - p_b^2) \left[ \frac{1}{\lambda_b^* - \lambda_b^*} - \frac{1}{\lambda_b - \lambda_b} \right]
\leq 6\bar{\delta} + |p_b^2 - p_b^2| \left[ \frac{6\bar{\delta}}{\lambda_b^* - \lambda_b^* (\lambda_b - \lambda_b)} \right]
\leq 6\bar{\delta} + |p_b^2 - p_b^2| \left[ \frac{6\bar{\delta}}{\lambda_b^* - \lambda_b^* (|\lambda_b^* - \lambda_b| - 6\bar{\delta})} \right]
\leq 6\bar{\delta} + |p_b^2 - p_b^2| \left[ \frac{12\delta}{|\lambda_b^* - \lambda_b^*|^2} \right]
\]
In the last inequality we use \(\bar{\delta} \leq \frac{|\lambda_1^* - \lambda_2^*|}{12}\).

Next, we calculate approximation error of eigenvectors. Note that \(\mathbb{E}\left(\frac{M - I}{2}\right) = p_1\beta_1^* \otimes \beta_1^* + p_2\beta_2^* \otimes \beta_2^*\), we have
\[
\{\lambda_1^*, \lambda_2^*\} = \left\{ \frac{1 + \kappa}{2}, \frac{1 - \kappa}{2} \right\}.
\]
Using lemma A.3, we have,
\[
\|v_b - v_b^*\|^2 \leq \frac{6\bar{\delta}^2 + 24\bar{\delta}}{\kappa (1 - \kappa)} \leq \frac{24\bar{\delta}}{\kappa (1 - \kappa)}, \ b = 1, 2.
\]
Then
\[
\|\beta_b^* - \beta_b\|_2 \leq \sqrt{\frac{1 - \Delta_b}{2}} v_b^* - \sqrt{\frac{1 - \Delta_b}{2}} v_b + \sqrt{\frac{1 + \Delta_b}{2}} v_{-b}^* - \sqrt{\frac{1 + \Delta_b}{2}} v_{-b}.
\]
Note that
\[
\left| \sqrt{\frac{1 - \Delta^*_b}{2}} \mathbf{v}_b^* - \sqrt{\frac{1 - \Delta_b}{2}} \mathbf{v}_b \right| = \sqrt{\frac{1 - \Delta^*_b}{2}} \mathbf{v}_b^* - \sqrt{\frac{1 - \Delta_b}{2}} \mathbf{v}_b + \sqrt{\frac{1 - \Delta^*_b}{2}} \mathbf{v}_b - \sqrt{\frac{1 - \Delta_b}{2}} \mathbf{v}_b \\
\leq \sqrt{\frac{1 - \Delta^*_b}{2}} \| \mathbf{v}_b - \mathbf{v}_b^* \|_2 + \sqrt{\frac{1 - \Delta_b}{2}} \| \mathbf{v}_b \|_2 \\
\leq \| \mathbf{v}_b - \mathbf{v}_b^* \|_2 + \sqrt{\frac{1 - \Delta_b}{2}} \| \mathbf{v}_b \|_2 \\
\leq \| \mathbf{v}_b - \mathbf{v}_b^* \|_2 + \frac{1}{2} \| \Delta_b - \Delta^*_b \|. 
\]

Plug the above result back to (14), we obtain
\[
\| \beta_b^* - \beta_b \|_2 \lesssim \sqrt{|\Delta_b - \Delta^*_b| + \sum_b \| \mathbf{v}_b - \mathbf{v}_b^* \|_2} \\
\leq \sqrt{\frac{\bar{\delta}}{\kappa(1 - \kappa)} + \frac{1}{\sqrt{\min\{p_1, p_2\}}}} \sqrt{\frac{\bar{\delta}}{\kappa^2}} \\
\leq \sqrt{\frac{\bar{\delta}}{\min\{p_1, p_2\}}} \times \sqrt{\frac{1}{\kappa(1 - \kappa)} + \frac{1}{\kappa^2}} \\
= \sqrt{\frac{\bar{\delta}}{\min\{p_1, p_2\}} \frac{1}{\kappa(1 - \kappa)}}. 
\]

By setting the above upper bound to be less than \( \hat{c} \min\{p_1, p_2\} \| \beta_1^* - \beta_2^* \|_2 \), we complete the proof.

**A.4. Proof of Proposition 4.5**

It’s equivalent to show that \( J_b = J_b^*_b, b = 1, 2 \). Let’s consider \( b = 1 \), that is for all \( p_1 \times |S_i| \) samples that are generated by \( y = \mathbf{x}^T \beta_1^* \). For simplicity, let \( \beta_1, \beta_2 \) denote \( \beta_1^{(t-1)}, \beta_2^{(t-1)} \), we need
\[
(x^T (\beta_1^* - \beta_1))^2 < (x^T (\beta_1^* - \beta_2))^2.
\]

From lemma 5.1,
\[
P \left[ (x^T (\beta_1^* - \beta_1))^2 < (x^T (\beta_1^* - \beta_2))^2 \right] \geq 1 - \frac{\| \beta_1^* - \beta_1 \|_2}{\| \beta_1^* - \beta_2 \|_2} \\
\geq 1 - 2 \frac{\| \beta_1^* - \beta_1 \|_2}{\| \beta_1^* - \beta_2 \|_2} \\
\geq 1 - \frac{2c_1}{k^2}. 
\]

Then we use union bound for \( p_1 \times |S_i| \) samples in \( J_1^* \),
\[
P \left[ (x_i^T (\beta_1^* - \beta_1))^2 < (x_i^T (\beta_1^* - \beta_2))^2, \text{for all } i \in J_1^* \right] \geq 1 - p_1 c_2 k \times \frac{2c_1}{k^2} \geq 1 - \frac{c'}{k}.
\]

So all samples are correctly clustered with high probability.

As \( \frac{1}{\min\{p_1, p_2\}} k < |S_i| \), number of samples in \( J_1 \) and \( J_2 \) are both greater than \( k \). Therefore, least square solution reveals the ground truth. In other words, \( err^{(t)} = 0 \).
A.5. Proof of Lemma 5.1

(1) Without loss of generality, we assume $T\{u, v\} = T\{e_1, e_2\}$. Let $x_1, x_2$ denote $x^T e_1, x^T e_2$. As $x_1, x_2$ are independent Gaussian random variables, we have $x_1 = A \cos \theta, x_2 = A \sin \theta$, where $A$ is Rayleigh random variable, and $\theta$ is uniformly distributed over $[0, 2\pi)$. Conditioning on $(x^T u)^2 > (x^T v)^2$, the range of $\theta$ is truncated to be $[\theta_0, \theta_0 + \alpha(u, v)] \cup [\theta_0 + \pi, \theta_0 + \pi + \alpha(u, v)]$ for some $\theta_0$. It is not hard to see the eigenvalues of covariance matrix of $(x_1, x_2)$ are $1 + \frac{\sin\alpha(u, v)}{\sigma(u, v)}, 1 - \frac{\sin\alpha(u, v)}{\sigma(u, v)}$.

As the rest if the eigenvalues of $\Sigma$ are $1$, this completes the proof.

(2) Note that

$$\mathbb{P} \left[ (x^T u)^2 > (x^T v)^2 \right] = \frac{\alpha(u, v)}{\pi}.$$ 

If $\|u\|_2 > \|v\|_2, \alpha(u, v) > \frac{\pi}{2}$, when $\|u\|_2 < \|v\|_2$,

$$\cos \alpha(u, v) \geq \frac{\|u\|_2^2 - \|u\|_2^2}{\|u\|_2^2 + \|v\|_2^2}.$$ 

Note that for any $\alpha \in [0, \pi/2)$, $\alpha \leq \frac{\pi}{2} \sin \alpha$. We have

$$\mathbb{P} \left[ (x^T u)^2 > (x^T v)^2 \right] \leq \frac{1}{2} \sin \alpha(u, v) \leq \frac{\|u\|_2^2}{\|u\|_2^2 + \|v\|_2^2} \leq \frac{\|u\|_2^2}{\|v\|_2^2}.$$ 

A.6. Supporting Lemmas

**Lemma A.1.** For any given $\delta > 0$, let $G$ denote the grid points, at resolution $\delta$, of the unit circle on the subspace spanned by the top two eigenvectors of $M$, formed with $|S_u|$ samples. Then, there exists an absolute constant $c$ such that if

$$|S_u| \geq c(1/\delta)^2 k \log^2 k,$$

where

$$\hat{\delta} = \frac{\delta^2}{684} (1 - \sqrt{1 - 4(1 - \langle \beta_{0}, \beta_{0}^* \rangle) p_1 p_2}),$$

then

$$\min_{a \in G} \|\beta_i^* - a\| \leq \delta, i = 1, 2,$$

with probability at least $1 - O \left( \frac{1}{\delta^2} \right)$.

**Proof.** In order to prove the result, we make use of standard concentration results.

Let $\Sigma = \mathbb{E}[M]$. We observe that $\mathbb{P} \left[ |y| > \sqrt{2\alpha \log k} \right] \leq n^{-\alpha}, \mathbb{P} \left[ |x|^2 \geq 3k \right] \leq e^{-k/3}$. Suppose $N$ is much less than $O \left( k^{10} \right)$, where the constant is arbitrarily chosen here. Set $\alpha = 12$. Then with probability at least $1 - O \left( \frac{1}{\delta^2} \right)$. The vectors $y, x$, are all supported in a ball with radius $\sqrt{2k \log k}$. Directly following theorem 5.44 in (Vershynin, 2010), we claim that when $N > C(1/\delta)^2 k \log^2 k$,

$$\|M - \Sigma\| \leq \hat{\delta} \|\Sigma\| \leq 3\hat{\delta}.$$ 

We use $\sigma_i(A)$ to denote the $i$'th biggest eigenvalue of the positive semidefinite matrix $A$. By simple algebraic calculation we get $\sigma_1(\Sigma) = 2 + \kappa, \sigma_2(\Sigma) = 2 - \kappa$, where $\kappa = \sqrt{1 - 4(1 - \langle \beta_{0}, \beta_{0}^* \rangle) p_1 p_2}$. The top two eigenvectors of $\Sigma$ are denoted as $v_1^*, v_2^*$. We use $v_1, v_2$ to denote the top two eigenvectors of $M$. Lemma A.3 yields that

$$\|v_i^* - \mathcal{P}_{T(v_1, v_2)} v_i^*\|_2^2 \leq \frac{12\hat{\delta}}{\sigma_2(M) - \sigma_3(M)} \leq \frac{12\hat{\delta}}{\sigma_2(\Sigma) - \sigma_3(\Sigma) - 6\delta} \leq \frac{12\hat{\delta}}{1 - \kappa - 6\delta} = \frac{24\hat{\delta}}{1 - \kappa}, i = 1, 2.$$
The last inequality holds when \( \hat{\delta} \leq \frac{1 - \kappa}{2} \). Using the fact that for any two vectors \( \mathbf{a}, \mathbf{b}, \| \mathbf{a} + \mathbf{b} \|_2^2 \leq 2 \| \mathbf{a} \|_2^2 + 2 \| \mathbf{b} \|_2^2 \), we conclude that

\[
\| \beta_i^* - \mathcal{P}_{T(v, S)} \beta_i^* \|_2^2 \leq \frac{48 \hat{\delta}}{1 - \kappa}, i = 1, 2.
\]

Let \( w = \| \beta_i^* - \mathcal{P}_{T(u, v)} \beta_i^* \|_2 \). Then, by simple geometric relation,

\[
\min_{\mathbf{a} \in S^{k-1}} \| \mathbf{a} - \beta_i^* \|_2 \leq 2 - 2 \sqrt{1 - w^2} \\
\leq 2w^2 \\
\leq (\frac{\varepsilon}{2})^2, i = 1, 2.
\]

Consider the \( \delta \)-resolution grid \( G \). We observe that for any point in \( S^{k-1} \cap T(u, v) \), there exists a point in \( G \) that is within \( \delta/2 \) away from it. By triangle inequality, we end up with

\[
\min_{\mathbf{a} \in W} \| \mathbf{a} - \beta_i^* \|_2 \leq \delta.
\]  \( \Box \)

**Lemma A.2.** Let \( \hat{\beta}_1, \hat{\beta}_2 \) be any two given vectors with error defined by \( \text{err} := \max_{i=1,2} \| \hat{\beta}_i - \beta_i^* \| \). There exist constants \( c_1, c_2 > 0 \) such that as long as we have enough testing samples,

\[
|S_+| \geq c_1 k / \min \{ p_1, p_2 \},
\]

then with probability at least \( 1 - O \left( e^{-c_1 k} \right) \)

\[
\sqrt{\frac{\mathcal{L}_+(\hat{\beta}_1, \hat{\beta}_2)}{|S_+|}} \leq 1.1 \text{err}
\]

and

\[
\sqrt{\frac{\mathcal{L}_+(\hat{\beta}_1, \hat{\beta}_2)}{|S_+|}} \geq \frac{1}{5} \sqrt{\min \{ p_1, p_2 \} \min \left\{ \text{err}, \frac{1}{2} \| \beta_1^* - \beta_2^* \|_2 \right\}}.
\]

**Proof.** Our notation here, namely, \( J_1, J_2, J_1^*, J_2^* \), is consistent with proof of Theorem 4.4. Note that we have:

\[
\mathcal{L}(\beta_1, \beta_2) = \sum_i \min z_i(y_i - \mathbf{x}_i^T \beta_1)^2 + (1 - z_i)(y_i - \mathbf{x}_i^T \beta_2)^2.
\]

For the upper bound, we assign label \( z_i \) as the true label. Then,

\[
\mathcal{L} \leq \sum_{i \in J_1^*} (\mathbf{x}_i^T (\beta_1^* - \beta_1))^2 + \sum_{i \in J_2^*} (\mathbf{x}_i^T (\beta_2^* - \beta_2))^2.
\]

When \( |S_+| \geq c_{\min} \frac{k}{\min \{ p_1, p_2 \}} \), then the number of samples in set \( J_1^*, J_2^* \) is also greater than \( Ck \). Following standard concentration results, there exist constants \( C, c_1 \), such that with probability greater than \( 1 - e^{-c_1 k} \), we have

\[
\| \frac{1}{p_j |S_+|} \sum_{i \in J_j^*} \mathbf{x}_i \mathbf{x}_i^T - \mathbf{I} \| \leq 0.21, j = 1, 2.
\]

We have

\[
\mathcal{L} \leq 1.21 p_1 |S_+| \| \beta_1 - \beta_1^* \|_2^2 + 1.21 p_2 |S_+| \| \beta_2 - \beta_2^* \|_2^2 \\
\leq 1.21 |S_+| \text{err}^2.
\]
Alternating Minimization for Mixed Linear Regression

For the lower bound, we observe that

$$\mathcal{L} = \sum_{i \in J \cap J_1^*} (x_i^T(\beta_1^* - \beta_1^*))^2 + \sum_{i \in J \cap J_2^*} (x_i^T(\beta_2^* - \beta_1^*))^2 + \sum_{i \in J_1 \cap J_2^*} (x_i^T(\beta_1^* - \beta_2^*))^2 + \sum_{i \in J_2 \cap J_2^*} (x_i^T(\beta_2^* - \beta_2^*))^2.$$  

First we consider the first term, $A_1$. Note a simple fact that $\|\beta_1 - \beta_1^*\|_2 < \|\beta_2 - \beta_1^*\|_2$ or $\|\beta_1 - \beta_1^*\|_2 > \|\beta_2 - \beta_1^*\|_2$. In the first case, from Lemma 5.1, $\mathbb{E}[|J_1 \cap J_2^*|] \geq \frac{1}{4} p_1 |\mathcal{S}_+|$. From Hoeffding’s inequality and concentration result (see proof of Lemma 5.1 for similar techniques), for any $\delta \in (0, 1 - \frac{2}{3})$, there exist constants $C', c'_1$, such that when $N \geq C' k / p_1$, with probability at least $1 - e^{-c'_1 k}$,

$$\sum_{i \in J \cap J_1^*} (x_i^T(\beta_1^* - \beta_1^*))^2 \geq \frac{1}{4} p_1 |\mathcal{S}_+| (1 - \frac{1}{4} \|\beta_1 - \beta_1^*\|_2^2).$$

In the second case, we have a similar result:

$$\sum_{i \in J \cap J_2^*} (x_i^T(\beta_2^* - \beta_1^*))^2 \geq \frac{1}{4} p_1 |\mathcal{S}_+| (1 - \frac{1}{4} \|\beta_2 - \beta_1^*\|_2^2).$$

Let $1 - \frac{2}{3} - \delta = 0.3$ and choose $C', c'_1$ to let the above results also hold for $A_2$. We then conclude that when $N > C' \frac{k}{\min\{p_1, p_2\}}$,

$$\mathcal{L} \geq \frac{0.3}{4} p_1 |\mathcal{S}_+| \min\{\|\beta_1 - \beta_1^*\|_2^2, \|\beta_2 - \beta_2^*\|_2^2\} + \frac{0.3}{4} p_2 |\mathcal{S}_+| \min\{\|\beta_1 - \beta_2^*\|_2^2, \|\beta_2 - \beta_2^*\|_2^2\}. \quad (19)$$

When $\|\beta_1 - \beta_1^*\|_2 < \|\beta_2 - \beta_2^*\|_2$ and $\|\beta_2 - \beta_2^*\|_2 < \|\beta_1 - \beta_2^*\|_2$, (19) implies

$$\mathcal{L} \geq \frac{1}{25} \min\{p_1, p_2\} |\mathcal{S}_+| |\text{err}|^2. \quad (20)$$

When $\|\beta_1 - \beta_1^*\|_2 > \|\beta_2 - \beta_2^*\|_2$ and $\|\beta_2 - \beta_2^*\|_2 < \|\beta_1 - \beta_2^*\|_2$, we have

$$\mathcal{L} \geq \frac{1}{25} \min\{p_1, p_2\} |\mathcal{S}_+| (||\beta_2 - \beta_2^*||_2^2 + ||\beta_2 - \beta_2^*||_2^2) \quad (21)$$

$$\geq \frac{1}{25} \min\{p_1, p_2\} |\mathcal{S}_+| \frac{1}{4} ||\beta_1^* - \beta_2^*||_2^2. \quad (22)$$

Note that it is impossible for $\|\beta_1 - \beta_1^*\|_2 > \|\beta_2 - \beta_1^*\|_2$ and $\|\beta_2 - \beta_2^*\|_2 > \|\beta_1 - \beta_2^*\|_2$ both to be true. Otherwise, we could switch the subscripts of the two $\beta$’s. Putting (20) and (22) together, we complete the proof.

**Lemma A.3.** Suppose symmetric matrix $\Sigma \in \mathbb{R}^{n \times n}$ has eigenvalues $\lambda_1 \geq \lambda_2 > \lambda_3...$ with corresponding normalized eigenvectors denoted as $u_1, u_2, u_3, ...$. Let $M$ be another symmetric matrix with eigenvalues: $\tilde{\lambda}_1 \geq \lambda_2 > \tilde{\lambda}_3...$ and eigenvectors $\tilde{u}_1, \tilde{u}_2, \tilde{u}_3, ...$. (a) Let $\text{span}\{u_1, u_2\}$ denote the hyperplane spanned by $u_1, u_2$. If $||M - \Sigma||_2 \leq \epsilon$, for $\epsilon < \frac{\lambda_2 - \lambda_3}{2}$ we have

$$||\tilde{u}_i - \Pi_{(u_1, u_2)} \tilde{u}_i||_2^2 \leq \frac{4\epsilon}{\lambda_2 - \lambda_3}, i = 1, 2. \quad (23)$$

Moreover, if $\lambda_1 \neq \lambda_2$,

$$||u_1 - \tilde{u}_1||_2^2 \leq \frac{4\epsilon}{\lambda_1 - \lambda_2} \quad (24)$$

$$||u_2 - \tilde{u}_2||_2^2 \leq \frac{4\epsilon}{\lambda_1 - \lambda_2} + \frac{8\epsilon}{\lambda_2 - \lambda_3} \quad (25)$$
Combining (26) and (27), using \( \tilde{\alpha} \) another upper bound \( \lambda \) we assume
We have
The last inequality follows from (31).
Suppose
We obtain the last inequality from (31) and (34).
Using some intermediate results, we derive the bounds for eigenvectors in the case \( \lambda \).
Through (30) and (34), we complete the proof of (23).
\[ \| u_1 - \tilde{u}_1 \|^2 = (1 - \alpha_1)^2 + \beta_1^2 + \gamma_1^2 \]
\[ = (1 - \alpha_1)^2 + 1 - \alpha_1^2 \]
\[ \leq 2(1 - \alpha_1^2) \]
\[ \leq \frac{4\varepsilon}{\lambda_1 - \lambda_2}. \]
The last inequality follows from (31).
Similarly,
\[ \| u_2 - \tilde{u}_2 \|^2 \leq 2(1 - \beta_2^2) \]
\[ = 2(\alpha_2^2 + \gamma_2^2) \]
\[ \leq 2(1 - \alpha_2^2 + \gamma_2^2) \]
\[ \leq \frac{4\varepsilon}{\lambda_1 - \lambda_2} + \frac{8\varepsilon}{\lambda_2 - \lambda_3}. \]
We obtain the last inequality from (31) and (34).