Electronic Journal of Statistics
ISSN: 1935-7524

Functional wavelet regression for linear function-on-function models

Ruiyan Luo

Division of Epidemiology and Biostatistics, Georgia State University School of Public Health, One Park Place, Atlanta, GA 30303
e-mail: rluo@gsu.edu

Xin Qi

Department of Mathematics and Statistics, Georgia State University, 30 Pryor Street, Atlanta, GA 30303
e-mail: xqi3@gsu.edu

Yanhong Wang

Department of Mathematics and Statistics, Georgia State University, 30 Pryor Street, Atlanta, GA 30303
e-mail: wangyanhongws@gmail.com

Contents

This supplementary material is organized as follows. Additional proofs of theorems can be found in Appendix B. The proofs of all technical lemmas are provided in Appendix C. Additional figures are provided in Appendix D.

1. Appendix B: Additional proofs of theorems

1.1. Proof of Theorem 4.2

- *Proof of Theorem 4.2 (a).*

The proof of Part (a) is broken into several steps.

- **Step 1: Provide two inequalities which play an important role.**

Due to the constraints of (3.7) and (3.10), we have

$$\boldsymbol{\alpha}_k^{\mathrm{T}}\mathbf{S}\boldsymbol{\alpha}_k = 1, \quad \widehat{\boldsymbol{\alpha}}_k^{\mathrm{T}}\mathbf{S}\widehat{\boldsymbol{\alpha}}_k = 1, \quad 1 \le k \le K.$$

Since $\widehat{\boldsymbol{\alpha}}_1$ is solution to

$$\max_{\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{S}\boldsymbol{\alpha}=1} \frac{\boldsymbol{\alpha}^{\mathrm{T}}\widehat{\mathbf{B}}\boldsymbol{\alpha}}{\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{S}\boldsymbol{\alpha} + \tau^{(1)}\|\boldsymbol{\alpha}\|^2_{\lambda^{(1)}}} = \max_{\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{S}\boldsymbol{\alpha}=1} \frac{\boldsymbol{\alpha}^{\mathrm{T}}\widehat{\mathbf{B}}\boldsymbol{\alpha}}{1 + \tau^{(1)}\|\boldsymbol{\alpha}\|^2_{\lambda^{(1)}}},$$

we have

$$\frac{\boldsymbol{\alpha}_1^{\mathrm{T}}\widehat{\mathbf{B}}\boldsymbol{\alpha}_1}{1 + \tau^{(1)}\|\boldsymbol{\alpha}_1\|^2_{\lambda^{(1)}}} \le \frac{\widehat{\boldsymbol{\alpha}}_1^{\mathrm{T}}\widehat{\mathbf{B}}\widehat{\boldsymbol{\alpha}}_1}{1 + \tau^{(1)}\|\widehat{\boldsymbol{\alpha}}_1\|^2_{\lambda^{(1)}}}. \tag{1.1}$$

On the other hand, because $\boldsymbol{\alpha}_1$ is the first eigenvector of the generalized eigenvalue problem (3.7), by Theorem 3.1, we have

$$\mu_1(\boldsymbol{\Xi}) = \boldsymbol{\alpha}_1^{\mathrm{T}}\mathbf{B}\boldsymbol{\alpha}_1 = \frac{\boldsymbol{\alpha}_1^{\mathrm{T}}\mathbf{B}\boldsymbol{\alpha}_1}{\boldsymbol{\alpha}_1^{\mathrm{T}}\mathbf{S}\boldsymbol{\alpha}_1} \ge \frac{\widehat{\boldsymbol{\alpha}}_1^{\mathrm{T}}\mathbf{B}\widehat{\boldsymbol{\alpha}}_1}{\widehat{\boldsymbol{\alpha}}_1^{\mathrm{T}}\mathbf{S}\widehat{\boldsymbol{\alpha}}_1} = \widehat{\boldsymbol{\alpha}}_1^{\mathrm{T}}\mathbf{B}\widehat{\boldsymbol{\alpha}}_1. \tag{1.2}$$

- **Step 2: Define an event Ω which has probability converging to one as $n, p \to \infty$. Then we will focus on the elements in this event.**

Then by the definition (3.9) of $\widehat{\mathbf{B}}$ and the definition of $\mathbf{B}$ in (3.6),

$$\begin{aligned}
\widehat{\mathbf{B}} &= \frac{1}{n^2}\mathbf{X}^{\mathrm{T}}\left\{\int_0^1 [\mathbf{Y}(t) - \bar{y}(t)\mathbf{1}_n][\mathbf{Y}(t) - \bar{y}(t)\mathbf{1}_n]^{\mathrm{T}}\,dt\right\}\mathbf{X} \qquad (1.3)\\
&= \frac{1}{n^2}\mathbf{X}^{\mathrm{T}}\left\{\int_0^1 (\mathbf{X}\boldsymbol{\beta}(t) + \boldsymbol{\varepsilon}(t) - \bar{\varepsilon}(t)\mathbf{1}_n)(\mathbf{X}\boldsymbol{\beta}(t) + \boldsymbol{\varepsilon}(t) - \bar{\varepsilon}(t)\mathbf{1}_n)^{\mathrm{T}}dt\right\}\mathbf{X}\\
&= \mathbf{Z}^{\mathrm{T}}\left\{\int_0^1 (\mathbf{Z}\boldsymbol{\beta}(t) + \boldsymbol{\varrho}(t))(\mathbf{Z}\boldsymbol{\beta}(t) + \boldsymbol{\varrho}(t))^{\mathrm{T}}dt\right\}\mathbf{Z}\\
&= \mathbf{B} + \int_0^1 (\mathbf{S}\boldsymbol{\beta}(t))(\mathbf{Z}^{\mathrm{T}}\boldsymbol{\varrho}(t))^{\mathrm{T}}dt + \int_0^1 (\mathbf{Z}^{\mathrm{T}}\boldsymbol{\varrho}(t))(\mathbf{S}\boldsymbol{\beta}(t))^{\mathrm{T}}dt + \int_0^1 (\mathbf{Z}^{\mathrm{T}}\boldsymbol{\varrho}(t))(\mathbf{Z}^{\mathrm{T}}\boldsymbol{\varrho}(t))^{\mathrm{T}}dt,
\end{aligned}$$

where $\boldsymbol{\varrho}(t)$ and $\boldsymbol{\varrho}(t)$ are defined in (B.1).

Lemma 1. *Suppose that Condition 2 holds and $p \geq 2$. Then for any $C > M_\epsilon/\sqrt{\log p}$, we have*

$$P\left(\max_{1\leq j\leq p}\|(\mathbf{Z}^{\mathrm{T}}\boldsymbol{\varrho}(t))_j\|_{L^2} > C\sqrt{\frac{\log p}{n}}\right) \leq p e^{M_\epsilon^2/2\sigma^2} p^{-C^2/4\sigma^2},$$

$$\text{and} \qquad P\left(\bigcup_{k=1}^{K}\left\{\|(\mathbf{Z}\boldsymbol{\alpha}_k)^{\mathrm{T}}\boldsymbol{\varrho}(t)\|_{L^2} > C\sqrt{\frac{\log p}{n}}\right\}\right) \leq K e^{M_\epsilon^2/2\sigma^2} p^{-C^2/4\sigma^2},$$

where $(\mathbf{Z}^{\mathrm{T}}\boldsymbol{\varrho}(t))_j$ denotes the j-th function in the vector $\mathbf{Z}^{\mathrm{T}}\boldsymbol{\varrho}(t)$.

Note that $\varpi = \sqrt{\log p/n}$. Define the event

$$\Omega = \left\{\max_{1\leq j\leq p}\|(\mathbf{Z}^{\mathrm{T}}\boldsymbol{\varrho}(t))_j\|_{L^2} \leq C_0\varpi\right\} \bigcap_{k=1}^{K}\left\{\|(\mathbf{Z}\boldsymbol{\alpha}_k)^{\mathrm{T}}\boldsymbol{\varrho}(t)\|_{L^2} \leq C_0\varpi\right\}, \tag{1.4}$$

where $C_0 = 2\max\{M_\epsilon/\sqrt{\log 2}, 2\sigma\}$. Hence, by Lemma 1, we have

$$P(\Omega) \geq 1 - (p+K)e^{M_\epsilon^2/2\sigma^2} p^{-C_0^2/4\sigma^2} \geq 1 - 2p e^{M_\epsilon^2/2\sigma^2} p^{-C_0^2/4\sigma^2},$$

where we use the fact that $K \leq p$.

- **Step 3: Provide upper bounds for $\|\widehat{\boldsymbol{\alpha}}_1\|_1$ and $\widehat{\boldsymbol{\alpha}}_1^{\mathrm{T}}\widehat{\mathbf{B}}\widehat{\boldsymbol{\alpha}}_1$.**

We first give a technical lemma, Lemma 2, and then provide the upper bounds in Lemma 3.

Lemma 2. *Under Condition 1, for any $1 \leq k \leq K$, we have $\|\boldsymbol{\alpha}_k\|_2 \leq 1/\kappa$. Under Condition 3, $\|\boldsymbol{\alpha}_k\|_1 \geq 1$ and $\|\widehat{\boldsymbol{\alpha}}_k\|_1 \geq 1$ for any $1 \leq k \leq K$. Moreover, if both the two conditions are satisfied, we have $\|\boldsymbol{\alpha}_k\|_1^2 \leq s/\kappa^2$.*

By the condition $\mu_1(\boldsymbol{\Xi}) \geq \hbar^2 C_0^2\varpi^2 s/\kappa^2$ and Lemma 2, we have

$$C_0\varpi\|\boldsymbol{\alpha}_1\|_1 \leq C_0\varpi\sqrt{s}/\kappa \leq \hbar^{-1}\mu_1(\boldsymbol{\Xi})^{1/2}, \quad \text{and hence,}$$

$$\tau^{(1)}\|\boldsymbol{\alpha}_1\|_{\lambda^{(1)}}^2 \leq \tau^{(1)}\|\boldsymbol{\alpha}_1\|_1^2 = \frac{A^{(1)}C_0\varpi\|\boldsymbol{\alpha}_1\|_1^2}{\|\boldsymbol{\alpha}_1\|_1\sqrt{\mu_1(\boldsymbol{\Xi})}} \leq A^{(1)}\hbar^{-1}\,. \tag{1.5}$$

Lemma 3. *In the event Ω, there exist $(A_1^L)'$ and $(\hbar_0)'$ only depending on c such that for any $A^{(1)} \geq (A_1^L)'$ and $\hbar \geq (\hbar_0)'$, we have*

$$\|\widehat{\boldsymbol{\alpha}}_1\|_1 \leq D_1\|\boldsymbol{\alpha}_1\|_1, \quad \widehat{\boldsymbol{\alpha}}_1^{\mathrm{T}}\widehat{\mathbf{B}}\widehat{\boldsymbol{\alpha}}_1 \leq (1+\hbar^{-1}D_1)^2\mu_1(\boldsymbol{\Xi}),$$
$$\boldsymbol{\alpha}_1^{\mathrm{T}}\widehat{\mathbf{B}}\boldsymbol{\alpha}_1 - \widehat{\boldsymbol{\alpha}}_1^{\mathrm{T}}\widehat{\mathbf{B}}\widehat{\boldsymbol{\alpha}}_1 \geq \frac{1}{2}c_2\mu_1(\boldsymbol{\Xi})\|\widehat{\boldsymbol{\gamma}}_1 - \boldsymbol{\gamma}_1\|_2^2 - \{4+\hbar^{-1}(1+D_1)\}\sqrt{\mu_1(\boldsymbol{\Xi})}C_0\varpi\|\boldsymbol{\alpha}_1 - \widehat{\boldsymbol{\alpha}}_1\|_1,$$

where $D_1 = \sqrt{6c}$.

- **Step 4: Derive the oracle inequalities in Part (a).**

Now by (1.1), we have $\boldsymbol{\alpha}_1^{\mathrm{T}}\widehat{\mathbf{B}}\boldsymbol{\alpha}_1(1+\tau^{(1)}\|\widehat{\boldsymbol{\alpha}}_1\|_{\lambda^{(1)}}^2) \leq \widehat{\boldsymbol{\alpha}}_1^{\mathrm{T}}\widehat{\mathbf{B}}\widehat{\boldsymbol{\alpha}}_1(1+\tau^{(1)}\|\boldsymbol{\alpha}_1\|_{\lambda^{(1)}}^2)$ which leads to,

$$\begin{aligned}
&\widehat{\boldsymbol{\alpha}}_1^{\mathrm{T}}\widehat{\mathbf{B}}\widehat{\boldsymbol{\alpha}}_1(\tau^{(1)}\|\boldsymbol{\alpha}_1\|_{\lambda^{(1)}}^2 - \tau^{(1)}\|\widehat{\boldsymbol{\alpha}}_1\|_{\lambda^{(1)}}^2) \\
\geq&(\boldsymbol{\alpha}_1^{\mathrm{T}}\widehat{\mathbf{B}}\boldsymbol{\alpha}_1 - \widehat{\boldsymbol{\alpha}}_1^{\mathrm{T}}\widehat{\mathbf{B}}\widehat{\boldsymbol{\alpha}}_1)(1+\tau^{(1)}\|\widehat{\boldsymbol{\alpha}}_1\|_{\lambda^{(1)}}^2) \\
\geq&\left(\frac{1}{2}c_2\mu_1(\boldsymbol{\Xi})\|\widehat{\boldsymbol{\gamma}}_1 - \boldsymbol{\gamma}_1\|_2^2 - \{4+\hbar^{-1}(1+D_1)\}C_0\sqrt{\mu_1(\boldsymbol{\Xi})}\varpi\|\boldsymbol{\alpha}_1 - \widehat{\boldsymbol{\alpha}}_1\|_1\right) \\
&\times(1+\tau^{(1)}\|\widehat{\boldsymbol{\alpha}}\|_{\lambda^{(1)}}^2) \geq \frac{1}{2}c_2\mu_1(\boldsymbol{\Xi})\|\widehat{\boldsymbol{\gamma}}_1 - \boldsymbol{\gamma}_1\|_2^2 \\
&-\{4+\hbar^{-1}(1+D_1)\}C_0\sqrt{\mu_1(\boldsymbol{\Xi})}\varpi\|\boldsymbol{\alpha}_1 - \widehat{\boldsymbol{\alpha}}_1\|_1(1+D_1^2A^{(1)}\hbar^{-1}),
\end{aligned} \tag{1.6}$$

where the third inequality follows from the last inequality in Lemma 3 and the last one is due to the first inequality in Lemma 3 and (1.5). We provide upper bounds for $\widehat{\boldsymbol{\alpha}}_1^{\mathrm{T}}\widehat{\mathbf{B}}\widehat{\boldsymbol{\alpha}}_1(\tau^{(1)}\|\boldsymbol{\alpha}_1\|_{\lambda^{(1)}}^2 - \tau^{(1)}\|\widehat{\boldsymbol{\alpha}}_1\|_{\lambda^{(1)}}^2)$ in the following lemma.

Lemma 4. *In the event Ω, we have two alternative cases:*

- **Case (a)**: $\|(\boldsymbol{\alpha}_1 - \widehat{\boldsymbol{\alpha}}_1)_{J_1}\|_1 < \lambda^{(1)}\|(\widehat{\boldsymbol{\alpha}}_1)_{J_1^c}\|_1$. *In this case, we have*

$$\begin{aligned}
&\widehat{\boldsymbol{\alpha}}_1^{\mathrm{T}}\widehat{\mathbf{B}}\widehat{\boldsymbol{\alpha}}_1(\tau^{(1)}\|\boldsymbol{\alpha}_1\|_{\lambda^{(1)}}^2 - \tau^{(1)}\|\widehat{\boldsymbol{\alpha}}_1\|_{\lambda^{(1)}}^2) \\
\leq&(1-2\hbar^{-1})A^{(1)}C_0\mu_1(\boldsymbol{\Xi})^{1/2}\varpi(\|(\boldsymbol{\alpha}_1 - \widehat{\boldsymbol{\alpha}}_1)_{J_1}\|_1 - \lambda^{(1)}\|(\widehat{\boldsymbol{\alpha}}_1)_{J_1^c}\|_1).
\end{aligned}$$

- **Case (b)**: $\|(\boldsymbol{\alpha}_1 - \widehat{\boldsymbol{\alpha}}_1)_{J_1}\|_1 \geq \lambda^{(1)}\|(\widehat{\boldsymbol{\alpha}}_1)_{J_1^c}\|_1$. *In this case, we have*

$$\begin{aligned}
&\widehat{\boldsymbol{\alpha}}_1^{\mathrm{T}}\widehat{\mathbf{B}}\widehat{\boldsymbol{\alpha}}_1(\tau^{(1)}\|\boldsymbol{\alpha}_1\|_{\lambda^{(1)}}^2 - \tau^{(1)}\|\widehat{\boldsymbol{\alpha}}_1\|_{\lambda^{(1)}}^2) \\
\leq&(1+\hbar^{-1}D_1)^2(1+D_1)A^{(1)}C_0\mu_1(\boldsymbol{\Xi})^{1/2}\varpi\|(\boldsymbol{\alpha}_1 - \widehat{\boldsymbol{\alpha}}_1)_{J_1}\|_1,
\end{aligned}$$

where $J_1 = J(\boldsymbol{\alpha}_1)$ *is the collection of the nonzero coordinates of* $\boldsymbol{\alpha}_1$ *and* J_1^c *is its complement set in the set* $\{1, 2, \cdots, p\}$.

In the following, we will consider the two cases separately.

Case (a). By (1.6) and Lemma 4,

$$\begin{aligned}
&\frac{1}{2}c_2\mu_1(\boldsymbol{\Xi})\|\widehat{\boldsymbol{\gamma}}_1 - \boldsymbol{\gamma}_1\|_2^2 \\
\leq &\{4 + \hbar^{-1}(1 + D_1)\}(1 + D_1^2 A^{(1)}\hbar^{-1})C_0\sqrt{\mu_1(\boldsymbol{\Xi})}\varpi\|\boldsymbol{\alpha}_1 - \widehat{\boldsymbol{\alpha}}_1\|_1 \\
&+ (1 - 2\hbar^{-1})A^{(1)}C_0\sqrt{\mu_1(\boldsymbol{\Xi})}\varpi\|(\widehat{\boldsymbol{\alpha}}_1 - \boldsymbol{\alpha}_1)_{J_1}\|_1 \\
&- (1 - 2\hbar^{-1})A^{(1)}C_0\sqrt{\mu_1(\boldsymbol{\Xi})}\varpi\lambda^{(1)}\|(\widehat{\boldsymbol{\alpha}}_1)_{J_1^c}\|_1,
\end{aligned}$$

which, together with the equality $\|\widehat{\boldsymbol{\alpha}}_1 - \boldsymbol{\alpha}_1\|_1 = \|(\widehat{\boldsymbol{\alpha}}_1 - \boldsymbol{\alpha}_1)_{J_1}\|_1 + \|(\widehat{\boldsymbol{\alpha}}_1)_{J_1^c}\|_1$, leads to

$$\begin{aligned}
0 \leq &\frac{1}{2}c_2\mu_1(\boldsymbol{\Xi})\|\widehat{\boldsymbol{\gamma}}_1 - \boldsymbol{\gamma}_1\|_2^2 \\
\leq &\Big[(1 - 2\hbar^{-1})A^{(1)} + \{4 + \hbar^{-1}(1 + D_1)\}(1 + D_1^2 A^{(1)}\hbar^{-1})\Big] \\
&\times \sqrt{\mu_1(\boldsymbol{\Xi})}C_0\varpi\|(\widehat{\boldsymbol{\alpha}}_1 - \boldsymbol{\alpha}_1)_{J_1}\|_1 \\
&- \Big[(1 - 2\hbar^{-1})A^{(1)}\lambda^{(1)} - \{4 + \hbar^{-1}(1 + D_1)\}(1 + D_1^2 A^{(1)}\hbar^{-1})\Big] \\
&\times \sqrt{\mu_1(\boldsymbol{\Xi})}C_0\varpi\|(\widehat{\boldsymbol{\alpha}}_1)_{J_1^c}\|_1.
\end{aligned} \tag{1.7}$$

Therefore, it follows from (1.7) that

$$\begin{aligned}
&\Big[(1 - 2\hbar^{-1})A^{(1)} + \{4 + \hbar^{-1}(1 + D_1)\}(1 + D_1^2 A^{(1)}\hbar^{-1})\Big] \\
&\times \sqrt{\mu_1(\boldsymbol{\Xi})}C_0\varpi\|(\widehat{\boldsymbol{\alpha}}_1 - \boldsymbol{\alpha}_1)_{J_1}\|_1 \\
\geq &\Big[(1 - 2\hbar^{-1})A^{(1)}\lambda^{(1)} - \{4 + \hbar^{-1}(1 + D_1)\}(1 + D_1^2 A^{(1)}\hbar^{-1})\Big] \\
&\times \sqrt{\mu_1(\boldsymbol{\Xi})}C_0\varpi\|(\widehat{\boldsymbol{\alpha}}_1)_{J_1^c}\|_1\ .
\end{aligned} \tag{1.8}$$

Note that

$$\lim_{\substack{A^{(1)}\to\infty \\ \hbar\to\infty}} \frac{(1 - 2\hbar^{-1})A^{(1)} + \{4 + \hbar^{-1}(1 + D_1)\}(1 + D_1^2 A^{(1)}\hbar^{-1})}{(1 - 2\hbar^{-1})A^{(1)}\lambda^{(1)} - \{4 + \hbar^{-1}(1 + D_1)\}(1 + D_1^2 A^{(1)}\hbar^{-1})} \to (\lambda^{(1)})^{-1},$$

and by (4.4), $(\lambda^{(1)})^{-1} < (c^{-1} + \delta_0)^{-1} < c$. Then by (1.8), there exist $(A_1^L)''$ and $(\hbar_0)''$ only depending on $D_1 = \sqrt{6c}$ (Lemma 3) and δ_0 such that for any

$A^{(1)} \geq (A_1^L)''$ and $\hbar \geq (\hbar_0)''$,

$$\|(\widehat{\boldsymbol{\alpha}}_1 - \boldsymbol{\alpha}_1)_{J_1^c}\|_1 = \|(\widehat{\boldsymbol{\alpha}}_1)_{J_1^c}\|_1 < c\|(\widehat{\boldsymbol{\alpha}}_1 - \boldsymbol{\alpha}_1)_{J_1}\|_1. \tag{1.9}$$

By Condition 1, (1.9) and the inequality $|J_1| = |J(\boldsymbol{\alpha}_1)| \leq |J(\boldsymbol{\beta}(t))| = \mathcal{M}(\boldsymbol{\beta}(t)) = s$, we have

$$\kappa\|(\widehat{\boldsymbol{\alpha}}_1 - \boldsymbol{\alpha}_1)_{J_1}\|_2 \leq \|\frac{\mathbf{X}}{\sqrt{n}}(\widehat{\boldsymbol{\alpha}}_1 - \boldsymbol{\alpha}_1)\|_2 = \|\mathbf{Z}(\widehat{\boldsymbol{\alpha}}_1 - \boldsymbol{\alpha}_1)\|_2 = \|\widehat{\boldsymbol{\gamma}}_1 - \boldsymbol{\gamma}_1\|_2, \tag{1.10}$$

which, together with (1.7) and the Cauchy-Schwarz inequality, leads to

$$\begin{aligned}
&\frac{1}{2}c_2\mu_1(\boldsymbol{\Xi})\|\widehat{\boldsymbol{\gamma}}_1 - \boldsymbol{\gamma}_1\|_2^2 \\
\leq &\Big[(1-2\hbar^{-1})A^{(1)} + \{4+\hbar^{-1}(1+D_1)\}(1+D_1^2A^{(1)}\hbar^{-1})\Big] \\
&\quad \times \sqrt{\mu_1(\boldsymbol{\Xi})}C_0\varpi\|(\widehat{\boldsymbol{\alpha}}_1 - \boldsymbol{\alpha}_1)_{J_1}\|_1 \\
\leq &\Big[(1-2\hbar^{-1})A^{(1)} + \{4+\hbar^{-1}(1+D_1)\}(1+D_1^2A^{(1)}\hbar^{-1})\Big] \\
&\quad \times \sqrt{\mu_1(\boldsymbol{\Xi})}C_0\varpi\sqrt{|J_1|}\|(\widehat{\boldsymbol{\alpha}}_1 - \boldsymbol{\alpha}_1)_{J_1}\|_2 \\
\leq &\Big[(1-2\hbar^{-1})A^{(1)} + \{4+\hbar^{-1}(1+D_1)\}(1+D_1^2A^{(1)}\hbar^{-1})\Big] \\
&\quad \times \sqrt{\mu_1(\boldsymbol{\Xi})}C_0\varpi\sqrt{s}\|\widehat{\boldsymbol{\gamma}}_1 - \boldsymbol{\gamma}_1\|_2/\kappa.
\end{aligned}$$

Then it follows that

$$\begin{aligned}
&\|\widehat{\boldsymbol{\gamma}}_1 - \boldsymbol{\gamma}_1\|_2 \leq D_2C_0\kappa^{-1}\mu_1(\boldsymbol{\Xi})^{-1/2}\varpi\sqrt{s}, \\
&\|\mathbf{Z}(\widehat{\boldsymbol{\alpha}}_1 - \boldsymbol{\alpha}_1)\|_2^2 = \|\widehat{\boldsymbol{\gamma}}_1 - \boldsymbol{\gamma}_1\|_2^2 \leq D_2^2C_0^2\kappa^{-2}\mu_1(\boldsymbol{\Xi})^{-1}\varpi^2 s, \\
&\|\mathbf{X}(\widehat{\boldsymbol{\alpha}}_1 - \boldsymbol{\alpha}_1)\|_2^2 = n\|\widehat{\boldsymbol{\gamma}}_1 - \boldsymbol{\gamma}_1\|_2^2 \leq nD_2^2C_0^2\kappa^{-2}\mu_1(\boldsymbol{\Xi})^{-1}\varpi^2 s,
\end{aligned} \tag{1.11}$$

where $D_2 = 2c_2^{-1}\left[(1-2\hbar^{-1})A^{(1)} + \{4+\hbar^{-1}(1+D_1)\}(1+D_1^2A^{(1)}\hbar^{-1})\right]$. By (1.9)-(1.11),

$$\begin{aligned}
&\|\widehat{\boldsymbol{\alpha}}_1 - \boldsymbol{\alpha}_1\|_1 = \|(\widehat{\boldsymbol{\alpha}}_1 - \boldsymbol{\alpha}_1)_{J_1^c}\|_1 + \|(\widehat{\boldsymbol{\alpha}}_1 - \boldsymbol{\alpha}_1)_{J_1}\|_1 \\
<&(1+c)\|(\widehat{\boldsymbol{\alpha}}_1 - \boldsymbol{\alpha}_1)_{J_1}\|_1 \leq (1+c)\sqrt{s}\|(\widehat{\boldsymbol{\alpha}}_1 - \boldsymbol{\alpha}_1)_{J_1}\|_2 \\
\leq&(1+c)\sqrt{s}\kappa^{-1}\|\widehat{\boldsymbol{\gamma}}_1 - \boldsymbol{\gamma}_1\|_2 \leq (1+c)D_2C_0\kappa^{-2}\mu_1(\boldsymbol{\Xi})^{-1/2}\varpi s.
\end{aligned} \tag{1.12}$$

Note that

$$\lim_{\substack{A^{(1)}\to\infty \\ \hbar\to\infty}} \frac{D_2}{A^{(1)}} = 2c_2^{-1} \leq 2c_2^{-1}(1+D_1) = 2c_2^{-1}(1+\sqrt{6c}). \tag{1.13}$$

Hence, there exist $(A_1^L)'''$ and $(\hbar_0)'''$ only depending on δ_0, c and c_2 such that for any $A^{(1)} \geq (A_1^L)'''$ and $\hbar \geq (\hbar_0)'''$, we have

$$D_2 \leq 4c_2^{-1}(1+\sqrt{6c})A^{(1)},$$

and the inequalities in Theorem 4.2 (a) follow from (1.11) and (1.12) in this case.

Case (b). The arguments are similar to Case (a). We summarize the results in the following lemma and provide the proof of the lemma in supplementary materials.

Lemma 5. *In the event Ω and Case (b), there exist $(A_1^L)''''$ and $(\hbar_0)''''$ only depending on δ_0, c and c_2 such that for any $A^{(1)} \geq (A_1^L)''''$ and $\hbar \geq (\hbar_0)''''$, the inequalities in Theorem 4.2 (a) hold.*

Finally, we choose

$$A_1^L = \max\{(A_1^L)', (A_1^L)'', (A_1^L)''', (A_1^L)'''' \quad h_0 = \max\{(h_0)', (h_0)'', (h_0)''', (h_0)''''\}.$$

It can be seen that A_1^L and h_0 only depend on c, c_2 and δ_0 and the inequalities in Theorem 4.2 (a) hold if $A^{(1)} \geq A_1^L$ and $\hbar \geq \hbar_0$. The proof of Theorem 4.2 (a) is completed.

- *Proof of Theorem 4.2 (b).*

We will proceed by induction to prove a set of inequalities from which Part (b) follows. In this proof, we only consider the elements in the event Ω defined in the proof of Part (a).

- **Step 1: Introduction of some notations and inequalities.**

By the first inequality in the condition (4.6), for any $1 \leq k \leq K$, we have

$$c_4^{-1}\|\boldsymbol{\alpha}_1\|_1 \leq \|\boldsymbol{\alpha}_k\|_1 \leq c_4\|\boldsymbol{\alpha}_1\|_1, \tag{1.14}$$

which together with the condition (4.3) and Lemma 2 lead to

$$\begin{aligned} C_0\varpi\|\boldsymbol{\alpha}_k\|_1 &\le C_0\varpi\sqrt{s}/\kappa \le \hbar^{-1}\mu_1(\boldsymbol{\Xi})^{1/2}, \\ \tau^{(k)}\|\boldsymbol{\alpha}_k\|^2_{\lambda^{(k)}} &\le \tau^{(k)}\|\boldsymbol{\alpha}_k\|_1^2 \le c_4^2\tau^{(k)}\|\boldsymbol{\alpha}_1\|_1^2 \\ &= \frac{c_4^2A^{(k)}C_0\varpi\|\boldsymbol{\alpha}_1\|_1^2}{\|\boldsymbol{\alpha}_1\|_1\sqrt{\mu_1(\boldsymbol{\Xi})}} \le c_4^2A^{(k)}\hbar^{-1}, \end{aligned} \tag{1.15}$$

for any $1 \le k \le K$. Note that

$$\begin{aligned} \boldsymbol{\alpha}_k^{\mathrm{T}}\mathbf{S}\boldsymbol{\alpha}_k &= \|\boldsymbol{\gamma}_k\|_2^2 = 1, \qquad \widehat{\boldsymbol{\alpha}}_k^{\mathrm{T}}\mathbf{S}\widehat{\boldsymbol{\alpha}}_k = \|\widehat{\boldsymbol{\gamma}}_k\|_2^2 = 1, \\ \boldsymbol{\alpha}_k^{\mathrm{T}}\mathbf{S}\boldsymbol{\alpha}_l &= \boldsymbol{\gamma}_k^{\mathrm{T}}\boldsymbol{\gamma}_l = 0, \qquad \widehat{\boldsymbol{\alpha}}_k^{\mathrm{T}}\mathbf{S}\widehat{\boldsymbol{\alpha}}_l = \widehat{\boldsymbol{\gamma}}_k^{\mathrm{T}}\widehat{\boldsymbol{\gamma}}_l = 0, \end{aligned} \tag{1.16}$$

for any $1 \le k \ne l \le K$. Define the following subspaces which are spanned by different sets of vectors. The first two are subspaces in $\mathbb{R}^n$ and the last two in $\mathbb{R}^p$.

$$\begin{aligned} &\mathbf{V}_k = \mathrm{span}\{\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2, \cdots, \boldsymbol{\gamma}_k\}, \quad \widehat{\mathbf{V}}_k = \mathrm{span}\{\widehat{\boldsymbol{\gamma}}_1, \widehat{\boldsymbol{\gamma}}_2, \cdots, \widehat{\boldsymbol{\gamma}}_k\}, \\ &\mathbf{W}_k = \mathrm{span}\{\mathbf{S}\boldsymbol{\alpha}_1, \mathbf{S}\boldsymbol{\alpha}_2, \cdots, \mathbf{S}\boldsymbol{\alpha}_k\}, \quad \widehat{\mathbf{W}}_k = \mathrm{span}\{\mathbf{S}\widehat{\boldsymbol{\alpha}}_1, \mathbf{S}\widehat{\boldsymbol{\alpha}}_2, \cdots, \mathbf{S}\widehat{\boldsymbol{\alpha}}_k\}, \end{aligned}$$

and

$$\mathbf{P}_k = \sum_{i=1}^k \boldsymbol{\gamma}_i\boldsymbol{\gamma}_i^{\mathrm{T}}, \quad \widehat{\mathbf{P}}_k = \sum_{i=1}^k \widehat{\boldsymbol{\gamma}}_i\widehat{\boldsymbol{\gamma}}_i^{\mathrm{T}}$$

be the orthogonal projection matrices onto $\mathbf{V}_k$ and $\widehat{\mathbf{V}}_k$, respectively. Let

$$\begin{aligned} \boldsymbol{\delta}_k &= \sum_{i=1}^{k-1}(\boldsymbol{\gamma}_k^{\mathrm{T}}\widehat{\boldsymbol{\gamma}}_i)\widehat{\boldsymbol{\alpha}}_i, \quad \boldsymbol{\beta}_k = \boldsymbol{\alpha}_k - \boldsymbol{\delta}_k, \\ \widehat{\boldsymbol{\delta}}_k &= \sum_{i=1}^{k-1}(\widehat{\boldsymbol{\gamma}}_k^{\mathrm{T}}\boldsymbol{\gamma}_i)\boldsymbol{\alpha}_i, \quad \widehat{\boldsymbol{\beta}}_k = \widehat{\boldsymbol{\alpha}}_k - \widehat{\boldsymbol{\delta}}_k, \end{aligned} \tag{1.17}$$

which play an important role in the proof. We have $\mathbf{Z}\boldsymbol{\delta}_k = \sum_{i=1}^{k-1}(\boldsymbol{\gamma}_k^{\mathrm{T}}\widehat{\boldsymbol{\gamma}}_i)\widehat{\boldsymbol{\gamma}}_i = \widehat{\mathbf{P}}_{k-1}\boldsymbol{\gamma}_k$ and similarly,

$$\mathbf{Z}\boldsymbol{\beta}_k = (\mathbf{I} - \widehat{\mathbf{P}}_{k-1})\boldsymbol{\gamma}_k, \quad \mathbf{Z}\widehat{\boldsymbol{\delta}}_k = \mathbf{P}_{k-1}\widehat{\boldsymbol{\gamma}}_k, \quad \mathbf{Z}\widehat{\boldsymbol{\beta}}_k = (\mathbf{I} - \mathbf{P}_{k-1})\widehat{\boldsymbol{\gamma}}_k, \tag{1.18}$$

for all $1 \le k \le K$. For any $1 \le i \le k-1$, $(\mathbf{S}\widehat{\boldsymbol{\alpha}}_i)^{\mathrm{T}}\boldsymbol{\beta}_k = \widehat{\boldsymbol{\alpha}}_i^{\mathrm{T}}\mathbf{Z}^{\mathrm{T}}\mathbf{Z}\boldsymbol{\beta}_k = \widehat{\boldsymbol{\gamma}}_i^{\mathrm{T}}(\mathbf{I} - \widehat{\mathbf{P}}_{k-1})\boldsymbol{\gamma}_k = 0$, therefore, we have

$$\boldsymbol{\beta}_k \perp \widehat{\mathbf{W}}_{k-1}, \quad \text{and similarly,} \quad \widehat{\boldsymbol{\beta}}_k \perp \mathbf{W}_{k-1}, \tag{1.19}$$

for any $1 \le k \le K$.

- **Step 2: Induction hypothesis.**

We will prove that we can find constants $\hbar_0$, $A_j^L < A_j^U$, $1 \le j \le K$, which only depend on δ_0, c, $c_2 \sim c_5$, such that for any $1 \le i \le K$, if $A_j^L \le A^{(j)} \le A_j^U$, $1 \le j < i$, $\hbar \ge \hbar_0$ and $A^{(i)} \ge A_i^L$, we have

$$\begin{aligned}
&\|\widehat{\boldsymbol{\alpha}}_i\|_1 \le D_{i,1}\kappa^{-1}\sqrt{s}, \\
&\|\widehat{\boldsymbol{\gamma}}_i - \boldsymbol{\gamma}_i\|_2^2 \le (A^{(i)})^2 D_{i,2} C_0^2 \kappa^{-2}\mu_1(\boldsymbol{\Xi})^{-1}\varpi^2 s, \\
&\|\widehat{\mathbf{P}}_i - \mathbf{P}_i\|_2^2 \le (A^{(i)})^2 D_{i,3} C_0^2 \kappa^{-2}\mu_1(\boldsymbol{\Xi})^{-1}\varpi^2 s, \\
&\|\widehat{\boldsymbol{\alpha}}_i - \boldsymbol{\alpha}_i\|_1 \le D_{i,4} A^{(i)} \mu_1(\boldsymbol{\Xi})^{-1/2}\kappa^{-2} C_0 \varpi s
\end{aligned} \tag{1.20}$$

where $D_{i,1} \sim D_{i,4}$ are constants only depending on δ_0, c and $c_2 \sim c_5$. We will proceed by induction.

- **Step 3: Proof of (1.20) for $i = 1$.**

When $i = 1$, by Theorem 4.2 (a), we can find constants A_1^L and $\hbar_0^{(1)}$ such that the inequalities in (1.20) except the third one hold for any $A^{(1)} \ge A_1^L$ and $\hbar \ge \hbar_0^{(1)}$ with $D_{1,1} = \sqrt{6c}$, $D_{1,2} = 16c_2^{-2}(1+\sqrt{6c})^2$ and $D_{1,4} = 4(1+c)(1+\sqrt{6c})c_2^{-1}$. The third inequality follows from

$$\begin{aligned}
&\|\widehat{\mathbf{P}}_1 - \mathbf{P}_1\|_2^2 = \|\widehat{\boldsymbol{\gamma}}_1\widehat{\boldsymbol{\gamma}}_1^{\mathrm{T}} - \boldsymbol{\gamma}_1\boldsymbol{\gamma}_1^{\mathrm{T}}\|^2 \\
\le& 2\|\widehat{\boldsymbol{\gamma}}_1 - \boldsymbol{\gamma}_1\|_2^2\|\widehat{\boldsymbol{\gamma}}_1\|_2^2 + 2\|\widehat{\boldsymbol{\gamma}}_1 - \boldsymbol{\gamma}_1\|_2^2\|\boldsymbol{\gamma}_1\|_2^2 \le 4\|\widehat{\boldsymbol{\gamma}}_1 - \boldsymbol{\gamma}_1\|_2^2
\end{aligned}$$

with $D_{1,3} = 4D_{1,4}$. Therefore, $D_{1,1} \sim D_{1,4}$ only depends on c and c_2. We arbitrarily choose a number A_1^U such that $A_1^U > A_1^L$.

- **Step 4: Inductive step, the proof of (1.20) for $i = k$.**

Now we assume that we have found constants $A_j^L < A_j^U$, $1 \le j \le k-1$ and $\hbar_0^{(k-1)}$ which only depend on δ_0, c, $c_2 \sim c_5$, such that (1.20) hold for any $i < k$

if $A_j^L \leq A^{(j)} \leq A_j^U$, $1 \leq j \leq i-1$, $\hbar \geq \hbar_0^{(k-1)}$ and $A^{(i)} \geq A_i^L$. Based on these assumptions, we will find A_k^L and $\hbar_0^{(k)} \geq \hbar_0^{(k-1)}$ such that (1.20) are true for $i = k$ when $A_j^L \leq A^{(j)} \leq A_j^U$, $1 \leq j \leq k-1$, $\hbar \geq \hbar_0^{(k)}$ and $A^{(k)} \geq A_k^L$. We choose an arbitrary $A_k^U > A_k^L$. Then by induction the claims (1.20) are true for all $1 \leq k \leq K$ with $\hbar_0 = \hbar_0^{(K-1)}$.

Define

$$b_{i,1} = D_{i,1}, \quad b_{i,2} = D_{i,2}(A_i^U)^2, \quad b_{i,3} = (A_i^U)^2 D_{i,3}, \quad 1 \leq i \leq k-1. \tag{1.21}$$

Then $b_{i,1}$, $b_{i,2}$ and $b_{i,3}$ only depend on δ_0, c and $c_2 \sim c_5$. It follows from (1.20) and (1.21) that for any $1 \leq i \leq k-1$,

$$\begin{aligned} &\|\widehat{\boldsymbol{\alpha}}_i\|_1 \leq b_{i,1}\kappa^{-1}\sqrt{s}, \quad \|\widehat{\boldsymbol{\gamma}}_i - \boldsymbol{\gamma}_i\|_2^2 \leq b_{i,2} C_0^2 \kappa^{-2} \mu_1(\boldsymbol{\Xi})^{-1} \varpi^2 s, \\ &\|\widehat{\mathbf{P}}_i - \mathbf{P}_i\|_2^2 \leq b_{i,3} C_0^2 \kappa^{-2} \mu_1(\boldsymbol{\Xi})^{-1} \varpi^2 s. \end{aligned} \tag{1.22}$$

for all $\hbar \geq \hbar_0^{(k-1)}$ and all $A_j^L \leq A^{(j)} \leq A_j^U$, $1 \leq j \leq i$.

We provide several inequalities related to $\boldsymbol{\delta}_k$ and $\boldsymbol{\beta}_k$. By (1.14), (1.16), (1.22) and the definitions of $\boldsymbol{\delta}_k$ and $\boldsymbol{\beta}_k$ in (1.17),

$$\begin{aligned} &\|\boldsymbol{\delta}_k\|_1 \leq \sum_{i=1}^{k-1} |\boldsymbol{\gamma}_k^{\mathrm{T}} \widehat{\boldsymbol{\gamma}}_i| \|\widehat{\boldsymbol{\alpha}}_i\|_1 = \sum_{i=1}^{k-1} |\boldsymbol{\gamma}_k^{\mathrm{T}} \widehat{\boldsymbol{\gamma}}_i - \boldsymbol{\gamma}_k^{\mathrm{T}} \boldsymbol{\gamma}_i| \|\widehat{\boldsymbol{\alpha}}_i\|_1 \\ &\leq \sum_{i=1}^{k-1} \|\widehat{\boldsymbol{\gamma}}_i - \boldsymbol{\gamma}_i\|_2 \|\widehat{\boldsymbol{\alpha}}_i\|_1 \leq \sum_{i=1}^{k-1} b_{i,2}^{1/2} C_0 \kappa^{-1} \mu_1(\boldsymbol{\Xi})^{-1/2} \varpi \sqrt{s} b_{i,1} \kappa^{-1} \sqrt{s} \\ &= M_{k,0} \kappa^{-2} \mu_1(\boldsymbol{\Xi})^{-1/2} C_0 \varpi s \leq M_{k,0} c_5 \|\boldsymbol{\alpha}_1\|_1, \end{aligned} \tag{1.23}$$

$$\begin{aligned} \text{and} \quad &\|\boldsymbol{\beta}_k\|_1 \leq \|\boldsymbol{\alpha}_k\|_1 + \|\boldsymbol{\delta}_k\|_1 \leq c_4 \|\boldsymbol{\alpha}_1\|_1 + \|\boldsymbol{\delta}_k\|_1 \\ &\leq M_{k,1} \|\boldsymbol{\alpha}_1\|_1 \leq M_{k,1} \sqrt{s}/\kappa, \end{aligned} \tag{1.24}$$

where

$$M_{k,0} = \sum_{i=1}^{k-1} b_{i,2}^{1/2} b_{i,1}, \quad M_{k,1} = \left(c_4 + M_{k,0} c_5\right), \tag{1.25}$$

the inequality in the third line is due to the second inequality in the condition

(4.6) and the last inequality follows from Lemma 2. (1.24) and (1.15) lead to

$$
\begin{aligned}
\tau^{(k)}\|\boldsymbol{\beta}_k\|_1^2 \leq & \frac{A^{(k)}C_0\varpi M_{k,1}^2\|\boldsymbol{\alpha}_1\|_1^2}{\|\boldsymbol{\alpha}_1\|_1\sqrt{\mu_1(\boldsymbol{\Xi})}} = A^{(k)}M_{k,1}^2\frac{C_0\varpi\|\boldsymbol{\alpha}_1\|_1}{\sqrt{\mu_1(\boldsymbol{\Xi})}} \\
\leq & A^{(k)}M_{k,1}^2\kappa^{-1}\frac{C_0\varpi\sqrt{s}}{\sqrt{\mu_1(\boldsymbol{\Xi})}} \leq A^{(k)}M_{k,1}^2\hbar^{-1}.
\end{aligned} \tag{1.26}
$$

Similar to the proof of Part (a), we next provide a key inequality (1.28) based on which the proof is. $\widehat{\boldsymbol{\alpha}}_k$ is the solution to the optimization problem (3.10). Because $\widehat{\boldsymbol{\alpha}}_k$ is the solution to (3.10) which has a scale-invariant objective function, it is also the solution to

$$
\max_{\boldsymbol{\alpha}\neq\mathbf{0},\quad \boldsymbol{\alpha}\perp\widehat{\mathbf{W}}_{k-1}} \frac{\boldsymbol{\alpha}^{\mathrm{T}}\widehat{\mathbf{B}}\boldsymbol{\alpha}}{\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{S}\boldsymbol{\alpha}+\tau^{(k)}\|\boldsymbol{\alpha}\|_{\lambda^{(k)}}^2},
$$

where we do not imose the constraint, $\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{S}\boldsymbol{\alpha}=1$. Therefore, since $\boldsymbol{\beta}_k\perp\widehat{\mathbf{W}}_{k-1}$ by (1.19), we have

$$
\frac{\boldsymbol{\beta}_k^{\mathrm{T}}\widehat{\mathbf{B}}\boldsymbol{\beta}_k}{\boldsymbol{\beta}_k^{\mathrm{T}}\mathbf{S}\boldsymbol{\beta}_k+\tau^{(k)}\|\boldsymbol{\beta}_k\|_{\lambda^{(k)}}^2} \leq \frac{\widehat{\boldsymbol{\alpha}}_k^{\mathrm{T}}\widehat{\mathbf{B}}\widehat{\boldsymbol{\alpha}}_k}{1+\tau^{(k)}\|\widehat{\boldsymbol{\alpha}}_k\|_{\lambda^{(k)}}^2}, \tag{1.27}
$$

which is one of the key inequalities in this proof.

<u>**Lemma**</u> **6.**

$$
\begin{aligned}
&\boldsymbol{\beta}_k^{\mathrm{T}}\mathbf{S}\boldsymbol{\beta}_k \leq 1, \\
&\boldsymbol{\beta}_k^{\mathrm{T}}\widehat{\mathbf{B}}\boldsymbol{\beta}_k \geq \mu_k(\boldsymbol{\Xi}) - N_{k,3}C_0\kappa^{-1}\mu_1(\boldsymbol{\Xi})^{1/2}\varpi\sqrt{s}, \\
&\widehat{\boldsymbol{\alpha}}_k^{\mathrm{T}}\widehat{\mathbf{B}}\widehat{\boldsymbol{\alpha}}_k \leq \mu_k(\boldsymbol{\Xi}) + (b_{k-1,3}\hbar^{-1}+1)\sqrt{\mu_1(\boldsymbol{\Xi})}\kappa^{-1}C_0\varpi\sqrt{s} \\
&\qquad + \frac{(1+\hbar^{-1})}{\sqrt{s}}\kappa\sqrt{\mu_1(\boldsymbol{\Xi})}C_0\varpi\|\widehat{\boldsymbol{\alpha}}_k\|_1^2,
\end{aligned}
$$

where $N_{k,3}=2(b_{k-1,3}\hbar^{-1}+M_{k,1})$.

By (1.27) and the first inequality in Lemma 6,

$$
\begin{aligned}
&(\boldsymbol{\beta}_k^{\mathrm{T}}\widehat{\mathbf{B}}\boldsymbol{\beta}_k)(1+\tau^{(k)}\|\widehat{\boldsymbol{\alpha}}_k\|_{\lambda^{(k)}}^2) \leq (\widehat{\boldsymbol{\alpha}}_k^{\mathrm{T}}\widehat{\mathbf{B}}\widehat{\boldsymbol{\alpha}}_k)(\boldsymbol{\beta}_k^{\mathrm{T}}\mathbf{S}\boldsymbol{\beta}_k+\tau^{(k)}\|\boldsymbol{\beta}_k\|_{\lambda^{(k)}}^2) \\
\leq&(\widehat{\boldsymbol{\alpha}}_k^{\mathrm{T}}\widehat{\mathbf{B}}\widehat{\boldsymbol{\alpha}}_k)(1+\tau^{(k)}\|\boldsymbol{\beta}_k\|_{\lambda^{(k)}}^2).
\end{aligned} \tag{1.28}
$$

Based on (1.28) and Lemma 6, we provide an upper bound for $\|\widehat{\boldsymbol{\alpha}}_k\|_1$ in the following lemma.

Lemma 7. *there exist $(A_k^L)'$ and $(\hbar_0)'$ only depending on δ_0, c and $c_2 \sim c_5$ such that for any $A^{(k)} \geq (A_k^L)'$ and $\hbar \geq (\hbar_0)'$, we have*

$$\|\widehat{\boldsymbol{\alpha}}_k\|_1 \leq D_{k,1}\kappa^{-1}\sqrt{s}. \tag{1.29}$$

where $D_{k,1} = 2M_{k,1}\sqrt{c_3 c}$.

Next, by (1.28), we have

$$\begin{aligned}&(\boldsymbol{\beta}_k^{\mathrm{T}}\widehat{\mathbf{B}}\boldsymbol{\beta}_k - \widehat{\boldsymbol{\alpha}}_k^{\mathrm{T}}\widehat{\mathbf{B}}\widehat{\boldsymbol{\alpha}}_k)(1+\tau^{(k)}\|\widehat{\boldsymbol{\alpha}}_k\|_{\lambda^{(k)}}^2)\\ \leq&(\widehat{\boldsymbol{\alpha}}_k^{\mathrm{T}}\widehat{\mathbf{B}}\widehat{\boldsymbol{\alpha}}_k)(\tau^{(k)}\|\boldsymbol{\beta}_k\|_{\lambda^{(k)}}^2 - \tau^{(k)}\|\widehat{\boldsymbol{\alpha}}_k\|_{\lambda^{(k)}}^2),\end{aligned} \tag{1.30}$$

which leads to the following lemma.

Lemma 8. *We have either*

$$\begin{aligned}&\frac{1}{2}c_2\mu_k(\boldsymbol{\Xi})\|\boldsymbol{\gamma}_k - \widehat{\boldsymbol{\gamma}}_k\|_2^2\\ \leq& N_{k,4}C_0^2\kappa^{-2}\varpi^2 s + N_{k,5}C_0\sqrt{\mu_1(\boldsymbol{\Xi})}\varpi\|\boldsymbol{\beta}_k - \widehat{\boldsymbol{\alpha}}_k\|_1\\ &+ N_{k,6}\mu_1(\boldsymbol{\Xi})[\tau^{(k)}\|\boldsymbol{\beta}_k\|_{\lambda^{(k)}}^2 - \tau^{(k)}\|\widehat{\boldsymbol{\alpha}}_k\|_{\lambda^{(k)}}^2].\end{aligned} \tag{1.31}$$

or

$$\begin{aligned}&\frac{1}{2}c_2\mu_k(\boldsymbol{\Xi})\|\boldsymbol{\gamma}_k - \widehat{\boldsymbol{\gamma}}_k\|_2^2\\ \leq& N_{k,4}C_0^2\kappa^{-2}\varpi^2 s + N_{k,5}C_0\sqrt{\mu_1(\boldsymbol{\Xi})}\varpi\|\boldsymbol{\beta}_k - \widehat{\boldsymbol{\alpha}}_k\|_1\\ &+ N_{k,7}\mu_1(\boldsymbol{\Xi})[\tau^{(k)}\|\boldsymbol{\beta}_k\|_{\lambda^{(k)}}^2 - \tau^{(k)}\|\widehat{\boldsymbol{\alpha}}_k\|_{\lambda^{(k)}}^2].\end{aligned} \tag{1.32}$$

where

$$\begin{aligned}N_{k,4} &= 3b_{k-1,3}N_{k,6}N_{k,7}^{-1}(1+A^{(k)}M_{k,1}^2\hbar^{-1}),\\ N_{k,5} &= [4+(M_{k,1}+D_{k,1})\hbar^{-1}]N_{k,6}N_{k,7}^{-1}(1+A^{(k)}M_{k,1}^2\hbar^{-1})\\ N_{k,6} &= 1+(b_{k-1,3}\hbar^{-1}+1)\hbar^{-1}+(1+\hbar^{-1})D_{k,1}^2\hbar^{-1},\\ N_{k,7} &= c_3 - 2(b_{k-1,3}\hbar^{-1}+M_{k,1})\hbar^{-1}.\end{aligned}$$

The two inequalities (1.31) and (1.32) leads to the same upper bounds except the constants. Hence, we only prove the results for the (1.31) and exactly the same arguments can be applied to (1.32).

Lemma 9. *By (1.31), we have*

$$\frac{1}{2}c_2\mu_k(\boldsymbol{\Xi})\|\boldsymbol{\gamma}_k-\widehat{\boldsymbol{\gamma}}_k\|_2^2 \le N_{k,1}C_0^2\kappa^{-2}\varpi^2 s + (N_{k,2}+N_{k,5})\mu_1(\boldsymbol{\Xi})^{1/2}C_0\varpi\|(\boldsymbol{\alpha}_k-\widehat{\boldsymbol{\alpha}}_k)_{J_k}\|_1 - (\lambda^{(k)}N_{k,2}-N_{k,5})\mu_1(\boldsymbol{\Xi})^{1/2}C_0\varpi\|(\widehat{\boldsymbol{\alpha}}_k)_{J_k^c}\|_1, \quad (1.33)$$

where

$$N_{k,1} = N_{k,4} + N_{k,5}M_{k,0} + 2A^{(k)}N_{k,6}M_{k,1}M_{k,0}$$
$$N_{k,2} = A^{(k)}N_{k,6}(2\|\boldsymbol{\alpha}_k\|_1/\|\boldsymbol{\alpha}_1\|_1).$$

We will consider the following two cases separately,

$$N_{k,1}C_0^2\kappa^{-2}\varpi^2 s \le \nu_0(N_{k,2}+N_{k,5})\mu_1(\boldsymbol{\Xi})^{1/2}C_0\varpi\|(\boldsymbol{\alpha}_k-\widehat{\boldsymbol{\alpha}}_k)_{J_k}\|_1$$

and

$$N_{k,1}C_0^2\kappa^{-2}\varpi^2 s > \nu_0(N_{k,2}+N_{k,5})\mu_1(\boldsymbol{\Xi})^{1/2}C_0\varpi\|(\boldsymbol{\alpha}_k-\widehat{\boldsymbol{\alpha}}_k)_{J_k}\|_1,$$

where

$$\nu_0 = (\delta_0/2)(c^{-1}+\delta_0/2)^{-1}.$$

- **Case 1:** $N_{k,1}C_0^2\kappa^{-2}\varpi^2 s \le \nu_0(N_{k,2}+N_{k,5})\mu_1(\boldsymbol{\Xi})^{1/2}C_0\varpi\|(\boldsymbol{\alpha}_k-\widehat{\boldsymbol{\alpha}}_k)_{J_k}\|_1$.

In this case, (1.33) leads to

$$\frac{1}{2}c_2\mu_k(\boldsymbol{\Xi})\|\boldsymbol{\gamma}_k-\widehat{\boldsymbol{\gamma}}_k\|_2^2 \quad (1.34)$$
$$\le(1+\nu_0)[N_{k,2}+N_{k,5}]\mu_1(\boldsymbol{\Xi})^{1/2}C_0\varpi\|(\boldsymbol{\alpha}_k-\widehat{\boldsymbol{\alpha}}_k)_{J_k}\|_1 - (\lambda^{(k)}N_{k,2}-N_{k,5})\mu_1(\boldsymbol{\Xi})^{1/2}C_0\varpi\|(\widehat{\boldsymbol{\alpha}}_k)_{J_k^c}\|_1.$$

Then we have

$$(\lambda^{(k)}N_{k,2}-N_{k,5})\mu_1(\boldsymbol{\Xi})^{1/2}C_0\varpi\|(\widehat{\boldsymbol{\alpha}}_k)_{J_k^c}\|_1 \le(1+\nu_0)[N_{k,2}+N_{k,5}]\mu_1(\boldsymbol{\Xi})^{1/2}C_0\varpi\|(\boldsymbol{\alpha}_k-\widehat{\boldsymbol{\alpha}}_k)_{J_k}\|_1. \quad (1.35)$$

By the definitions of $N_{k,5}$ and $N_{k,2}$ in Lemmas 8 and 9, the facts that $b_{i,1}$, $b_{i,2}$ $b_{i,3}$, $M_{k,0}$, $M_{k,1}$, and $D_{k,1}$, $1 \leq i \leq k-1$, only depend on δ_0, c and $c_2 \sim c_5$ (see (1.21), (1.25) and Lemma 7) and the inequality (1.14), we have

$$\lim_{\substack{A^{(k)}\to\infty \\ \hbar\to\infty}} \frac{(1+\nu_0)[N_{k,2}+N_{k,5}]}{\lambda^{(k)}N_{k,2}-N_{k,5}} = (1+\nu_0)(\lambda^{(k)})^{-1}$$
$$\leq (1+\nu_0)(c+\delta_0)^{-1} = (c^{-1}+\delta_0/2)^{-1} < c. \quad (1.36)$$

Therefore, by (1.35) and (1.36), as $A^{(k)}$ and $\hbar$ are large enough, we have $\|(\widehat{\boldsymbol{\alpha}}_k)_{J_k^c}\|_1 \leq c\|(\widehat{\boldsymbol{\alpha}}_k - \boldsymbol{\alpha}_k)_{J_k}\|_1$. Note that $(\boldsymbol{\alpha}_k)_{J_k^c} = \mathbf{0}$, hence

$$\|(\widehat{\boldsymbol{\alpha}}_k - \boldsymbol{\alpha}_k)_{J_k^c}\|_1 \leq c\|(\widehat{\boldsymbol{\alpha}}_k - \boldsymbol{\alpha}_k)_{J_k}\|_1, \quad (1.37)$$

which together with $|J_k| = |J(\boldsymbol{\alpha}_k)| \leq |J(\boldsymbol{\beta}(t))| = \mathcal{M}(\boldsymbol{\beta}(t)) = s$ and Condition 1, imply

$$\kappa\|(\widehat{\boldsymbol{\alpha}}_k - \boldsymbol{\alpha}_k)_{J_k}\|_2 \leq \|\frac{\mathbf{X}}{\sqrt{n}}(\widehat{\boldsymbol{\alpha}}_k - \boldsymbol{\alpha}_k)\|_2$$
$$= \|\mathbf{Z}(\widehat{\boldsymbol{\alpha}}_k - \boldsymbol{\alpha}_k)\|_2 = \|\widehat{\boldsymbol{\gamma}}_k - \boldsymbol{\gamma}_k\|_2. \quad (1.38)$$

By (1.38), (1.34) and the Cauchy-Schwarz inequality lead to

$$\frac{1}{2}c_2\mu_k(\boldsymbol{\Xi})\|\boldsymbol{\gamma}_k - \widehat{\boldsymbol{\gamma}}_k\|_2^2$$
$$\leq (1+\nu_0)[N_{k,2}+N_{k,5}]\mu_1(\boldsymbol{\Xi})^{1/2}C_0\varpi\|(\boldsymbol{\alpha}_k - \widehat{\boldsymbol{\alpha}}_k)_{J_k}\|_1$$
$$\leq (1+\nu_0)[N_{k,2}+N_{k,5}]\mu_1(\boldsymbol{\Xi})^{1/2}C_0\varpi\sqrt{|J_k|}\|(\widehat{\boldsymbol{\alpha}}_k - \boldsymbol{\alpha}_k)_{J_k}\|_2$$
$$\leq (1+\nu_0)[N_{k,2}+N_{k,5}]\mu_1(\boldsymbol{\Xi})^{1/2}C_0\varpi\sqrt{s}\kappa^{-1}\|\widehat{\boldsymbol{\gamma}}_k - \boldsymbol{\gamma}_k\|_2$$

Hence, by the inequality above and $\mu_k(\boldsymbol{\Xi}) \geq c_3\mu_1(\boldsymbol{\Xi})$ (Condition 3), we have $\|\widehat{\boldsymbol{\gamma}}_1 - \boldsymbol{\gamma}_1\|_2 \leq 2c_2^{-1}c_3(1+\nu_0)[N_{k,2}+N_{k,5}]\kappa^{-1}\mu_1(\boldsymbol{\Xi})^{-1/2}C_0\varpi\sqrt{s}$. Because

$$\lim_{\substack{A^{(k)}\to\infty \\ \hbar\to\infty}} \frac{2c_2^{-1}c_3(1+\nu_0)[N_{k,2}+N_{k,5}]}{A^{(k)}}$$
$$= 2c_2^{-1}c_3(1+\nu_0)(2\|\boldsymbol{\alpha}_k\|_1/\|\boldsymbol{\alpha}_1\|_1)$$
$$< 2c_2^{-1}c_3(1+2\nu_0)(2\|\boldsymbol{\alpha}_k\|_1/\|\boldsymbol{\alpha}_1\|_1) \leq 4c_2^{-1}c_3(1+2\nu_0)c_4,$$

where the last inequality is due to the inequality (1.14). Therefore, as $A^{(k)}$ and

$\hbar$ are large enough, we have

$$\|\widehat{\boldsymbol{\gamma}}_k - \boldsymbol{\gamma}_k\|_2 \le 4c_2^{-1}c_3(1+2\nu_0)c_4A^{(k)}\kappa^{-1}\mu_1(\boldsymbol{\Xi})^{-1/2}C_0\varpi\sqrt{s}, \tag{1.39}$$

$$\text{and hence} \quad \|\widehat{\boldsymbol{\gamma}}_k - \boldsymbol{\gamma}_k\|_2^2 \le 16c_2^{-2}c_3^2(1+2\nu_0)^2c_4^2(A^{(k)})^2\kappa^{-2}\mu_1(\boldsymbol{\Xi})^{-1}C_0^2\varpi^2 s.$$

Now by (1.37), (1.38) and (1.39),

$$\begin{aligned}
&\|\widehat{\boldsymbol{\alpha}}_k - \boldsymbol{\alpha}_k\|_1 = \|(\widehat{\boldsymbol{\alpha}}_k)_{J_k^c}\|_1 + \|(\widehat{\boldsymbol{\alpha}}_k - \boldsymbol{\alpha}_k)_{J_k}\|_1 \\
<&(1+c)\|(\widehat{\boldsymbol{\alpha}}_k - \boldsymbol{\alpha}_k)_{J_k}\|_1 \\
\le&(1+c)\sqrt{s}\|(\widehat{\boldsymbol{\alpha}}_k - \boldsymbol{\alpha}_k)_{J_k}\|_2 \le (1+c)\sqrt{s}\kappa^{-1}\|\widehat{\boldsymbol{\gamma}}_k - \boldsymbol{\gamma}_k\|_2 \\
\le&4(1+c)c_2^{-1}c_3c_4(1+2\nu_0)A^{(k)}\kappa^{-2}\mu_k(\boldsymbol{\Xi})^{-1/2}C_0\varpi s.
\end{aligned} \tag{1.40}$$

- **Case 2:** $N_{k,1}C_0^2\kappa^{-2}\varpi^2 s > \nu_0(N_{k,2}+N_{k,5})\mu_1(\boldsymbol{\Xi})^{1/2}C_0\varpi\|(\boldsymbol{\alpha}_k - \widehat{\boldsymbol{\alpha}}_k)_{J_k}\|_1$.

In this case, the inequality above leads to

$$\|(\boldsymbol{\alpha}_k - \widehat{\boldsymbol{\alpha}}_k)_{J_k}\|_1 \le \frac{N_{k,1}}{\nu_0(N_{k,2}+N_{k,5})}\mu_1(\boldsymbol{\Xi})^{-1/2}\kappa^{-2}C_0\varpi s. \tag{1.41}$$

Because

$$\begin{aligned}
&\lim_{\substack{A^{(k)}\to\infty\\ \hbar\to\infty}} \frac{N_{k,5}}{N_{k,2}} = 0, \\
&\lim_{\substack{A^{(k)}\to\infty\\ \hbar\to\infty}} \frac{N_{k,1}}{\nu_0(N_{k,2}+N_{k,5})} = \nu_0^{-1}M_{k,1}M_{k,0}(2\|\boldsymbol{\alpha}_k\|_1/\|\boldsymbol{\alpha}_1\|_1)^{-1} \\
&< \nu_0^{-1}M_{k,1}M_{k,0}c_4,
\end{aligned}$$

we have, as $A^{(k)}$ and $\hbar$ are large enough,

$$\begin{aligned}
&\lambda^{(k)}N_{k,2} - N_{k,5} > 0, \\
&\|(\boldsymbol{\alpha}_k - \widehat{\boldsymbol{\alpha}}_k)_{J_k}\|_1 \le \nu_0^{-1}M_{k,1}M_{k,0}c_4\mu_1(\boldsymbol{\Xi})^{-1/2}\kappa^{-2}C_0\varpi s,
\end{aligned} \tag{1.42}$$

In this case, (1.33) gives

$$\begin{aligned}
&\frac{1}{2}c_2\mu_k(\boldsymbol{\Xi})\|\boldsymbol{\gamma}_k - \widehat{\boldsymbol{\gamma}}_k\|_2^2 + (\lambda^{(k)}N_{k,2} - N_{k,5})\mu_1(\boldsymbol{\Xi})^{1/2}C_0\varpi\|(\widehat{\boldsymbol{\alpha}}_k)_{J_k^c}\|_1 \\
\le&(1+\nu_0^{-1})N_{k,1}C_0^2\kappa^{-2}\varpi^2 s,
\end{aligned} \tag{1.43}$$

which together with the first inequality in (1.42) lead to

$$\begin{aligned}
&\|\boldsymbol{\gamma}_k - \widehat{\boldsymbol{\gamma}}_k\|_2^2 \le 2c_2^{-1}\mu_k(\boldsymbol{\Xi})^{-1}(1+\nu_0^{-1})N_{k,1}C_0^2\kappa^{-2}\varpi^2 s \\
\le & 2c_2^{-1}c_3\mu_1(\boldsymbol{\Xi})^{-1}(1+\nu_0^{-1})N_{k,1}C_0^2\kappa^{-2}\varpi^2 s,
\end{aligned} \tag{1.44}$$

Because

$$\lim_{\substack{A^{(k)}\to\infty \\ \hbar\to\infty}} \frac{N_{k,1}}{A^{(k)}} = 2M_{k,1}M_{k,0} < 4M_{k,1}M_{k,0},$$

as $A^{(k)}$ and $\hbar$ are large enough, we have

$$\begin{aligned}
&\|\boldsymbol{\gamma}_k - \widehat{\boldsymbol{\gamma}}_k\|_2^2 \le 8c_2^{-1}c_3(1+\nu_0^{-1})M_{k,1}M_{k,0}A^{(k)}\mu_1(\boldsymbol{\Xi})^{-1}C_0^2\kappa^{-2}\varpi^2 s \\
\le & 8c_2^{-1}c_3(1+\nu_0^{-1})M_{k,1}M_{k,0}(A^{(k)})^2\mu_1(\boldsymbol{\Xi})^{-1}C_0^2\kappa^{-2}\varpi^2 s,
\end{aligned} \tag{1.45}$$

By (1.43),

$$\begin{aligned}
&(\lambda^{(k)}N_{k,2} - N_{k,5})\mu_1(\boldsymbol{\Xi})^{1/2}C_0\varpi\|(\widehat{\boldsymbol{\alpha}}_k)_{J_k^c}\|_1 \\
\le & (1+\nu_0^{-1})N_{k,1}C_0^2\kappa^{-2}\varpi^2 s.
\end{aligned} \tag{1.46}$$

Because

$$\begin{aligned}
\lim_{\substack{A^{(k)}\to\infty \\ \hbar\to\infty}} \frac{(1+\nu_0^{-1})N_{k,1}}{(\lambda^{(k)}N_{k,2} - N_{k,5})} &= (1+\nu_0^{-1})(\lambda^{(k)})^{-1}M_{k,1}M_{k,0}(2\|\boldsymbol{\alpha}_k\|_1/\|\boldsymbol{\alpha}_1\|_1)^{-1} \\
&< (1+\nu_0^{-1})cM_{k,1}M_{k,0}c_4,
\end{aligned} \tag{1.47}$$

as $A^{(k)}$ and $\hbar$ are large enough, we have

$$\|(\widehat{\boldsymbol{\alpha}}_k)_{J_k^c}\|_1 \le (1+\nu_0^{-1})cM_{k,1}M_{k,0}c_4\mu_1(\boldsymbol{\Xi})^{-1/2}\kappa^{-2}C_0\varpi s$$

which together with (1.42) lead to

$$\begin{aligned}
&\|\widehat{\boldsymbol{\alpha}}_k - \boldsymbol{\alpha}_k\|_1 = \|(\widehat{\boldsymbol{\alpha}}_k - \boldsymbol{\alpha}_k)_{J_k}\|_1 + \|(\widehat{\boldsymbol{\alpha}}_k)_{J_k^c}\|_1 \\
< & (1+2\nu_0^{-1})cM_{k,1}M_{k,0}c_4\mu_1(\boldsymbol{\Xi})^{-1/2}\kappa^{-2}C_0\varpi s \\
\le & (1+2\nu_0^{-1})cM_{k,1}M_{k,0}c_4A^{(k)}\mu_1(\boldsymbol{\Xi})^{-1/2}\kappa^{-2}C_0\varpi s.
\end{aligned} \tag{1.48}$$

Now we combine the results for the two cases. By (1.39) and (1.45), in both the two cases, we have

$$\|\boldsymbol{\gamma}_k - \widehat{\boldsymbol{\gamma}}_k\|_2^2 \le D_{k,2}(A^{(k)})^2\kappa^{-2}\mu_1(\boldsymbol{\Xi})^{-1}C_0^2\varpi^2 s, \tag{1.49}$$

where $D_{k,2} = 16c_2^{-2}c_3^2(1+2\nu_0)^2c_4^2 + 8c_2^{-1}c_3(1+\nu_0^{-1})M_{k,1}M_{k,0}$ which only depends on δ_0, c and $c_2 \sim c_5$. Now by (1.49) and (1.22), as $A^{(k)}$ and $\hbar$ are large enough,

$$\begin{aligned}
&\|\widehat{\mathbf{P}}_k - \mathbf{P}_{k-1}\|^2 = \|(\widehat{\mathbf{P}}_{k-1} + \widehat{\boldsymbol{\gamma}}_k^{\mathrm{T}}\widehat{\boldsymbol{\gamma}}_k) - (\mathbf{P}_{k-1} - \boldsymbol{\gamma}_k^{\mathrm{T}}\boldsymbol{\gamma}_k)\|^2 \qquad (1.50)\\
\le& 2\|\widehat{\mathbf{P}}_{k-1} - \mathbf{P}_{k-1}\|^2 + 2\|\widehat{\boldsymbol{\gamma}}_k^{\mathrm{T}}\widehat{\boldsymbol{\gamma}}_k - \boldsymbol{\gamma}_k^{\mathrm{T}}\boldsymbol{\gamma}_k\|^2\\
\le& 2\|\widehat{\mathbf{P}}_{k-1} - \mathbf{P}_{k-1}\|^2 + 4\|\widehat{\boldsymbol{\gamma}}_k\|_2^2\|\widehat{\boldsymbol{\gamma}}_k - \boldsymbol{\gamma}_k\|_2^2 + 4\|\widehat{\boldsymbol{\gamma}}_k - \boldsymbol{\gamma}_k\|_2^2\|\boldsymbol{\gamma}_k\|_2^2\\
=& 2\|\widehat{\mathbf{P}}_{k-1} - \mathbf{P}_{k-1}\|^2 + 4\|\widehat{\boldsymbol{\gamma}}_k - \boldsymbol{\gamma}_k\|_2^2 + 4\|\widehat{\boldsymbol{\gamma}}_k - \boldsymbol{\gamma}_k\|_2^2\\
\le& 2b_{k-1,3}\kappa^{-2}\mu_1(\boldsymbol{\Xi})^{-1}C_0^2\varpi^2 s + 8\|\widehat{\boldsymbol{\gamma}}_k - \boldsymbol{\gamma}_k\|_2^2\\
\le& 2b_{k-1,3}(A^{(k)})^2\kappa^{-2}\mu_1(\boldsymbol{\Xi})^{-1}C_0^2\varpi^2 s + 8D_{k,2}(A^{(k)})^2\kappa^{-2}\mu_1(\boldsymbol{\Xi})^{-1}C_0^2\varpi^2 s\\
=& D_{k,3}(A^{(k)})^2\kappa^{-2}\mu_1(\boldsymbol{\Xi})^{-1}C_0^2\varpi^2 s,
\end{aligned}$$

where $D_{k,3} = 2b_{k-1,3} + 8D_{k,2}$. Finally, by (1.40) and (1.48), in both the two cases, we have

$$\begin{aligned}
\|\widehat{\boldsymbol{\alpha}}_k - \boldsymbol{\alpha}_k\|_1 \le& 4(1+c)c_2^{-1}c_3c_4(1+2\nu_0)A^{(k)}\kappa^{-2}\mu_k(\boldsymbol{\Xi})^{-1/2}C_0\varpi s\\
&+ (1+2\nu_0^{-1})cM_{k,1}M_{k,0}c_4A^{(k)}\mu_1(\boldsymbol{\Xi})^{-1/2}\kappa^{-2}C_0\varpi s\\
=& D_{k,4}A^{(k)}\mu_1(\boldsymbol{\Xi})^{-1/2}\kappa^{-2}C_0\varpi s.
\end{aligned}$$

where $D_{k,4} = 4(1+c)c_2^{-1}c_3c_4(1+2\nu_0) + (1+2\nu_0^{-1})cM_{k,1}M_{k,0}c_4$.

Hence, we have proved the claims (1.20) for $i = k$. By induction, the claims are true for all $1 \le i \le K$.

1.2. Proof of Theorem 4.3

In this proof, we only consider the elements in the event Ω. Since $\widehat{w}_k(t)$ is the solution to (3.15), we have

$$\|\widehat{w}_k(t) - \widehat{w}_k^0(t)\|_{L^2}^2 + \eta\|\widehat{w}_k''(t)\|_{L^2}^2 \le \|w_k(t) - \widehat{w}_k^0(t)\|_{L^2}^2 + \eta\|w_k''(t)\|_{L^2}^2. \qquad (1.51)$$

Then by (1.51)

$$
\begin{aligned}
&\|\widehat{w}_k(t)-w_k(t)\|_{L^2}^2 \le 2\|\widehat{w}_k(t)-\widehat{w}_k^0(t)\|_{L^2}^2+2\|w_k(t)-\widehat{w}_k^0(t)\|_{L^2}^2\\
\le& 2\|w_k(t)-\widehat{w}_k^0(t)\|_{L^2}^2+2\eta\|w_k''(t)\|_{L^2}^2+2\|w_k(t)-\widehat{w}_k^0(t)\|_{L^2}^2\\
=& 4\|w_k(t)-\widehat{w}_k^0(t)\|_{L^2}^2+2\eta\|w_k''(t)\|_{L^2}^2 \qquad (1.52)
\end{aligned}
$$

We first estimate $\|w_k(t)-\widehat{w}_k^0(t)\|_{L^2}^2$. Because for each k, $\widehat{\mathbf{t}}_k$ has sample mean zero,

$$
\begin{aligned}
\widehat{w}_k^0(t) &= \frac{1}{n}\widehat{\mathbf{t}}_k^{\mathrm{T}}\mathbf{Y}(t)=\frac{1}{n}\widehat{\mathbf{t}}_k^{\mathrm{T}}\left[\mathbf{Y}(t)-\bar{y}(t)\mathbf{1}_n\right]=\frac{1}{n}\widehat{\mathbf{t}}_k^{\mathrm{T}}\left\{\mathbf{X}\boldsymbol{\beta}(t)+\boldsymbol{\varepsilon}(t)-\bar{\epsilon}(t)\mathbf{1}_n\right\}\\
&= \frac{1}{n}\widehat{\mathbf{t}}_k^{\mathrm{T}}\left\{\mathbf{t}_1w_1(t)+\mathbf{t}_2w_2(t)+\cdots+\mathbf{t}_Kw_K(t)+\boldsymbol{\varepsilon}(t)-\bar{\epsilon}(t)\mathbf{1}_n\right\}\\
&= \frac{1}{n}\sqrt{n}\widehat{\boldsymbol{\gamma}}_k^{\mathrm{T}}\left\{\sqrt{n}\boldsymbol{\gamma}_1w_1(t)+\sqrt{n}\boldsymbol{\gamma}_2w_2(t)+\cdots+\sqrt{n}\boldsymbol{\gamma}_Kw_K(t)+\sqrt{n}\boldsymbol{\varrho}(t)\right\}\\
&= \widehat{\boldsymbol{\gamma}}_k^{\mathrm{T}}\boldsymbol{\gamma}_kw_k(t)+\sum_{i\neq k}\widehat{\boldsymbol{\gamma}}_k^{\mathrm{T}}\boldsymbol{\gamma}_iw_i(t)+\widehat{\boldsymbol{\gamma}}_k^{\mathrm{T}}\boldsymbol{\varrho}(t)\\
&= \widehat{\boldsymbol{\gamma}}_k^{\mathrm{T}}\boldsymbol{\gamma}_kw_k(t)+\sum_{i\neq k}\widehat{\boldsymbol{\gamma}}_k^{\mathrm{T}}\boldsymbol{\gamma}_iw_i(t)+\widehat{\boldsymbol{\alpha}}_k^{\mathrm{T}}(\mathbf{Z}^{\mathrm{T}}\boldsymbol{\varrho}(t)), \qquad (1.53)
\end{aligned}
$$

where in the last equality is because $\widehat{\boldsymbol{\gamma}}_k=\mathbf{Z}\widehat{\boldsymbol{\alpha}}_k$. Note that $1-\widehat{\boldsymbol{\gamma}}_k^{\mathrm{T}}\boldsymbol{\gamma}_k=\|\widehat{\boldsymbol{\gamma}}_k-\boldsymbol{\gamma}_k\|_2^2/2$ and $w_1(t),\cdots,w_K(t)$ are orthogonal to each other. The middle term in the right hand side of the last equality in (1.53) satisfies

$$
\begin{aligned}
&\|\sum_{i\neq k}\widehat{\boldsymbol{\gamma}}_k^{\mathrm{T}}\boldsymbol{\gamma}_iw_i(t)\|_{L^2}^2=\sum_{i\neq k}(\widehat{\boldsymbol{\gamma}}_k^{\mathrm{T}}\boldsymbol{\gamma}_i)^2\|w_i(t)\|_{L^2}^2=\sum_{i\neq k}(\widehat{\boldsymbol{\gamma}}_k^{\mathrm{T}}\boldsymbol{\gamma}_i)^2\mu_i(\boldsymbol{\Xi})\\
&\le \mu_1(\boldsymbol{\Xi})\sum_{i\neq k}(\widehat{\boldsymbol{\gamma}}_k^{\mathrm{T}}\boldsymbol{\gamma}_i)^2\le\mu_1(\boldsymbol{\Xi})[1-(\widehat{\boldsymbol{\gamma}}_k^{\mathrm{T}}\boldsymbol{\gamma}_k)^2]\\
&\le 2\mu_1(\boldsymbol{\Xi})[1-\widehat{\boldsymbol{\gamma}}_k^{\mathrm{T}}\boldsymbol{\gamma}_k]=\mu_1(\boldsymbol{\Xi})\|\widehat{\boldsymbol{\gamma}}_k-\boldsymbol{\gamma}_k\|_2^2, \qquad (1.54)
\end{aligned}
$$

where the first inequality in the last line is because $|1+\widehat{\boldsymbol{\gamma}}_k^{\mathrm{T}}\boldsymbol{\gamma}_k|\le 2$. Then by (1.53) and (1.54), we have

$$
\begin{aligned}
\|w_k(t)-\widehat{w}_k^0(t)\|_{L^2} &\le (1-\widehat{\boldsymbol{\gamma}}_k^{\mathrm{T}}\boldsymbol{\gamma}_k)\|w_k(t)\|_{L^2}+\|\sum_{i\neq k}\widehat{\boldsymbol{\gamma}}_k^{\mathrm{T}}\boldsymbol{\gamma}_iw_i(t)\|_{L^2}+\|\widehat{\boldsymbol{\alpha}}_k^{\mathrm{T}}(\mathbf{Z}^{\mathrm{T}}\boldsymbol{\varrho}(t))\|_{L^2}\\
&\le (1-\widehat{\boldsymbol{\gamma}}_k^{\mathrm{T}}\boldsymbol{\gamma}_k)\|w_k(t)\|_{L^2}+\|\sum_{i\neq k}\widehat{\boldsymbol{\gamma}}_k^{\mathrm{T}}\boldsymbol{\gamma}_iw_i(t)\|_{L^2}+\max_{1\le j\le p}\|(\mathbf{Z}^{\mathrm{T}}\boldsymbol{\varrho}(t))_j\|_{L^2}\|\widehat{\boldsymbol{\alpha}}_k\|_1\\
&\le \frac{1}{2}\mu_1(\boldsymbol{\Xi})^{1/2}\|\widehat{\boldsymbol{\gamma}}_k-\boldsymbol{\gamma}_k\|_2^2+\mu_1(\boldsymbol{\Xi})^{1/2}\|\widehat{\boldsymbol{\gamma}}_k-\boldsymbol{\gamma}_k\|_2+C_0\varpi\|\widehat{\boldsymbol{\alpha}}_k\|_1, \qquad (1.55)
\end{aligned}
$$

where in the last inequality, we use the inequality $\max_{1\le j\le p}\|(\mathbf{Z}^{\mathrm{T}}\boldsymbol{\varrho}(t))_j\|_{L^2}\le C_0\varpi$ in the event Ω (see the definition (1.4)). By (1.55) and Theorem 4.2 (b),

$$\begin{aligned}\|w_k(t)-\widehat{w}_k^0(t)\|_{L^2}\le&\frac{1}{2}\mu_1(\boldsymbol{\Xi})^{-1/2}D_{k,2}(A^{(k)})^2C_0^2\kappa^{-2}\varpi^2s+D_{k,2}^{1/2}(A^{(k)})C_0\kappa^{-1}\varpi\sqrt{s}\\&+C_0\varpi D_{k,1}\sqrt{s}\kappa^{-1}\\\le&\frac{1}{2}D_{k,2}(A^{(k)})^2\hbar^{-1}C_0\kappa^{-1}\varpi\sqrt{s}+D_{k,2}^{1/2}(A^{(k)})C_0\kappa^{-1}\varpi\sqrt{s}\\&+C_0\varpi D_{k,1}\sqrt{s}\kappa^{-1}\\=&\widetilde{L}_{k,1}C_0\varpi\sqrt{s}\kappa^{-1},\end{aligned}\tag{1.56}$$

where $\widetilde{L}_{k,1}=\frac{1}{2}D_{k,2}(A^{(k)})^2\hbar^{-1}+D_{k,2}^{1/2}(A^{(k)})+D_{k,1}$ and the second inequality is due to the condition $\mu_1(\boldsymbol{\Xi})\ge\hbar^2C_0^2\varpi^2s/\kappa^2$. Now we estimate $2\eta\|w_k''(t)\|_{L^2}^2$ in (1.52). It follows from the equalities: $\boldsymbol{\alpha}_k^{\mathrm{T}}\mathbf{S}\boldsymbol{\alpha}_k=1$ and $\boldsymbol{\alpha}_k^{\mathrm{T}}\mathbf{S}\boldsymbol{\alpha}_j=0$ if $j\ne k$, that

$$\boldsymbol{\alpha}_k^{\mathrm{T}}\mathbf{S}\boldsymbol{\beta}(t)=\boldsymbol{\alpha}_k^{\mathrm{T}}\mathbf{S}\left\{\boldsymbol{\alpha}_1w_1(t)+\cdots+\boldsymbol{\alpha}_Kw_K(t)\right\}=w_k(t),\tag{1.57}$$

which leads to

$$\begin{aligned}|w_k''(t)|=|\boldsymbol{\alpha}_k^{\mathrm{T}}\mathbf{S}\boldsymbol{\beta}''(t)|\le\|\mathbf{S}_{ij}\|_\infty\|\boldsymbol{\alpha}_k\|_1\sum_{j=1}^p|\boldsymbol{\beta}''(t)|\\\le\|\boldsymbol{\alpha}_k\|_1\sum_{j=1}^p|\boldsymbol{\beta}''(t)|\le\kappa^{-1}\sqrt{s}\sum_{j=1}^p|\boldsymbol{\beta}''(t)|\le\kappa^{-1}\sqrt{s}(\max_{1\le j\le p}|\beta_j''(t)|)s\le\kappa^{-1}C_\beta s^{3/2},\end{aligned}\tag{1.58}$$

where the first inequality in the second line is because $\|\mathbf{S}\|_\infty=1$ by Condition 3, the second inequality follows Lemma 2, the third inequality is because there are only s nonzero functions among $\{\boldsymbol{\beta}_1,\cdots,\boldsymbol{\beta}_p\}$ and the last one is due to the uniform smoothness property (**??**) with

$$C_\beta=\max_{1\le q\le Q}\int_0^1\int_0^1\left(\frac{\partial^2\beta_q(t,s)}{\partial t^2}\right)^2dsdt.$$

Now by (1.58) and the inequality $\eta\le C_0C_\beta^{-1}s^{-1}$, we have

$$2\eta\|w_k''(t)\|_{L^2}^2=2\eta\int_0^1|w_k''(t)|^2dt\le2\eta\kappa^{-1}C_\beta s^{3/2}\le2C_0\varpi\sqrt{s}\kappa^{-1},$$

which, together with (1.52) and (1.56), implies

$$\begin{aligned}\|\widehat{w}_k(t)-w_k(t)\|_{L^2}^2\le4\|w_k(t)-\widehat{w}_k^0(t)\|_{L^2}^2+2\eta\|w_k''(t)\|_{L^2}^2\\\le4\widetilde{L}_{k,1}C_0\varpi\sqrt{s}\kappa^{-1}+2C_0\varpi\sqrt{s}\kappa^{-1}\le L_{k,1}C_0\varpi\sqrt{s}\kappa^{-1},\end{aligned}\tag{1.59}$$

where $L_{k,1} = 4\widetilde{L}_{k,1} + 2$. Now by (**??**), (1.59) and Theorem 4.2(b),

$$
\begin{aligned}
&\|\widehat{\boldsymbol{\beta}}_{K_0}(t) - \boldsymbol{\beta}_{K_0}(t)\|_{1,2}^2 \\
\leq &\|\sum_{k=1}^{K_0} \widehat{\boldsymbol{\alpha}}_k \widehat{w}_k(t) - \sum_{k=1}^{K_0} \boldsymbol{\alpha}_k w_k(t)\|_{1,2} \leq \sum_{k=1}^{K_0} \|\widehat{\boldsymbol{\alpha}}_k \widehat{w}_k(t) - \boldsymbol{\alpha}_k w_k(t)\|_{1,2} \\
\leq &\sum_{k=1}^{K_0} \|\widehat{\boldsymbol{\alpha}}_k \left[\widehat{w}_k(t) - w_k(t)\right] \|_{1,2} + \sum_{k=1}^{K_0} \|(\widehat{\boldsymbol{\alpha}}_k - \boldsymbol{\alpha}_k) w_k(t)\|_{1,2} \\
\leq &\sum_{k=1}^{K_0} \|\widehat{\boldsymbol{\alpha}}_k\|_1 \|\widehat{w}_k(t) - w_k(t)\|_{L^2} + \sum_{k=1}^{K_0} \|\widehat{\boldsymbol{\alpha}}_k - \boldsymbol{\alpha}_k\|_1 \|w_k(t)\|_{L^2} \\
\leq &\sum_{k=1}^{K_0} D_{k,1} L_{k,1} C_0 \kappa^{-2} \varpi s + \sum_{k=1}^{K_0} D_{k,4}(A^{(k)}) C_0 \kappa^{-2} \mu_1(\boldsymbol{\Xi})^{-1/2} \varpi s \mu_1(\boldsymbol{\Xi})^{1/2} \\
= &L_{K_0,3} C_0 \kappa^{-2} \varpi s,
\end{aligned}
$$

where $L_{K_0,3} = \sum_{k=1}^{K_0} D_{k,1} L_{k,1} + \sum_{k=1}^{K_0} D_{k,4}(A^{(k)})$. In particular, when $K_0 = K$, we have

$$
\|\widehat{\boldsymbol{\beta}}(t) - \boldsymbol{\beta}(t)\|_{1,2} \leq L_{K,3} C_0 \kappa^{-2} \varpi s,
$$

Finally, by (1.59),

$$
\begin{aligned}
&\|\mathbf{X}\widehat{\boldsymbol{\beta}}_{K_0}(t) - \mathbf{X}\boldsymbol{\beta}_{K_0}(t)\|_{L^2}^2 \\
= &\|\sum_{k=1}^{K_0} \mathbf{X}\widehat{\boldsymbol{\alpha}}_k \widehat{w}_k(t) - \sum_{k=1}^{K_0} \mathbf{X}\boldsymbol{\alpha}_k w_k(t)\|_{L^2} \leq \sum_{k=1}^{K_0} \|\mathbf{X}\widehat{\boldsymbol{\alpha}}_k \widehat{w}_k(t) - \mathbf{X}\boldsymbol{\alpha}_k w_k(t)\|_{L^2} \\
\leq &\sum_{k=1}^{K_0} \|\mathbf{X}\widehat{\boldsymbol{\alpha}}_k\|_2 \|\widehat{w}_k(t) - w_k(t)\|_{L^2} + \sum_{k=1}^{K_0} \|\mathbf{X}(\widehat{\boldsymbol{\alpha}}_k - \boldsymbol{\alpha}_k)\|_{L^2} \|w_k(t)\|_{L^2} \\
\leq &\sum_{k=1}^{K_0} \sqrt{n} L_{k,1} C_0 \kappa^{-1} \varpi \sqrt{s} + \sum_{k=1}^{K_0} \sqrt{n} D_{k,2}^{1/2}(A^{(k)}) C_0 \kappa^{-1} \mu_1(\boldsymbol{\Xi})^{-1/2} \varpi \sqrt{s} \mu_1(\boldsymbol{\Xi})^{1/2} \\
= &\sqrt{n} L_{K_0,4} C_0 \kappa^{-1} \varpi \sqrt{s},
\end{aligned}
$$

where $L_{K_0,4} = \sum_{k=1}^{K_0} L_{k,1} + \sum_{k=1}^{K_0} D_{k,2}^{1/2}(A^{(k)})$ and the fourth line is due to $\|\mathbf{X}\widehat{\boldsymbol{\alpha}}_k\|_2^2 = \widehat{\boldsymbol{\alpha}}_k^{\mathrm{T}} \mathbf{X}^{\mathrm{T}} \mathbf{X} \widehat{\boldsymbol{\alpha}}_k = n \widehat{\boldsymbol{\alpha}}_k^{\mathrm{T}} \mathbf{S} \widehat{\boldsymbol{\alpha}}_k = 1$. In particular, when $K_0 = K$, we have

$$
\|\mathbf{X}\widehat{\boldsymbol{\beta}}(t) - \mathbf{X}\boldsymbol{\beta}(t)\|_{L^2} \leq \sqrt{n} L_{K,4} C_0 \kappa^{-1} \varpi \sqrt{s}.
$$

2. Appendix D: Proofs of technical lemmas

In Appendix D, we provide the proofs for all technical lemmas.

Proof of Lemma 1.

For any $1 \le j \le p$,

$$(\mathbf{Z}^{\mathrm{T}}\boldsymbol{\varrho}(t))_j = \frac{1}{\sqrt{n}}(\mathbf{X}^{\mathrm{T}}\boldsymbol{\varrho}(t))_j = \frac{1}{n}\sum_{i=1}^{n}\mathbf{X}_{ij}(\varepsilon_i(t) - \bar{\varepsilon}(t))$$
$$= \frac{1}{n}\sum_{i=1}^{n}\mathbf{X}_{ij}\varepsilon_i(t) - \bar{\mathbf{X}}_{\cdot j}\bar{\varepsilon}(t) = \frac{1}{n}\sum_{i=1}^{n}\mathbf{X}_{ij}\varepsilon_i(t)$$

where $\mathbf{X}_{ij}$ is the (i,j)-th entry of $\mathbf{X}$ and $\bar{\mathbf{X}}_{\cdot j}$ is the mean of the j-th column of $\mathbf{X}$ which is equal to zero. Therefore, because $\varepsilon_1(t)(t), \cdots, \varepsilon_n(t)$ are i.i.d. Gaussian random functions by Condition 2, by the equalities above, $(\mathbf{Z}^{\mathrm{T}}\boldsymbol{\varrho}(t))_j$ has the same distribution as $\varepsilon_1(t)\sqrt{\sum_{i=1}^{n}\mathbf{X}_{ij}^2/n^2} = \varepsilon_1(t)\sqrt{\mathbf{S}_{jj}/n} = \varepsilon_1(t)/\sqrt{n}$, where $\mathbf{S}_{jj}$ is the j-th diagonal element of $\mathbf{S}$ which is equal to 1 by Condition 3. By the inequality in Lemma 3.1 in Section 3.1 of Ledoux and Talagrand [1] for the tail probability of Gaussian variables, we have for any $x > M_\epsilon$,

$$P\left(\|(\mathbf{Z}^{\mathrm{T}}\boldsymbol{\varrho}(t))_j\|_{L^2} > \frac{x}{\sqrt{n}}\right) = P\left(\|\frac{\varepsilon_1(t)}{\sqrt{n}}\|_{L^2} > \frac{x}{\sqrt{n}}\right) = P\left(\|\varepsilon_1(t)\|_{L^2} > x\right) \le e^{-\frac{(x-M_\epsilon)^2}{2\sigma^2}}$$
$$= e^{M_\epsilon^2/2\sigma^2}e^{-\frac{(x-M_\epsilon)^2+M_\epsilon^2}{2\sigma^2}} \le e^{M_\epsilon^2/2\sigma^2}e^{-\frac{x^2}{4\sigma^2}}. \tag{2.1}$$

For any $C > M_\epsilon/\sqrt{\log p}$, let $x = C\sqrt{\log p}$. Then $x > M_\epsilon$ and by (2.1), we have

$$P\left(\max_{1\le j\le p}\|(\mathbf{Z}^{\mathrm{T}}\boldsymbol{\varrho}(t))_j\|_{L^2} > C\sqrt{\frac{\log p}{n}}\right)$$
$$= P\left(\max_{1\le j\le p}\|(\mathbf{Z}^{\mathrm{T}}\boldsymbol{\varrho}(t))_j\|_{L^2} > \frac{x}{\sqrt{n}}\right) \le \sum_{j=1}^{p} P\left(\|(\mathbf{Z}^{\mathrm{T}}\boldsymbol{\varrho}(t))_j\|_{L^2} > \frac{x}{\sqrt{n}}\right)$$
$$\le p e^{M_\epsilon^2/2\sigma^2}e^{-x^2/4\sigma^2} = p e^{M_\epsilon^2/2\sigma^2}p^{-C^2/4\sigma^2}. \tag{2.2}$$

On the other hand, for any $1 \le k \le K$,

$$\boldsymbol{\varrho}(t)^{\mathrm{T}}(\mathbf{Z}\boldsymbol{\alpha}_k) = \frac{1}{\sqrt{n}}(\boldsymbol{\varepsilon}(t) - \bar{\varepsilon}(t)\mathbf{1}_n)^{\mathrm{T}}(\mathbf{Z}\boldsymbol{\alpha}_k) = \frac{1}{\sqrt{n}}\boldsymbol{\varepsilon}(t)^{\mathrm{T}}(\mathbf{Z}\boldsymbol{\alpha}_k) - \frac{1}{\sqrt{n}}\bar{\varepsilon}(t)\mathbf{1}_n^{\mathrm{T}}(\mathbf{Z}\boldsymbol{\alpha}_k)$$
$$= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(\mathbf{Z}\boldsymbol{\alpha}_k)_i\varepsilon_i(t),$$

where $(\mathbf{Z}\boldsymbol{\alpha}_k)_i$ is the i-th coordinate of the vector $\mathbf{Z}\boldsymbol{\alpha}_k$ and last equality is because the column means of $\mathbf{Z} = \mathbf{X}/\sqrt{n}$ are all zero. Note that $\sum_{i=1}^{n}(\mathbf{Z}\boldsymbol{\alpha}_k)_i\varepsilon_i(t)/\sqrt{n}$ has the same distribution as $\|(\mathbf{Z}\boldsymbol{\alpha}_k)/\sqrt{n}\|_{L^2}\varepsilon_1(t) = \varepsilon_1(t)/\sqrt{n}$, where we use the equality: $\|\mathbf{Z}\boldsymbol{\alpha}_k\|_{L^2}^2 = \boldsymbol{\alpha}_k^{\mathrm{T}}\mathbf{Z}^{\mathrm{T}}\mathbf{Z}\boldsymbol{\alpha}_k = \boldsymbol{\alpha}_k^{\mathrm{T}}\mathbf{S}\boldsymbol{\alpha}_k = 1$. Similar to (2.1), for $x = C\sqrt{\log p}$, we have

$$\begin{aligned}&P\left(\|(\mathbf{Z}\boldsymbol{\alpha}_k)^{\mathrm{T}}\boldsymbol{\varrho}(t)\|_{L^2} > \frac{x}{\sqrt{n}}\right) = P\left(\|\frac{\varepsilon_1(t)}{\sqrt{n}}\|_{L^2} > \frac{x}{\sqrt{n}}\right)\\ \le& e^{M_\epsilon^2/2\sigma^2}e^{-x^2/4\sigma^2} = e^{M_\epsilon^2/2\sigma^2}p^{-C^2/4\sigma^2}.\end{aligned}$$

Hence

$$P\left(\bigcup_{k=1}^{K}\left\{\|(\mathbf{Z}\boldsymbol{\alpha}_k)^{\mathrm{T}}\boldsymbol{\varrho}(t)\|_{L^2} > C\sqrt{\frac{\log p}{n}}\right\}\right) \le Ke^{M_\epsilon^2/2\sigma^2}p^{-C^2/4\sigma^2}.$$

Proof of Lemma 2.

Given $1 \le k \le K$, let $J_k = J(\boldsymbol{\alpha}_k)$. Then by (4.1), $|J_k| \le \mathcal{M}(\boldsymbol{\beta}(t)) = s$ and $(\boldsymbol{\alpha}_k)_{J_k^c} = \mathbf{0}$. By Condition 1,

$$\kappa \le \frac{\|\mathbf{X}\boldsymbol{\alpha}_k\|_2}{\sqrt{n}\|(\boldsymbol{\alpha}_k)_{J_k}\|_2} = \frac{\|\mathbf{Z}\boldsymbol{\alpha}_k\|_2}{\|(\boldsymbol{\alpha}_k)_{J_k}\|_2} = \frac{\|\mathbf{Z}\boldsymbol{\alpha}_k\|_2}{\|\boldsymbol{\alpha}_k\|_2}.$$

Therefore, $\kappa^2\|\boldsymbol{\alpha}_k\|_2^2 \le \|\mathbf{Z}\boldsymbol{\alpha}_k\|_2^2 = \boldsymbol{\alpha}_k^{\mathrm{T}}\mathbf{Z}^{\mathrm{T}}\mathbf{Z}\boldsymbol{\alpha}_k = \boldsymbol{\alpha}_k^{\mathrm{T}}\mathbf{S}\boldsymbol{\alpha}_k = 1$, which leads to the first inequality in the lemma.

Under Condition 3, all the diagonal elements of $\mathbf{S} = \mathbf{X}^{\mathrm{T}}\mathbf{X}/n$ are equal to 1. Then it is easy to see that the absolute values of all the elements of $\mathbf{S}$ are less than or equal to 1. That is, $\|\mathbf{S}\|_\infty \le 1$. Therefore, we have $1 = \boldsymbol{\alpha}_k^{\mathrm{T}}\mathbf{S}\boldsymbol{\alpha}_k \le \|\mathbf{S}\|_\infty\|\boldsymbol{\alpha}_k\|_1^2 \le \|\boldsymbol{\alpha}_k\|_1^2$ and $1 = \widehat{\boldsymbol{\alpha}}_k^{\mathrm{T}}\mathbf{S}\widehat{\boldsymbol{\alpha}}_k \le \|\mathbf{S}\|_\infty\|\widehat{\boldsymbol{\alpha}}_k\|_1^2 \le \|\widehat{\boldsymbol{\alpha}}_k\|_1^2$.

As for the last inequality, if both Conditions 3 and 1 are satisfied, by the Cauchy-Schwarz inequality, $\|\boldsymbol{\alpha}_k\|_1^2 \le s\|\boldsymbol{\alpha}_k\|_2^2 \le s/\kappa^2$.

Proof of Lemma 3.

In this proof, we only consider the element in Ω. By (1.3),

$$
\begin{aligned}
\boldsymbol{\alpha}_1^{\mathrm{T}}\widehat{\mathbf{B}}\boldsymbol{\alpha}_1 &= \boldsymbol{\alpha}_1^{\mathrm{T}}\mathbf{B}\boldsymbol{\alpha}_1 + 2\boldsymbol{\alpha}_1^{\mathrm{T}}\left[\int_0^1 (\mathbf{S}\boldsymbol{\beta}(t))(\mathbf{Z}^{\mathrm{T}}\boldsymbol{\varrho}(t))^{\mathrm{T}} dt\right]\boldsymbol{\alpha}_1 + \boldsymbol{\alpha}_1^{\mathrm{T}}\left[\int_0^1 (\mathbf{Z}^{\mathrm{T}}\boldsymbol{\varrho}(t))(\mathbf{Z}^{\mathrm{T}}\boldsymbol{\varrho}(t))^{\mathrm{T}} dt\right]\boldsymbol{\alpha}_1 \\
&\geq \mu_1(\boldsymbol{\Xi}) - 2\|\boldsymbol{\alpha}_1^{\mathrm{T}}\mathbf{S}\boldsymbol{\beta}(t)\|_{L^2}\|(\mathbf{Z}^{\mathrm{T}}\boldsymbol{\varrho}(t))^{\mathrm{T}}\boldsymbol{\alpha}_1\|_{L^2} + \|(\mathbf{Z}^{\mathrm{T}}\boldsymbol{\varrho}(t))^{\mathrm{T}}\boldsymbol{\alpha}_1\|_{L^2}^2 \\
&\geq \mu_1(\boldsymbol{\Xi}) - 2\sqrt{\mu_1(\boldsymbol{\Xi})}\|(\mathbf{Z}^{\mathrm{T}}\boldsymbol{\varrho}(t))^{\mathrm{T}}\boldsymbol{\alpha}_1\|_{L^2} \\
&\geq \mu_1(\boldsymbol{\Xi}) - 2\sqrt{\mu_1(\boldsymbol{\Xi})}\left[\max_{1\leq j\leq p}\|(\mathbf{Z}^{\mathrm{T}}\boldsymbol{\varrho}(t))_j\|_{L^2}\right]\|\boldsymbol{\alpha}_1\|_1 \\
&\geq \mu_1(\boldsymbol{\Xi}) - 2\sqrt{\mu_1(\boldsymbol{\Xi})}C_0\varpi\|\boldsymbol{\alpha}_1\|_1,
\end{aligned} \tag{2.3}
$$

where the inequality in the second line is due to the Cauchy-Schwarz inequality and the inequality in the third line is because $\|\boldsymbol{\alpha}_1^{\mathrm{T}}\mathbf{S}\boldsymbol{\beta}(t)\|_{L^2} = \sqrt{\boldsymbol{\alpha}_1^{\mathrm{T}}\mathbf{S}\left(\int_0^1 \boldsymbol{\beta}(t)\boldsymbol{\beta}(t)^{\mathrm{T}}dt\right)\mathbf{S}\boldsymbol{\alpha}_1} = \sqrt{\boldsymbol{\alpha}_1^{\mathrm{T}}\mathbf{B}\boldsymbol{\alpha}_1} = \sqrt{\mu_1(\boldsymbol{\Xi})}$ by Theorem 3.1(a), and the inequality in the fourth line is because

$$
\begin{aligned}
\|(\mathbf{Z}^{\mathrm{T}}\boldsymbol{\varrho}(t))^{\mathrm{T}}\boldsymbol{\alpha}_1\|_{L^2} &= \|\sum_{j=1}^p (\mathbf{Z}^{\mathrm{T}}\boldsymbol{\varrho}(t))_j(\boldsymbol{\alpha}_1)_j\|_{L^2} \leq \sum_{j=1}^p \|(\mathbf{Z}^{\mathrm{T}}\boldsymbol{\varrho}(t))_j\|_{L^2}|(\boldsymbol{\alpha}_1)_j| \\
&\leq \left[\max_{1\leq j\leq p}\|(\mathbf{Z}^{\mathrm{T}}\boldsymbol{\varrho}(t))_j\|_{L^2}\right]\|\boldsymbol{\alpha}_1\|_1 \leq C_0\varpi\|\boldsymbol{\alpha}_1\|_1,
\end{aligned} \tag{2.4}
$$

where $(\boldsymbol{\alpha}_1)_j$ is the j-th coordinate of the vector $\boldsymbol{\alpha}_1$ and the last inequality is due to the definition of Ω. By the similar arguments as in (2.3) and (2.4), we have

$$
\begin{aligned}
\widehat{\boldsymbol{\alpha}}_1^{\mathrm{T}}\widehat{\mathbf{B}}\widehat{\boldsymbol{\alpha}}_1 &= \widehat{\boldsymbol{\alpha}}_1^{\mathrm{T}}\mathbf{B}\widehat{\boldsymbol{\alpha}}_1 + 2\widehat{\boldsymbol{\alpha}}_1^{\mathrm{T}}\left[\int_0^1 (\mathbf{S}\boldsymbol{\beta}(t))(\mathbf{Z}^{\mathrm{T}}\boldsymbol{\varrho}(t))^{\mathrm{T}} dt\right]\widehat{\boldsymbol{\alpha}}_1 + \widehat{\boldsymbol{\alpha}}_1^{\mathrm{T}}\left[\int_0^1 (\mathbf{Z}^{\mathrm{T}}\boldsymbol{\varrho}(t))(\mathbf{Z}^{\mathrm{T}}\boldsymbol{\varrho}(t))^{\mathrm{T}} dt\right]\widehat{\boldsymbol{\alpha}}_1 \\
&\leq \mu_1(\boldsymbol{\Xi}) + 2\|\widehat{\boldsymbol{\alpha}}_1^{\mathrm{T}}\mathbf{S}\boldsymbol{\beta}(t)\|_{L^2}\|(\mathbf{Z}^{\mathrm{T}}\boldsymbol{\varrho}(t))^{\mathrm{T}}\widehat{\boldsymbol{\alpha}}_1\|_{L^2} + \|(\mathbf{Z}^{\mathrm{T}}\boldsymbol{\varrho}(t))^{\mathrm{T}}\widehat{\boldsymbol{\alpha}}_1\|_{L^2}^2 \\
&\leq \mu_1(\boldsymbol{\Xi}) + 2\sqrt{\widehat{\boldsymbol{\alpha}}_1^{\mathrm{T}}\mathbf{B}\widehat{\boldsymbol{\alpha}}_1}C_0\varpi\|\widehat{\boldsymbol{\alpha}}_1\|_1 + C_0^2\varpi^2\|\widehat{\boldsymbol{\alpha}}_1\|_1^2 \\
&\leq \mu_1(\boldsymbol{\Xi}) + 2\sqrt{\mu_1(\boldsymbol{\Xi})}C_0\varpi\|\widehat{\boldsymbol{\alpha}}_1\|_1 + C_0^2\varpi^2\|\widehat{\boldsymbol{\alpha}}_1\|_1^2 \\
&= (\sqrt{\mu_1(\boldsymbol{\Xi})} + C_0\varpi\|\widehat{\boldsymbol{\alpha}}_1\|_1)^2,
\end{aligned} \tag{2.5}
$$

where the inequality in the fourth line is due to (1.2). By (1.1), (2.3) and (2.5), we have

$$\begin{aligned}
&\left(\mu_1(\boldsymbol{\Xi}) - 2\sqrt{\mu_1(\boldsymbol{\Xi})}C_0\varpi\|\boldsymbol{\alpha}_1\|_1\right)(1+\tau^{(1)}\|\widehat{\boldsymbol{\alpha}}_1\|^2_{\lambda^{(1)}}) \qquad (2.6)\\
\le&\boldsymbol{\alpha}_1^{\mathrm{T}}\widehat{\mathbf{B}}\boldsymbol{\alpha}_1(1+\tau^{(1)}\|\widehat{\boldsymbol{\alpha}}_1\|^2_{\lambda^{(1)}}) \le \widehat{\boldsymbol{\alpha}}_1^{\mathrm{T}}\widehat{\mathbf{B}}\widehat{\boldsymbol{\alpha}}_1(1+\tau^{(1)}\|\boldsymbol{\alpha}_1\|^2_{\lambda^{(1)}})\\
\le&\left(\mu_1(\boldsymbol{\Xi}) + 2\sqrt{\mu_1(\boldsymbol{\Xi})}C_0\varpi\|\widehat{\boldsymbol{\alpha}}_1\|_1 + C_0^2\varpi^2\|\widehat{\boldsymbol{\alpha}}_1\|_1^2\right)(1+\tau^{(1)}\|\boldsymbol{\alpha}_1\|^2_{\lambda^{(1)}}).
\end{aligned}$$

By (1.5) and the inequality $\lambda^{(1)}\|\widehat{\boldsymbol{\alpha}}_1\|_1^2 \le \|\widehat{\boldsymbol{\alpha}}_1\|^2_{\lambda^{(1)}} \le \|\widehat{\boldsymbol{\alpha}}_1\|_1^2$, the left hand side of the first inequality in (2.6)

$$\begin{aligned}
&\left(\mu_1(\boldsymbol{\Xi}) - 2\sqrt{\mu_1(\boldsymbol{\Xi})}C_0\varpi\|\boldsymbol{\alpha}_1\|_1\right)(1+\tau^{(1)}\|\widehat{\boldsymbol{\alpha}}_1\|^2_{\lambda^{(1)}}) \qquad (2.7)\\
=&\ \mu_1(\boldsymbol{\Xi}) - 2\sqrt{\mu_1(\boldsymbol{\Xi})}C_0\varpi\|\boldsymbol{\alpha}_1\|_1 + \left(\mu_1(\boldsymbol{\Xi}) - 2\sqrt{\mu_1(\boldsymbol{\Xi})}C_0\varpi\|\boldsymbol{\alpha}_1\|_1\right)\tau^{(1)}\|\widehat{\boldsymbol{\alpha}}_1\|^2_{\lambda^{(1)}}\\
\ge&\ \mu_1(\boldsymbol{\Xi}) - 2\sqrt{\mu_1(\boldsymbol{\Xi})}C_0\varpi\|\boldsymbol{\alpha}_1\|_1 + \left(\mu_1(\boldsymbol{\Xi}) - 2\sqrt{\mu_1(\boldsymbol{\Xi})}\sqrt{\mu_1(\boldsymbol{\Xi})}\hbar^{-1}\right)\tau^{(1)}\|\widehat{\boldsymbol{\alpha}}_1\|^2_{\lambda^{(1)}}\\
=&\ \mu_1(\boldsymbol{\Xi}) - 2\sqrt{\mu_1(\boldsymbol{\Xi})}C_0\varpi\|\boldsymbol{\alpha}_1\|_1 + (1-2\hbar^{-1})\mu_1(\boldsymbol{\Xi})\tau^{(1)}\|\widehat{\boldsymbol{\alpha}}_1\|^2_{\lambda^{(1)}}\\
\ge&\ \mu_1(\boldsymbol{\Xi}) - 2\sqrt{\mu_1(\boldsymbol{\Xi})}C_0\varpi\|\boldsymbol{\alpha}_1\|_1 + \lambda^{(1)}(1-2\hbar^{-1})\mu_1(\boldsymbol{\Xi})\tau^{(1)}\|\widehat{\boldsymbol{\alpha}}_1\|_1^2\\
\ge&\ \mu_1(\boldsymbol{\Xi}) - 2\sqrt{\mu_1(\boldsymbol{\Xi})}C_0\varpi\|\boldsymbol{\alpha}_1\|_1 + c^{-1}(1-2\hbar^{-1})\mu_1(\boldsymbol{\Xi})\tau^{(1)}\|\widehat{\boldsymbol{\alpha}}_1\|_1^2.
\end{aligned}$$

where we assume that $\hbar$ is large enough so that $1 - 2\hbar^{-1} > 0$, and the last inequality is due to $\lambda^{(1)} > c^{-1}$ by the second condition in (4.4). The right hand side of the last inequality in (2.6)

$$\begin{aligned}
&\left(\mu_1(\boldsymbol{\Xi}) + 2\sqrt{\mu_1(\boldsymbol{\Xi})}C_0\varpi\|\widehat{\boldsymbol{\alpha}}_1\|_1 + C_0^2\varpi^2\|\widehat{\boldsymbol{\alpha}}_1\|_1^2\right)(1+\tau^{(1)}\|\boldsymbol{\alpha}_1\|^2_{\lambda^{(1)}})\\
=&\mu_1(\boldsymbol{\Xi}) + \mu_1(\boldsymbol{\Xi})\tau^{(1)}\|\boldsymbol{\alpha}_1\|^2_{\lambda^{(1)}}\\
&+\left(2\sqrt{\mu_1(\boldsymbol{\Xi})}C_0\varpi\|\widehat{\boldsymbol{\alpha}}_1\|_1 + C_0^2\varpi^2\|\widehat{\boldsymbol{\alpha}}_1\|_1^2\right)(1+\tau^{(1)}\|\boldsymbol{\alpha}_1\|^2_{\lambda^{(1)}})\\
\le&\mu_1(\boldsymbol{\Xi}) + \mu_1(\boldsymbol{\Xi})\tau^{(1)}\|\boldsymbol{\alpha}_1\|_1^2\\
&+\left(2\sqrt{\mu_1(\boldsymbol{\Xi})}C_0\varpi\|\widehat{\boldsymbol{\alpha}}_1\|_1 + C_0^2\varpi^2\|\widehat{\boldsymbol{\alpha}}_1\|_1^2\right)(1+A^{(1)}\hbar^{-1}), \qquad (2.8)
\end{aligned}$$

where the last inequality follows from the second inequality in (1.5). Combining (2.6), (2.7) and (2.8), we obtain

$$\begin{aligned}
&-2\sqrt{\mu_1(\boldsymbol{\Xi})}C_0\varpi\|\boldsymbol{\alpha}_1\|_1 + c^{-1}(1-2\hbar^{-1})\mu_1(\boldsymbol{\Xi})\tau^{(1)}\|\widehat{\boldsymbol{\alpha}}_1\|_1^2\\
\le&\mu_1(\boldsymbol{\Xi})\tau^{(1)}\|\boldsymbol{\alpha}_1\|_1^2 + \left(2\sqrt{\mu_1(\boldsymbol{\Xi})}C_0\varpi\|\widehat{\boldsymbol{\alpha}}_1\|_1 + C_0^2\varpi^2\|\widehat{\boldsymbol{\alpha}}_1\|_1^2\right)(1+A^{(1)}\hbar^{-1})
\end{aligned}$$

where we have canceled a $\mu_1(\mathbf{\Xi})$ on each side. Multiplying $\|\boldsymbol{\alpha}_1\|_1$ on both sides of the above inequality and shifting the first term on the left to the right, by the definition of $\tau^{(1)}$, we obtain

$$\begin{aligned}
&c^{-1}(1-2\hbar^{-1})\sqrt{\mu_1(\mathbf{\Xi})}A^{(1)}C_0\varpi\|\widehat{\boldsymbol{\alpha}}_1\|_1^2\\
\leq& 2\sqrt{\mu_1(\mathbf{\Xi})}C_0\varpi\|\boldsymbol{\alpha}_1\|_1^2+\sqrt{\mu_1(\mathbf{\Xi})}A^{(1)}C_0\varpi\|\boldsymbol{\alpha}_1\|_1^2\\
&+\left(2\sqrt{\mu_1(\mathbf{\Xi})}C_0\varpi\|\widehat{\boldsymbol{\alpha}}_1\|_1\|\boldsymbol{\alpha}_1\|_1+C_0^2\varpi^2\|\boldsymbol{\alpha}_1\|_1\|\widehat{\boldsymbol{\alpha}}_1\|_1^2\right)(1+A^{(1)}\hbar^{-1})\\
\leq&\sqrt{\mu_1(\mathbf{\Xi})}C_0(2+A^{(1)})\varpi\|\boldsymbol{\alpha}_1\|_1^2\\
&+\left(2\sqrt{\mu_1(\mathbf{\Xi})}C_0\varpi\frac{\|\widehat{\boldsymbol{\alpha}}_1\|_1^2+\|\boldsymbol{\alpha}_1\|_1^2}{2}+C_0\varpi\sqrt{\mu_1(\mathbf{\Xi})}\hbar^{-1}\|\widehat{\boldsymbol{\alpha}}_1\|_1^2\right)(1+A^{(1)}\hbar^{-1})\\
=&\sqrt{\mu_1(\mathbf{\Xi})}C_0[3+A^{(1)}(1+\hbar^{-1})]\varpi\|\boldsymbol{\alpha}_1\|_1^2\\
&+\sqrt{\mu_1(\mathbf{\Xi})}\varpi(1+\hbar^{-1})\|\widehat{\boldsymbol{\alpha}}_1\|_1^2C_0(1+A^{(1)}\hbar^{-1}),
\end{aligned}$$

where the third inequality follows from the first inequality in (1.5). The above inequality, after a simplification, gives

$$\begin{aligned}
&\left\{[c^{-1}(1-2\hbar^{-1})-(1+\hbar^{-1})\hbar^{-1}]A^{(1)}-(1+\hbar^{-1})\right\}\|\widehat{\boldsymbol{\alpha}}_1\|_1^2\\
\leq&[3+A^{(1)}(1+\hbar^{-1})]\|\boldsymbol{\alpha}_1\|_1^2,
\end{aligned}\tag{2.9}$$

where we have canceled $\sqrt{\mu_1(\mathbf{\Xi})}C_0\varpi$ on both sides. Because

$$\lim_{\substack{A^{(1)}\to\infty\\ \hbar\to\infty}}\frac{3+A^{(1)}(1+\hbar^{-1})}{[c^{-1}(1-2\hbar^{-1})-(1+\hbar^{-1})\hbar^{-1}]A^{(1)}-(1+\hbar^{-1})}=c.$$

Therefore, there exist $(A_1^L)'$ and $(\hbar_0)'$ only depending on c such that for any $A^{(1)}\geq(A_1^L)'$ and $\hbar\geq(\hbar_0)'$, we have

$$\|\widehat{\boldsymbol{\alpha}}_1\|_1\leq D_1\|\boldsymbol{\alpha}_1\|_1,\tag{2.10}$$

where $D_1=\sqrt{6c}$. Therefore, we obtained the first inequality in the lemma. To prove the second one, by (2.5) and (2.10),

$$\begin{aligned}
\widehat{\boldsymbol{\alpha}}_1^{\mathrm{T}}\widehat{\mathbf{B}}\widehat{\boldsymbol{\alpha}}_1&\leq(\sqrt{\mu_1(\mathbf{\Xi})}+C_0\varpi\|\widehat{\boldsymbol{\alpha}}_1\|_1)^2\leq(\sqrt{\mu_1(\mathbf{\Xi})}+C_0\varpi D_1\|\boldsymbol{\alpha}_1\|_1)^2\\
&\leq(\sqrt{\mu_1(\mathbf{\Xi})}+D_1\hbar^{-1}\sqrt{\mu_1(\mathbf{\Xi})})^2=(1+\hbar^{-1}D_1)^2\mu_1(\mathbf{\Xi}),
\end{aligned}$$

where the first inequality in the second line is due to the first inequality in (1.5).

For the last inequality in the lemma, by (1.3),

$$
\begin{aligned}
&\boldsymbol{\alpha}_1^{\mathrm{T}}\widehat{\mathbf{B}}\boldsymbol{\alpha}_1 - \widehat{\boldsymbol{\alpha}}_1^{\mathrm{T}}\widehat{\mathbf{B}}\widehat{\boldsymbol{\alpha}}_1 \qquad (2.11)\\
=&(\boldsymbol{\alpha}_1^{\mathrm{T}}\mathbf{B}\boldsymbol{\alpha}_1 - \widehat{\boldsymbol{\alpha}}_1^{\mathrm{T}}\mathbf{B}\widehat{\boldsymbol{\alpha}}_1) + \boldsymbol{\alpha}_1^{\mathrm{T}}(\widehat{\mathbf{B}}-\mathbf{B})\boldsymbol{\alpha}_1 - \widehat{\boldsymbol{\alpha}}_1^{\mathrm{T}}(\widehat{\mathbf{B}}-\mathbf{B})\widehat{\boldsymbol{\alpha}}_1\\
=&(\boldsymbol{\alpha}_1^{\mathrm{T}}\mathbf{B}\boldsymbol{\alpha}_1 - \widehat{\boldsymbol{\alpha}}_1^{\mathrm{T}}\mathbf{B}\widehat{\boldsymbol{\alpha}}_1)\\
&+2\left(\boldsymbol{\alpha}_1^{\mathrm{T}}\left[\int_0^1(\mathbf{S}\boldsymbol{\beta}(t))(\mathbf{Z}^{\mathrm{T}}\boldsymbol{\varrho}(t))^{\mathrm{T}}dt\right]\boldsymbol{\alpha}_1 - \widehat{\boldsymbol{\alpha}}_1^{\mathrm{T}}\left[\int_0^1(\mathbf{S}\boldsymbol{\beta}(t))(\mathbf{Z}^{\mathrm{T}}\boldsymbol{\varrho}(t))^{\mathrm{T}}dt\right]\widehat{\boldsymbol{\alpha}}_1\right)\\
&+\left(\boldsymbol{\alpha}_1^{\mathrm{T}}\left[\int_0^1(\mathbf{Z}^{\mathrm{T}}\boldsymbol{\varrho}(t))(\mathbf{Z}^{\mathrm{T}}\boldsymbol{\varrho}(t))^{\mathrm{T}}dt\right]\boldsymbol{\alpha}_1 - \widehat{\boldsymbol{\alpha}}_1^{\mathrm{T}}\left[\int_0^1(\mathbf{Z}^{\mathrm{T}}\boldsymbol{\varrho}(t))(\mathbf{Z}^{\mathrm{T}}\boldsymbol{\varrho}(t))^{\mathrm{T}}dt\right]\widehat{\boldsymbol{\alpha}}_1\right)\\
\geq&(\boldsymbol{\alpha}_1^{\mathrm{T}}\mathbf{B}\boldsymbol{\alpha}_1 - \widehat{\boldsymbol{\alpha}}_1^{\mathrm{T}}\mathbf{B}\widehat{\boldsymbol{\alpha}}_1)\\
&-2\left|\boldsymbol{\alpha}_1^{\mathrm{T}}\left[\int_0^1(\mathbf{S}\boldsymbol{\beta}(t))(\mathbf{Z}^{\mathrm{T}}\boldsymbol{\varrho}(t))^{\mathrm{T}}dt\right]\boldsymbol{\alpha}_1 - \widehat{\boldsymbol{\alpha}}_1^{\mathrm{T}}\left[\int_0^1(\mathbf{S}\boldsymbol{\beta}(t))(\mathbf{Z}^{\mathrm{T}}\boldsymbol{\varrho}(t))^{\mathrm{T}}dt\right]\widehat{\boldsymbol{\alpha}}_1\right|\\
&-\left|\boldsymbol{\alpha}_1^{\mathrm{T}}\left[\int_0^1(\mathbf{Z}^{\mathrm{T}}\boldsymbol{\varrho}(t))(\mathbf{Z}^{\mathrm{T}}\boldsymbol{\varrho}(t))^{\mathrm{T}}dt\right]\boldsymbol{\alpha}_1 - \widehat{\boldsymbol{\alpha}}_1^{\mathrm{T}}\left[\int_0^1(\mathbf{Z}^{\mathrm{T}}\boldsymbol{\varrho}(t))(\mathbf{Z}^{\mathrm{T}}\boldsymbol{\varrho}(t))^{\mathrm{T}}dt\right]\widehat{\boldsymbol{\alpha}}_1\right|.
\end{aligned}
$$

We will estimate the three terms on the right hand side of the last equality above. To estimate $(\boldsymbol{\alpha}_1^{\mathrm{T}}\mathbf{B}\boldsymbol{\alpha}_1 - \widehat{\boldsymbol{\alpha}}_1^{\mathrm{T}}\mathbf{B}\widehat{\boldsymbol{\alpha}}_1)$, note that $\boldsymbol{\alpha}_1^{\mathrm{T}}\mathbf{B}\boldsymbol{\alpha}_1 = \boldsymbol{\alpha}_1^{\mathrm{T}}\mathbf{Z}^{\mathrm{T}}\boldsymbol{\Xi}\mathbf{Z}\boldsymbol{\alpha}_1 = \boldsymbol{\gamma}_1^{\mathrm{T}}\boldsymbol{\Xi}\boldsymbol{\gamma}_1$, $\widehat{\boldsymbol{\alpha}}_1^{\mathrm{T}}\mathbf{B}\widehat{\boldsymbol{\alpha}}_1 = \widehat{\boldsymbol{\gamma}}_1^{\mathrm{T}}\boldsymbol{\Xi}\widehat{\boldsymbol{\gamma}}_1$, $\|\boldsymbol{\gamma}_1\|_2 = \|\widehat{\boldsymbol{\gamma}}_1\|_2 = 1$ and the following eigen-decomposition

$$
\boldsymbol{\Xi} = \sum_{k=1}^{K}\mu_k(\boldsymbol{\Xi})\boldsymbol{\gamma}_k\boldsymbol{\gamma}_k^{\mathrm{T}}. \qquad (2.12)
$$

We have

$$
\begin{aligned}
&\boldsymbol{\alpha}_1^{\mathrm{T}}\mathbf{B}\boldsymbol{\alpha}_1 - \widehat{\boldsymbol{\alpha}}_1^{\mathrm{T}}\mathbf{B}\widehat{\boldsymbol{\alpha}}_1 = \boldsymbol{\gamma}_1^{\mathrm{T}}\boldsymbol{\Xi}\boldsymbol{\gamma}_1 - \widehat{\boldsymbol{\gamma}}_1^{\mathrm{T}}\boldsymbol{\Xi}\widehat{\boldsymbol{\gamma}}_1 \qquad (2.13)\\
&= \mu_1(\boldsymbol{\Xi}) - \sum_{k=1}^{K}\mu_k(\boldsymbol{\Xi})(\widehat{\boldsymbol{\gamma}}_1^{\mathrm{T}}\boldsymbol{\gamma}_k)^2\\
&\geq \mu_1(\boldsymbol{\Xi}) - \mu_1(\boldsymbol{\Xi})(\widehat{\boldsymbol{\gamma}}_1^{\mathrm{T}}\boldsymbol{\gamma}_1)^2 - \mu_2(\boldsymbol{\Xi})\sum_{k=2}^{K}(\widehat{\boldsymbol{\gamma}}_1^{\mathrm{T}}\boldsymbol{\gamma}_k)^2\\
&= [\mu_1(\boldsymbol{\Xi}) - \mu_2(\boldsymbol{\Xi})][1-(\widehat{\boldsymbol{\gamma}}_1^{\mathrm{T}}\boldsymbol{\gamma}_1)^2] + \mu_2(\boldsymbol{\Xi})[1-\sum_{k=1}^{K}(\widehat{\boldsymbol{\gamma}}_1^{\mathrm{T}}\boldsymbol{\gamma}_k)^2]\\
&\geq [\mu_1(\boldsymbol{\Xi}) - \mu_2(\boldsymbol{\Xi})][1-(\widehat{\boldsymbol{\gamma}}_1^{\mathrm{T}}\boldsymbol{\gamma}_1)^2] + \mu_2(\boldsymbol{\Xi})[1-\|\widehat{\boldsymbol{\gamma}}_1\|_2^2]\\
&= [\mu_1(\boldsymbol{\Xi}) - \mu_2(\boldsymbol{\Xi})][1-(\widehat{\boldsymbol{\gamma}}_1^{\mathrm{T}}\boldsymbol{\gamma}_1)^2] \geq c_2\mu_1(\boldsymbol{\Xi})[1-(\widehat{\boldsymbol{\gamma}}_1^{\mathrm{T}}\boldsymbol{\gamma}_1)^2]\\
&\geq c_2\mu_1(\boldsymbol{\Xi})[1-(\widehat{\boldsymbol{\gamma}}_1^{\mathrm{T}}\boldsymbol{\gamma}_1)] = c_2\mu_1(\boldsymbol{\Xi})\frac{1}{2}[2-2(\widehat{\boldsymbol{\gamma}}_1^{\mathrm{T}}\boldsymbol{\gamma}_1)]\\
&= \frac{1}{2}c_2\mu_1(\boldsymbol{\Xi})[\|\widehat{\boldsymbol{\gamma}}_1\|_2^2 + \|\boldsymbol{\gamma}_1\|_2^2 - 2(\widehat{\boldsymbol{\gamma}}_1^{\mathrm{T}}\boldsymbol{\gamma}_1)] = \frac{1}{2}c_2\mu_1(\boldsymbol{\Xi})\|\widehat{\boldsymbol{\gamma}}_1 - \boldsymbol{\gamma}_1\|_2^2,
\end{aligned}
$$

where the last inequality in the sixth line is due to Condition 3 and the first inequality in the seventh line is because we have assumed that $\widehat{\boldsymbol{\gamma}}_1^{\mathrm{T}}\boldsymbol{\gamma}_1 \geq 0$.

For the second term on the right hand side of last inequality of (2.11),

$$
\begin{aligned}
&2\left|\boldsymbol{\alpha}_1^{\mathrm{T}}\left[\int_0^1(\mathbf{S}\boldsymbol{\beta}(t))(\mathbf{Z}^{\mathrm{T}}\boldsymbol{\varrho}(t))^{\mathrm{T}}dt\right]\boldsymbol{\alpha}_1-\widehat{\boldsymbol{\alpha}}_1^{\mathrm{T}}\left[\int_0^1(\mathbf{S}\boldsymbol{\beta}(t))(\mathbf{Z}^{\mathrm{T}}\boldsymbol{\varrho}(t))^{\mathrm{T}}dt\right]\widehat{\boldsymbol{\alpha}}_1\right| \\
=&2\left|(\boldsymbol{\alpha}_1-\widehat{\boldsymbol{\alpha}}_1)^{\mathrm{T}}\left[\int_0^1(\mathbf{S}\boldsymbol{\beta}(t))(\mathbf{Z}^{\mathrm{T}}\boldsymbol{\varrho}(t))^{\mathrm{T}}dt\right]\boldsymbol{\alpha}_1+\widehat{\boldsymbol{\alpha}}_1^{\mathrm{T}}\left[\int_0^1(\mathbf{S}\boldsymbol{\beta}(t))(\mathbf{Z}^{\mathrm{T}}\boldsymbol{\varrho}(t))^{\mathrm{T}}dt\right](\boldsymbol{\alpha}_1-\widehat{\boldsymbol{\alpha}}_1)\right| \\
\leq&2\|(\boldsymbol{\alpha}_1-\widehat{\boldsymbol{\alpha}}_1)^{\mathrm{T}}\mathbf{S}\boldsymbol{\beta}(t)\|_{L^2}\|(\mathbf{Z}\boldsymbol{\alpha}_1)^{\mathrm{T}}\boldsymbol{\varrho}(t)\|_{L^2}+2\|\widehat{\boldsymbol{\alpha}}_1^{\mathrm{T}}\mathbf{S}\boldsymbol{\beta}(t)\|_{L^2}\|(\mathbf{Z}^{\mathrm{T}}\boldsymbol{\varrho}(t))^{\mathrm{T}}(\boldsymbol{\alpha}_1-\widehat{\boldsymbol{\alpha}}_1)\|_{L^2},
\end{aligned}
$$

where

$$
\begin{aligned}
&\|(\boldsymbol{\alpha}_1-\widehat{\boldsymbol{\alpha}}_1)^{\mathrm{T}}\mathbf{S}\boldsymbol{\beta}(t)\|_{L^2}=\sqrt{(\boldsymbol{\alpha}_1-\widehat{\boldsymbol{\alpha}}_1)^{\mathrm{T}}\mathbf{S}\left(\int_0^1\boldsymbol{\beta}(t)\boldsymbol{\beta}(t)^{\mathrm{T}}dt\right)\mathbf{S}(\boldsymbol{\alpha}_1-\widehat{\boldsymbol{\alpha}}_1)} \\
=&\sqrt{(\boldsymbol{\alpha}_1-\widehat{\boldsymbol{\alpha}}_1)^{\mathrm{T}}\mathbf{B}(\boldsymbol{\alpha}_1-\widehat{\boldsymbol{\alpha}}_1)}=\sqrt{(\boldsymbol{\gamma}_1-\widehat{\boldsymbol{\gamma}}_1)^{\mathrm{T}}\boldsymbol{\Xi}(\boldsymbol{\gamma}_1-\widehat{\boldsymbol{\gamma}}_1)} \\
\leq&\sqrt{\mu_1(\boldsymbol{\Xi})(\boldsymbol{\gamma}_1-\widehat{\boldsymbol{\gamma}}_1)^{\mathrm{T}}(\boldsymbol{\gamma}_1-\widehat{\boldsymbol{\gamma}}_1)}=\sqrt{\mu_1(\boldsymbol{\Xi})(\boldsymbol{\alpha}_1-\widehat{\boldsymbol{\alpha}}_1)^{\mathrm{T}}\mathbf{S}(\boldsymbol{\alpha}_1-\widehat{\boldsymbol{\alpha}}_1)} \\
\leq&\sqrt{\mu_1(\boldsymbol{\Xi})\|\mathbf{S}\|_\infty\|\boldsymbol{\alpha}_1-\widehat{\boldsymbol{\alpha}}_1\|_1^2}=\sqrt{\mu_1(\boldsymbol{\Xi})}\|\boldsymbol{\alpha}_1-\widehat{\boldsymbol{\alpha}}_1\|_1
\end{aligned}
$$

and by (1.2), $\|\widehat{\boldsymbol{\alpha}}_1^{\mathrm{T}}\mathbf{S}\boldsymbol{\beta}(t)\|_{L^2}=\sqrt{\widehat{\boldsymbol{\alpha}}_1^{\mathrm{T}}\mathbf{B}\widehat{\boldsymbol{\alpha}}_1}\leq\sqrt{\mu_1(\boldsymbol{\Xi})}$. By the definition (1.4) of Ω, $\|(\mathbf{Z}^{\mathrm{T}}\boldsymbol{\varrho}(t))^{\mathrm{T}}(\boldsymbol{\alpha}_1-\widehat{\boldsymbol{\alpha}}_1)\|_{L^2}\leq\max_{1\leq j\leq p}\|(\mathbf{Z}^{\mathrm{T}}\boldsymbol{\varrho}(t))_j\|_{L^2}\|\boldsymbol{\alpha}_1-\widehat{\boldsymbol{\alpha}}_1\|_1\leq C_0\varpi\|\boldsymbol{\alpha}_1-\widehat{\boldsymbol{\alpha}}_1\|_1$, and similarly, $\|(\mathbf{Z}\boldsymbol{\alpha}_1)^{\mathrm{T}}\boldsymbol{\varrho}(t)\|_{L^2}\leq C_0\varpi$. Hence, the inequalities above lead to

$$
\begin{aligned}
&2\left|\boldsymbol{\alpha}_1^{\mathrm{T}}\left[\int_0^1(\mathbf{S}\boldsymbol{\beta}(t))(\mathbf{Z}^{\mathrm{T}}\boldsymbol{\varrho}(t))^{\mathrm{T}}dt\right]\boldsymbol{\alpha}_1-\widehat{\boldsymbol{\alpha}}_1^{\mathrm{T}}\left[\int_0^1(\mathbf{S}\boldsymbol{\beta}(t))(\mathbf{Z}^{\mathrm{T}}\boldsymbol{\varrho}(t))^{\mathrm{T}}dt\right]\widehat{\boldsymbol{\alpha}}_1\right| \qquad (2.14)\\
\leq&2\sqrt{\mu_1(\boldsymbol{\Xi})}\|\boldsymbol{\alpha}_1-\widehat{\boldsymbol{\alpha}}_1\|_1C_0\varpi+2\sqrt{\mu_1(\boldsymbol{\Xi})}C_0\varpi\|\boldsymbol{\alpha}_1-\widehat{\boldsymbol{\alpha}}_1\|_1 \\
\leq&4\sqrt{\mu_1(\boldsymbol{\Xi})}C_0\varpi\|\boldsymbol{\alpha}_1-\widehat{\boldsymbol{\alpha}}_1\|_1
\end{aligned}
$$

For the third term on the right hand side of last inequality of (2.11), we have

$$\begin{aligned}
&\left|\boldsymbol{\alpha}_1^{\mathrm{T}}\left[\int_0^1(\mathbf{Z}^{\mathrm{T}}\boldsymbol{\varrho}(t))(\mathbf{Z}^{\mathrm{T}}\boldsymbol{\varrho}(t))^{\mathrm{T}}dt\right]\boldsymbol{\alpha}_1-\widehat{\boldsymbol{\alpha}}_1^{\mathrm{T}}\left[\int_0^1(\mathbf{Z}^{\mathrm{T}}\boldsymbol{\varrho}(t))(\mathbf{Z}^{\mathrm{T}}\boldsymbol{\varrho}(t))^{\mathrm{T}}dt\right]\widehat{\boldsymbol{\alpha}}_1\right| \\
=&\left|(\boldsymbol{\alpha}_1-\widehat{\boldsymbol{\alpha}}_1)^{\mathrm{T}}\left[\int_0^1(\mathbf{Z}^{\mathrm{T}}\boldsymbol{\varrho}(t))(\mathbf{Z}^{\mathrm{T}}\boldsymbol{\varrho}(t))^{\mathrm{T}}dt\right](\boldsymbol{\alpha}_1+\widehat{\boldsymbol{\alpha}}_1)\right| \\
\le&\left(\max_{1\le j\le p}\|(\mathbf{Z}^{\mathrm{T}}\boldsymbol{\varrho}(t))_j\|_{L^2}\right)^2\|\boldsymbol{\alpha}_1-\widehat{\boldsymbol{\alpha}}_1\|_1\|\boldsymbol{\alpha}_1+\widehat{\boldsymbol{\alpha}}_1\|_1 \\
\le&C_0^2\varpi^2\|\boldsymbol{\alpha}_1-\widehat{\boldsymbol{\alpha}}_1\|_1(\|\boldsymbol{\alpha}_1\|_1+\|\widehat{\boldsymbol{\alpha}}_1\|_1) \\
\le&C_0^2\varpi^2\|\boldsymbol{\alpha}_1-\widehat{\boldsymbol{\alpha}}_1\|_1(1+D_1)\|\boldsymbol{\alpha}_1\|_1 \\
=&\left[C_0\varpi\|\boldsymbol{\alpha}_1\|_1\right](1+D_1)C_0\varpi\|\boldsymbol{\alpha}_1-\widehat{\boldsymbol{\alpha}}_1\|_1 \\
\le&\hbar^{-1}\sqrt{\mu_1(\boldsymbol{\Xi})}(1+D_1)C_0\varpi\|\boldsymbol{\alpha}_1-\widehat{\boldsymbol{\alpha}}_1\|_1 \\
\le&\sqrt{\mu_1(\boldsymbol{\Xi})}[\hbar^{-1}(1+D_1)]C_0\varpi\|\boldsymbol{\alpha}_1-\widehat{\boldsymbol{\alpha}}_1\|_1, \qquad (2.15)
\end{aligned}$$

where the inequality in the fifth line follows from $\|\widehat{\boldsymbol{\alpha}}_1\|_1\le D_1\|\boldsymbol{\alpha}_1\|_1$, and the inequality in the seventh line is due to (1.5). Combining (2.11)-(2.15) gives

$$\begin{aligned}
&\boldsymbol{\alpha}_1^{\mathrm{T}}\widehat{\mathbf{B}}\boldsymbol{\alpha}_1-\widehat{\boldsymbol{\alpha}}_1^{\mathrm{T}}\widehat{\mathbf{B}}\widehat{\boldsymbol{\alpha}}_1 \\
\ge&\frac{1}{2}c_2\mu_1(\boldsymbol{\Xi})\|\widehat{\boldsymbol{\gamma}}_1-\boldsymbol{\gamma}_1\|_{L^2}^2-\{4+\hbar^{-1}(1+D_1)\}\sqrt{\mu_1(\boldsymbol{\Xi})}C_0\varpi\|\boldsymbol{\alpha}_1-\widehat{\boldsymbol{\alpha}}_1\|_1.
\end{aligned}$$

We have proved the lemma.

Proof of Lemma 4.

Note that

$$\begin{aligned}
&\tau^{(1)}\|\boldsymbol{\alpha}_1\|_{\lambda^{(1)}}^2-\tau^{(1)}\|\widehat{\boldsymbol{\alpha}}_1\|_{\lambda^{(1)}}^2 \qquad (2.16)\\
=&\frac{A^{(1)}C_0\varpi}{\|\boldsymbol{\alpha}_1\|_1\sqrt{\mu_1(\boldsymbol{\Xi})}}\left[(1-\lambda^{(1)})(\|\boldsymbol{\alpha}_1\|_2^2-\|\widehat{\boldsymbol{\alpha}}_1\|_2^2)+\lambda^{(1)}(\|\boldsymbol{\alpha}_1\|_1^2-\|\widehat{\boldsymbol{\alpha}}_1\|_1^2)\right].
\end{aligned}$$

Because $\|(\boldsymbol{\alpha}_1)_{J_1^c}\|_1=0$, $\|(\boldsymbol{\alpha}_1)_{J_1}\|_1=\|\boldsymbol{\alpha}_1\|_1$, $\|(\boldsymbol{\alpha}_1)_{J_1^c}\|_2=0$ and $\|(\boldsymbol{\alpha}_1)_{J_1}\|_2=$

$\|\boldsymbol{\alpha}_1\|_2$, we have

$$\begin{aligned}
&\frac{A^{(1)}C_0\varpi}{\|\boldsymbol{\alpha}_1\|_1\sqrt{\mu_1(\boldsymbol{\Xi})}}(1-\lambda^{(1)})(\|\boldsymbol{\alpha}_1\|_2^2-\|\widehat{\boldsymbol{\alpha}}_1\|_2^2) \qquad (2.17)\\
=&\frac{A^{(1)}C_0\varpi}{\|\boldsymbol{\alpha}_1\|_1\sqrt{\mu_1(\boldsymbol{\Xi})}}(1-\lambda^{(1)})(\|(\boldsymbol{\alpha}_1)_{J_1}\|_2^2-\|(\widehat{\boldsymbol{\alpha}}_1)_{J_1}\|_2^2-\|(\widehat{\boldsymbol{\alpha}}_1)_{J_1^c}\|_2^2)\\
\le&\frac{A^{(1)}C_0\varpi}{\|\boldsymbol{\alpha}_1\|_1\sqrt{\mu_1(\boldsymbol{\Xi})}}(1-\lambda^{(1)})(\|(\boldsymbol{\alpha}_1)_{J_1}\|_2^2-\|(\widehat{\boldsymbol{\alpha}}_1)_{J_1}\|_2^2)\\
=&\frac{A^{(1)}C_0\varpi}{\|\boldsymbol{\alpha}_1\|_1\sqrt{\mu_1(\boldsymbol{\Xi})}}(1-\lambda^{(1)})(\|(\boldsymbol{\alpha}_1)_{J_1}\|_2-\|(\widehat{\boldsymbol{\alpha}}_1)_{J_1}\|_2)(\|(\boldsymbol{\alpha}_1)_{J_1}\|_2+\|(\widehat{\boldsymbol{\alpha}}_1)_{J_1}\|_2)\\
\le&\frac{A^{(1)}C_0\varpi}{\|\boldsymbol{\alpha}_1\|_1\sqrt{\mu_1(\boldsymbol{\Xi})}}(1-\lambda^{(1)})\|(\widehat{\boldsymbol{\alpha}}_1-\boldsymbol{\alpha}_1)_{J_1}\|_2(\|\boldsymbol{\alpha}_1\|_2+\|\widehat{\boldsymbol{\alpha}}_1\|_2)\\
\le&\frac{A^{(1)}C_0\varpi}{\|\boldsymbol{\alpha}_1\|_1\sqrt{\mu_1(\boldsymbol{\Xi})}}(1-\lambda^{(1)})\|(\widehat{\boldsymbol{\alpha}}_1-\boldsymbol{\alpha}_1)_{J_1}\|_1(\|\boldsymbol{\alpha}_1\|_1+\|\widehat{\boldsymbol{\alpha}}_1\|_1)\\
=&\frac{A^{(1)}C_0\varpi}{\sqrt{\mu_1(\boldsymbol{\Xi})}}(1-\lambda^{(1)})\|(\widehat{\boldsymbol{\alpha}}_1-\boldsymbol{\alpha}_1)_{J_1}\|_1(1+\|\widehat{\boldsymbol{\alpha}}_1\|_1/\|\boldsymbol{\alpha}_1\|_1),
\end{aligned}$$

and

$$\begin{aligned}
&\frac{A^{(1)}C_0\varpi}{\|\boldsymbol{\alpha}_1\|_1\sqrt{\mu_1(\boldsymbol{\Xi})}}\lambda^{(1)}(\|\boldsymbol{\alpha}_1\|_1^2-\|\widehat{\boldsymbol{\alpha}}_1\|_1^2) \qquad (2.18)\\
=&\frac{A^{(1)}C_0\varpi}{\|\boldsymbol{\alpha}_1\|_1\sqrt{\mu_1(\boldsymbol{\Xi})}}\lambda^{(1)}(\|\boldsymbol{\alpha}_1\|_1-\|\widehat{\boldsymbol{\alpha}}_1\|_1)(\|\boldsymbol{\alpha}_1\|_1+\|\widehat{\boldsymbol{\alpha}}_1\|_1)\\
=&\frac{A^{(1)}C_0\varpi}{\sqrt{\mu_1(\boldsymbol{\Xi})}}\lambda^{(1)}(\|(\boldsymbol{\alpha}_1)_{J_1}\|_1-\|(\widehat{\boldsymbol{\alpha}}_1)_{J_1}\|_1-\|(\widehat{\boldsymbol{\alpha}}_1)_{J_1^c}\|_1)(1+\|\widehat{\boldsymbol{\alpha}}_1\|_1/\|\boldsymbol{\alpha}_1\|_1)\\
\le&\frac{A^{(1)}C_0\varpi}{\sqrt{\mu_1(\boldsymbol{\Xi})}}(\lambda^{(1)}\|(\boldsymbol{\alpha}_1-\widehat{\boldsymbol{\alpha}}_1)_{J_1}\|_1-\lambda^{(1)}\|(\widehat{\boldsymbol{\alpha}}_1)_{J_1^c}\|_1)(1+\|\widehat{\boldsymbol{\alpha}}_1\|_1/\|\boldsymbol{\alpha}_1\|_1).
\end{aligned}$$

By (2.16)-(2.18), we have

$$\begin{aligned}
&\tau^{(1)}\|\boldsymbol{\alpha}_1\|_{\lambda^{(1)}}^2-\tau^{(1)}\|\widehat{\boldsymbol{\alpha}}_1\|_{\lambda^{(1)}}^2 \qquad (2.19)\\
\le&\frac{A^{(1)}C_0\varpi}{\sqrt{\mu_1(\boldsymbol{\Xi})}}(\|(\boldsymbol{\alpha}_1-\widehat{\boldsymbol{\alpha}}_1)_{J_1}\|_1-\lambda^{(1)}\|(\widehat{\boldsymbol{\alpha}}_1)_{J_1^c}\|_1)(1+\|\widehat{\boldsymbol{\alpha}}_1\|_1/\|\boldsymbol{\alpha}_1\|_1).
\end{aligned}$$

In the following, we will consider Cases (a) and (b), separately.

Case (a): $\|(\boldsymbol{\alpha}_1-\widehat{\boldsymbol{\alpha}}_1)_{J_1}\|_1<\lambda^{(1)}\|(\widehat{\boldsymbol{\alpha}}_1)_{J_1^c}\|_1$.

In this case, by (2.19), we have $\tau^{(1)}\|\boldsymbol{\alpha}_1\|^2_{\lambda^{(1)}} - \tau^{(1)}\|\widehat{\boldsymbol{\alpha}}_1\|^2_{\lambda^{(1)}} < 0$. By the following inequality (which is the first inequality of (1.6)),

$$\widehat{\boldsymbol{\alpha}}_1^{\mathrm{T}}\widehat{\mathbf{B}}\widehat{\boldsymbol{\alpha}}_1(\tau^{(1)}\|\boldsymbol{\alpha}_1\|^2_{\lambda^{(1)}} - \tau^{(1)}\|\widehat{\boldsymbol{\alpha}}_1\|^2_{\lambda^{(1)}}) \geq (\boldsymbol{\alpha}_1^{\mathrm{T}}\widehat{\mathbf{B}}\boldsymbol{\alpha}_1 - \widehat{\boldsymbol{\alpha}}_1^{\mathrm{T}}\widehat{\mathbf{B}}\widehat{\boldsymbol{\alpha}}_1)(1 + \tau^{(1)}\|\widehat{\boldsymbol{\alpha}}_1\|^2_{\lambda^{(1)}}),$$

we have $\boldsymbol{\alpha}_1^{\mathrm{T}}\widehat{\mathbf{B}}\boldsymbol{\alpha}_1 < \widehat{\boldsymbol{\alpha}}_1^{\mathrm{T}}\widehat{\mathbf{B}}\widehat{\boldsymbol{\alpha}}_1$. Now by (2.3) and (1.5),

$$\widehat{\boldsymbol{\alpha}}_1^{\mathrm{T}}\widehat{\mathbf{B}}\widehat{\boldsymbol{\alpha}}_1 > \boldsymbol{\alpha}_1^{\mathrm{T}}\widehat{\mathbf{B}}\boldsymbol{\alpha}_1 \geq \mu_1(\boldsymbol{\Xi}) - 2\sqrt{\mu_1(\boldsymbol{\Xi})}C_0\varpi\|\boldsymbol{\alpha}_1\|_1 \geq (1 - 2\hbar^{-1})\mu_1(\boldsymbol{\Xi})$$

which together with (2.19) give

$$\begin{aligned}
&\widehat{\boldsymbol{\alpha}}_1^{\mathrm{T}}\widehat{\mathbf{B}}\widehat{\boldsymbol{\alpha}}_1(\tau^{(1)}\|\boldsymbol{\alpha}_1\|^2_{\lambda^{(1)}} - \tau^{(1)}\|\widehat{\boldsymbol{\alpha}}_1\|^2_{\lambda^{(1)}})\\
\leq&(1 - 2\hbar^{-1})\mu_1(\boldsymbol{\Xi})\frac{A^{(1)}C_0\varpi}{\sqrt{\mu_1(\boldsymbol{\Xi})}}\left[\|(\boldsymbol{\alpha}_1 - \widehat{\boldsymbol{\alpha}}_1)_{J_1}\|_1 - \lambda^{(1)}\|(\widehat{\boldsymbol{\alpha}}_1)_{J_1^c}\|_1\right](1 + \|\widehat{\boldsymbol{\alpha}}_1\|_1/\|\boldsymbol{\alpha}_1\|_1)\\
\leq&(1 - 2\hbar^{-1})A^{(1)}C_0\mu_1(\boldsymbol{\Xi})^{1/2}\varpi(\|(\boldsymbol{\alpha}_1 - \widehat{\boldsymbol{\alpha}}_1)_{J_1}\|_1 - \lambda^{(1)}\|(\widehat{\boldsymbol{\alpha}}_1)_{J_1^c}\|_1) \leq 0,
\end{aligned}$$

where we use $1 + \|\widehat{\boldsymbol{\alpha}}_1\|_1/\|\boldsymbol{\alpha}_1\|_1 \geq 1$.

Case (b): $\|(\boldsymbol{\alpha}_1 - \widehat{\boldsymbol{\alpha}}_1)_{J_1}\|_1 \geq \lambda^{(1)}\|(\widehat{\boldsymbol{\alpha}}_1)_{J_1^c}\|_1$.

By Lemma 3, we have $\|\widehat{\boldsymbol{\alpha}}_1\|_1 \leq D_1\|\boldsymbol{\alpha}_1\|_1$ and $\widehat{\boldsymbol{\alpha}}_1^{\mathrm{T}}\widehat{\mathbf{B}}\widehat{\boldsymbol{\alpha}}_1 \leq (1+\hbar^{-1}D_1)^2\mu_1(\boldsymbol{\Xi})$. Therefore, by (2.19), we have

$$\begin{aligned}
&\widehat{\boldsymbol{\alpha}}_1^{\mathrm{T}}\widehat{\mathbf{B}}\widehat{\boldsymbol{\alpha}}_1(\tau^{(1)}\|\boldsymbol{\alpha}_1\|^2_{\lambda^{(1)}} - \tau^{(1)}\|\widehat{\boldsymbol{\alpha}}_1\|^2_{\lambda^{(1)}})\\
\leq&\widehat{\boldsymbol{\alpha}}_1^{\mathrm{T}}\widehat{\mathbf{B}}\widehat{\boldsymbol{\alpha}}_1\frac{A^{(1)}C_0\varpi}{\sqrt{\mu_1(\boldsymbol{\Xi})}}(\|(\boldsymbol{\alpha}_1 - \widehat{\boldsymbol{\alpha}}_1)_{J_1}\|_1 - \lambda^{(1)}\|(\widehat{\boldsymbol{\alpha}}_1)_{J_1^c}\|_1)(1 + \|\widehat{\boldsymbol{\alpha}}_1\|_1/\|\boldsymbol{\alpha}_1\|_1)\\
\leq&(1+\hbar^{-1}D_1)^2A^{(1)}C_0\mu_1(\boldsymbol{\Xi})^{1/2}\varpi(\|(\boldsymbol{\alpha}_1 - \widehat{\boldsymbol{\alpha}}_1)_{J_1}\|_1 - \lambda^{(1)}\|(\widehat{\boldsymbol{\alpha}}_1)_{J_1^c}\|_1)(1 + D_1)\\
\leq&(1+\hbar^{-1}D_1)^2(1+D_1)A^{(1)}C_0\mu_1(\boldsymbol{\Xi})^{1/2}\varpi\|(\boldsymbol{\alpha}_1 - \widehat{\boldsymbol{\alpha}}_1)_{J_1}\|_1.
\end{aligned}$$

The proof is completed.

Proof of Lemma 5.

Because in this case,

$$\begin{aligned}
&\|(\widehat{\boldsymbol{\alpha}}_1 - \boldsymbol{\alpha}_1)_{J_1^c}\|_1 = \|(\widehat{\boldsymbol{\alpha}}_1)_{J_1^c}\|_1\\
\leq&(\lambda^{(1)})^{-1}\|(\widehat{\boldsymbol{\alpha}}_1 - \boldsymbol{\alpha}_1)_{J_1}\|_1 < c\|(\widehat{\boldsymbol{\alpha}}_1 - \boldsymbol{\alpha}_1)_{J_1}\|_1.
\end{aligned} \tag{2.20}$$

by Condition 1,

$$\kappa\|(\widehat{\boldsymbol{\alpha}}_1-\boldsymbol{\alpha}_1)_{J_1}\|_2 \le \|\frac{\mathbf{X}}{\sqrt{n}}(\widehat{\boldsymbol{\alpha}}_1-\boldsymbol{\alpha}_1)\|_2 = \|\mathbf{Z}(\widehat{\boldsymbol{\alpha}}_1-\boldsymbol{\alpha}_1)\|_2 = \|\widehat{\boldsymbol{\gamma}}_1-\boldsymbol{\gamma}_1\|_2. \tag{2.21}$$

Moreover, because $\|\widehat{\boldsymbol{\alpha}}_1-\boldsymbol{\alpha}_1\|_1 = \|(\widehat{\boldsymbol{\alpha}}_1-\boldsymbol{\alpha}_1)_{J_1^c}\|_1 + \|(\widehat{\boldsymbol{\alpha}}_1-\boldsymbol{\alpha}_1)_{J_1}\|_1 < (1+c)\|(\widehat{\boldsymbol{\alpha}}_1-\boldsymbol{\alpha}_1)_{J_1}\|_1$, by (1.6), Lemma 4 and (2.21), we have

$$\begin{aligned}
&\frac{1}{2}c_2\mu_1(\boldsymbol{\Xi})\|\widehat{\boldsymbol{\gamma}}_1-\boldsymbol{\gamma}_1\|_2^2\\
&\le \Big[(1+\hbar^{-1}D_1)^2(1+D_1)A^{(1)} + (1+c)\{4+\hbar^{-1}(1+D_1)\}(1+D_1^2A^{(1)}\hbar^{-1})\Big]\\
&\qquad \times \sqrt{\mu_1(\boldsymbol{\Xi})}C_0\varpi\|(\widehat{\boldsymbol{\alpha}}_1-\boldsymbol{\alpha}_1)_{J_1}\|_1\\
&\le \Big[(1+\hbar^{-1}D_1)^2(1+D_1)A^{(1)} + (1+c)\{4+\hbar^{-1}(1+D_1)\}(1+D_1^2A^{(1)}\hbar^{-1})\Big]\\
&\qquad \times \sqrt{\mu_1(\boldsymbol{\Xi})}C_0\varpi\sqrt{s}\|(\widehat{\boldsymbol{\alpha}}_1-\boldsymbol{\alpha}_1)_{J_1}\|_2\\
&\le \Big[(1+\hbar^{-1}D_1)^2(1+D_1)A^{(1)} + (1+c)\{4+\hbar^{-1}(1+D_1)\}(1+D_1^2A^{(1)}\hbar^{-1})\Big]\\
&\qquad \times \sqrt{\mu_1(\boldsymbol{\Xi})}C_0\varpi\sqrt{s}\|\widehat{\boldsymbol{\gamma}}_1-\boldsymbol{\gamma}_1\|_2/\kappa.
\end{aligned}$$

Similar to Case (a), we obtain

$$\begin{aligned}
&\|\widehat{\boldsymbol{\gamma}}_1-\boldsymbol{\gamma}_1\|_2 \le D_3C_0\kappa^{-1}\mu_1(\boldsymbol{\Xi})^{-1/2}\varpi\sqrt{s},\\
&\|\mathbf{Z}(\widehat{\boldsymbol{\alpha}}_1-\boldsymbol{\alpha}_1)\|_2^2 = \|\widehat{\boldsymbol{\gamma}}_1-\boldsymbol{\gamma}_1\|_2^2 \le D_3^2C_0^2\kappa^{-2}\mu_1(\boldsymbol{\Xi})^{-1}\varpi^2 s,\\
&\|\mathbf{X}(\widehat{\boldsymbol{\alpha}}_1-\boldsymbol{\alpha}_1)\|_2^2 = n\|\widehat{\boldsymbol{\gamma}}_1-\boldsymbol{\gamma}_1\|_2^2 \le nD_3^2C_0^2\kappa^{-2}\mu_1(\boldsymbol{\Xi})^{-1}\varpi^2 s,\\
&\|\widehat{\boldsymbol{\alpha}}_1-\boldsymbol{\alpha}_1\|_1 \le (1+c)\|(\widehat{\boldsymbol{\alpha}}_1-\boldsymbol{\alpha}_1)_{J_1}\|_1 \le (1+c)\sqrt{s}\|(\widehat{\boldsymbol{\alpha}}_1-\boldsymbol{\alpha}_1)_{J_1}\|_2\\
&\le (1+c)\sqrt{s}\kappa^{-1}\|\widehat{\boldsymbol{\gamma}}_1-\boldsymbol{\gamma}_1\|_2 \le (1+c)D_3C_0\kappa^{-2}\mu_1(\boldsymbol{\Xi})^{-1/2}\varpi s
\end{aligned}$$

where

$$\begin{aligned}
D_3 =& 2c_2^{-1}\Big[(1+\hbar^{-1}D_1)^2(1+D_1)A^{(1)}\\
&+(1+c)\{4+\hbar^{-1}(1+D_1)\}(1+D_1^2A^{(1)}\hbar^{-1})\Big].
\end{aligned}$$

Note that

$$\lim_{\substack{A^{(1)}\to\infty\\ \hbar\to\infty}} \frac{D_3}{A^{(1)}} \le 2c_2^{-1}(1+D_1) = 2c_2^{-1}(1+\sqrt{6c}). \tag{2.22}$$

Hence, there exist $(A_1^L)''''$ and $(\hbar_0)''''$ only depending on δ_0, c and c_2 such that for any $A^{(1)} \ge (A_1^L)''''$ and $\hbar \ge (\hbar_0)''''$, we have

$$D_3 \le 4c_2^{-1}(1+\sqrt{6c})A^{(1)},$$

and the inequalities in Theorem 4.2 (a) follows in this case.

Proof of Lemma 6.

By (1.18) and the similar arguments as in (2.3), we can obtain

$$\begin{aligned}
&\boldsymbol{\beta}_k^{\mathrm{T}}\widehat{\mathbf{B}}\boldsymbol{\beta}_k \geq \boldsymbol{\beta}_k^{\mathrm{T}}\mathbf{B}\boldsymbol{\beta}_k - 2\sqrt{\mu_1(\boldsymbol{\Xi})}C_0\varpi\|\boldsymbol{\beta}_k\|_1 \\
=&\boldsymbol{\beta}_k^{\mathrm{T}}\mathbf{Z}^{\mathrm{T}}\boldsymbol{\Xi}\mathbf{Z}\boldsymbol{\beta}_k - 2\sqrt{\mu_1(\boldsymbol{\Xi})}C_0\varpi\|\boldsymbol{\beta}_k\|_1 \\
=&\boldsymbol{\gamma}_k^{\mathrm{T}}(\mathbf{I}-\widehat{\mathbf{P}}_{k-1})\boldsymbol{\Xi}(\mathbf{I}-\widehat{\mathbf{P}}_{k-1})\boldsymbol{\gamma}_k - 2\sqrt{\mu_1(\boldsymbol{\Xi})}C_0\varpi\|\boldsymbol{\beta}_k\|_1 \\
=&\boldsymbol{\gamma}_k^{\mathrm{T}}\boldsymbol{\Xi}\boldsymbol{\gamma}_k - 2\boldsymbol{\gamma}_k^{\mathrm{T}}\widehat{\mathbf{P}}_{k-1}\boldsymbol{\Xi}\boldsymbol{\gamma}_k + \boldsymbol{\gamma}_k^{\mathrm{T}}\widehat{\mathbf{P}}_{k-1}\boldsymbol{\Xi}\widehat{\mathbf{P}}_{k-1}\boldsymbol{\gamma}_k - 2\sqrt{\mu_1(\boldsymbol{\Xi})}C_0\varpi\|\boldsymbol{\beta}_k\|_1 \\
\geq&\boldsymbol{\gamma}_k^{\mathrm{T}}\boldsymbol{\Xi}\boldsymbol{\gamma}_k - 2\boldsymbol{\gamma}_k^{\mathrm{T}}\widehat{\mathbf{P}}_{k-1}\boldsymbol{\Xi}\boldsymbol{\gamma}_k - 2\sqrt{\mu_1(\boldsymbol{\Xi})}C_0\varpi\|\boldsymbol{\beta}_k\|_1 \\
=&\mu_k(\boldsymbol{\Xi}) - 2\mu_k(\boldsymbol{\Xi})\boldsymbol{\gamma}_k^{\mathrm{T}}\widehat{\mathbf{P}}_{k-1}\boldsymbol{\gamma}_k - 2\sqrt{\mu_1(\boldsymbol{\Xi})}C_0\varpi\|\boldsymbol{\beta}_k\|_1 \\
=&\mu_k(\boldsymbol{\Xi}) - 2\mu_k(\boldsymbol{\Xi})\|\widehat{\mathbf{P}}_{k-1}\boldsymbol{\gamma}_k\|_2^2 - 2\sqrt{\mu_1(\boldsymbol{\Xi})}C_0\varpi\|\boldsymbol{\beta}_k\|_1 \\
\geq&\mu_k(\boldsymbol{\Xi}) - 2\mu_1(\boldsymbol{\Xi})\|\widehat{\mathbf{P}}_{k-1}\boldsymbol{\gamma}_k\|_2^2 - 2\sqrt{\mu_1(\boldsymbol{\Xi})}M_{k,1}\kappa^{-1}C_0\varpi\sqrt{s},
\end{aligned} \tag{2.23}$$

where the last equality is due to (1.24). Because $\mathbf{P}_{k-1}\boldsymbol{\gamma}_k = \mathbf{0}$,

$$\begin{aligned}
&\|\widehat{\mathbf{P}}_{k-1}\boldsymbol{\gamma}_k\|_2^2 = \|(\widehat{\mathbf{P}}_{k-1}-\mathbf{P}_{k-1})\boldsymbol{\gamma}_k\|_2^2 \leq \|\widehat{\mathbf{P}}_{k-1}-\mathbf{P}_{k-1}\|^2 \\
\leq&b_{k-1,3}C_0^2\kappa^{-2}\mu_1(\boldsymbol{\Xi})^{-1}\varpi^2 s = b_{k-1,3}C_0\kappa^{-2}\mu_1(\boldsymbol{\Xi})^{-1/2}\varpi\sqrt{s}\frac{C_0\varpi\sqrt{s}}{\mu_1(\boldsymbol{\Xi})^{1/2}} \\
\leq&b_{k-1,3}C_0\kappa^{-2}\mu_1(\boldsymbol{\Xi})^{-1/2}\varpi\sqrt{s}\hbar^{-1}\kappa = b_{k-1,3}\hbar^{-1}C_0\kappa^{-1}\mu_1(\boldsymbol{\Xi})^{-1/2}\varpi\sqrt{s},
\end{aligned} \tag{2.24}$$

where the first inequality in the second line follows from (1.22) and the first inequality in the last line is due to (1.15). Combining (2.24) and (2.23) gives the second inequality in the lemma.

$$\boldsymbol{\beta}_k^{\mathrm{T}}\widehat{\mathbf{B}}\boldsymbol{\beta}_k \geq \mu_k(\boldsymbol{\Xi}) - N_{k,3}C_0\kappa^{-1}\mu_1(\boldsymbol{\Xi})^{1/2}\varpi\sqrt{s}, \tag{2.25}$$

where $N_{k,3} = 2(b_{k-1,3}\hbar^{-1} + M_{k,1})$. On the other hand,

$$\begin{aligned}
&\boldsymbol{\beta}_k^{\mathrm{T}}\mathbf{S}\boldsymbol{\beta}_k = \boldsymbol{\beta}_k^{\mathrm{T}}\mathbf{Z}^{\mathrm{T}}\mathbf{Z}\boldsymbol{\beta}_k = \boldsymbol{\gamma}_k^{\mathrm{T}}(\mathbf{I}-\widehat{\mathbf{P}}_{k-1})(\mathbf{I}-\widehat{\mathbf{P}}_{k-1})\boldsymbol{\gamma}_k \\
=&\boldsymbol{\gamma}_k^{\mathrm{T}}(\mathbf{I}-\widehat{\mathbf{P}}_{k-1})\boldsymbol{\gamma}_k = \|\boldsymbol{\gamma}_k\|_2^2 - \boldsymbol{\gamma}_k^{\mathrm{T}}\widehat{\mathbf{P}}_{k-1}\boldsymbol{\gamma}_k \\
=&\|\boldsymbol{\gamma}_k\|_2^2 - \|\widehat{\mathbf{P}}_{k-1}\boldsymbol{\gamma}_k\|_2^2 \leq \|\boldsymbol{\gamma}_k\|_2^2 = 1,
\end{aligned} \tag{2.26}$$

which is the first inequality in the lemma. Now we prove the third inequality in the lemma. By the similar arguments as in (2.5), we have

$$
\begin{aligned}
&\widehat{\boldsymbol{\alpha}}_k^{\mathrm{T}}\widehat{\mathbf{B}}\widehat{\boldsymbol{\alpha}}_k \qquad (2.27)\\
\le&\widehat{\boldsymbol{\alpha}}_k^{\mathrm{T}}\mathbf{B}\widehat{\boldsymbol{\alpha}}_k + 2\sqrt{\mu_1(\boldsymbol{\Xi})}C_0\varpi\|\widehat{\boldsymbol{\alpha}}_k\|_1 + C_0^2\varpi^2\|\widehat{\boldsymbol{\alpha}}_k\|_1^2\\
=&\widehat{\boldsymbol{\alpha}}_k^{\mathrm{T}}\mathbf{B}\widehat{\boldsymbol{\alpha}}_k + 2\sqrt{\mu_1(\boldsymbol{\Xi})}\frac{C_0\varpi}{\sqrt{s}}\sqrt{s}\|\widehat{\boldsymbol{\alpha}}_k\|_1 + (C_0\varpi)C_0\varpi\|\widehat{\boldsymbol{\alpha}}_k\|_1^2\\
\le&\widehat{\boldsymbol{\alpha}}_k^{\mathrm{T}}\mathbf{B}\widehat{\boldsymbol{\alpha}}_k + \sqrt{\mu_1(\boldsymbol{\Xi})}\frac{C_0\varpi}{\sqrt{s}}[\kappa^{-1}s + \kappa\|\widehat{\boldsymbol{\alpha}}_k\|_1^2] + (C_0\varpi)C_0\varpi\|\widehat{\boldsymbol{\alpha}}_k\|_1^2\\
\le&\widehat{\boldsymbol{\alpha}}_k^{\mathrm{T}}\mathbf{B}\widehat{\boldsymbol{\alpha}}_k + \sqrt{\mu_1(\boldsymbol{\Xi})}\frac{C_0\varpi}{\sqrt{s}}[\kappa^{-1}s + \kappa\|\widehat{\boldsymbol{\alpha}}_k\|_1^2] + \frac{\kappa\mu_1(\boldsymbol{\Xi})^{1/2}\hbar^{-1}}{\sqrt{s}}C_0\varpi\|\widehat{\boldsymbol{\alpha}}_k\|_1^2\\
=&\widehat{\boldsymbol{\alpha}}_k^{\mathrm{T}}\mathbf{B}\widehat{\boldsymbol{\alpha}}_k + \sqrt{\mu_1(\boldsymbol{\Xi})}C_0\kappa^{-1}\varpi\sqrt{s} + \frac{(1+\hbar^{-1})}{\sqrt{s}}\kappa\sqrt{\mu_1(\boldsymbol{\Xi})}C_0\varpi\|\widehat{\boldsymbol{\alpha}}_k\|_1^2,
\end{aligned}
$$

where the inequality in the fourth line is due to the inequality: $a^2 + b^2 \ge 2ab$, and the inequality in the fourth line follows from (1.15). We estimate $\widehat{\boldsymbol{\alpha}}_k^{\mathrm{T}}\mathbf{B}\widehat{\boldsymbol{\alpha}}_k$. By the eigen-decomposition of $\boldsymbol{\Xi}$,

$$
\begin{aligned}
&\widehat{\boldsymbol{\alpha}}_k^{\mathrm{T}}\mathbf{B}\widehat{\boldsymbol{\alpha}}_k = \widehat{\boldsymbol{\gamma}}_k^{\mathrm{T}}\boldsymbol{\Xi}\widehat{\boldsymbol{\gamma}}_k = \widehat{\boldsymbol{\gamma}}_k^{\mathrm{T}}(\sum_{i=1}^{K}\mu_i(\boldsymbol{\Xi})\boldsymbol{\gamma}_i\boldsymbol{\gamma}_i^{\mathrm{T}})\widehat{\boldsymbol{\gamma}}_k = \sum_{i=1}^{K}\mu_i(\boldsymbol{\Xi})(\boldsymbol{\gamma}_i^{\mathrm{T}}\widehat{\boldsymbol{\gamma}}_k)^2 \qquad (2.28)\\
\le&\mu_1(\boldsymbol{\Xi})\sum_{i=1}^{k-1}(\boldsymbol{\gamma}_i^{\mathrm{T}}\widehat{\boldsymbol{\gamma}}_k)^2 + \mu_k(\boldsymbol{\Xi})\sum_{i=k}^{K}(\boldsymbol{\gamma}_i^{\mathrm{T}}\widehat{\boldsymbol{\gamma}}_k)^2\\
=&\mu_1(\boldsymbol{\Xi})\|\mathbf{P}_{k-1}\widehat{\boldsymbol{\gamma}}_k\|_2^2 + \mu_k(\boldsymbol{\Xi})\sum_{i=k}^{K}(\boldsymbol{\gamma}_i^{\mathrm{T}}\widehat{\boldsymbol{\gamma}}_k)^2\\
\le&\mu_1(\boldsymbol{\Xi})\|\mathbf{P}_{k-1}\widehat{\boldsymbol{\gamma}}_k\|_2^2 + \mu_k(\boldsymbol{\Xi})\|\widehat{\boldsymbol{\gamma}}_k\|_2^2\\
\le&b_{k-1,3}\hbar^{-1}C_0\kappa^{-1}\mu_1(\boldsymbol{\Xi})^{1/2}\varpi\sqrt{s} + \mu_k(\boldsymbol{\Xi}),
\end{aligned}
$$

where the last inequality follows from (2.24). Combining (2.28) and (2.27) gives

$$
\begin{aligned}
&\widehat{\boldsymbol{\alpha}}_k^{\mathrm{T}}\widehat{\mathbf{B}}\widehat{\boldsymbol{\alpha}}_k \le \mu_k(\boldsymbol{\Xi}) + b_{k-1,3}\hbar^{-1}C_0\kappa^{-1}\mu_1(\boldsymbol{\Xi})^{1/2}\varpi\sqrt{s} + \sqrt{\mu_1(\boldsymbol{\Xi})}C_0\kappa^{-1}\varpi\sqrt{s}\\
&+ \frac{(1+\hbar^{-1})}{\sqrt{s}}\kappa\sqrt{\mu_1(\boldsymbol{\Xi})}C_0\varpi\|\widehat{\boldsymbol{\alpha}}_k\|_1^2\\
=&\mu_k(\boldsymbol{\Xi}) + (b_{k-1,3}\hbar^{-1} + 1)\sqrt{\mu_1(\boldsymbol{\Xi})}\kappa^{-1}C_0\varpi\sqrt{s}\\
&+ \frac{(1+\hbar^{-1})}{\sqrt{s}}\kappa\sqrt{\mu_1(\boldsymbol{\Xi})}C_0\varpi\|\widehat{\boldsymbol{\alpha}}_k\|_1^2
\end{aligned}
$$

The proof is completed.

<u>Proof of Lemma</u> 7.

Consider the left hand side of the first inequality in (1.28). We first note that by the definition of $\tau^{(k)}$,

$$\begin{aligned}\tau^{(k)}\|\widehat{\boldsymbol{\alpha}}_k\|^2_{\lambda^{(k)}} &= \frac{A^{(k)}C_0\varpi}{\|\boldsymbol{\alpha}_1\|_1\sqrt{\mu_1(\boldsymbol{\Xi})}}\|\widehat{\boldsymbol{\alpha}}_k\|^2_{\lambda^{(k)}} \\ \geq&\frac{A^{(k)}C_0\varpi}{\|\boldsymbol{\alpha}_1\|_1\sqrt{\mu_1(\boldsymbol{\Xi})}}\lambda^{(k)}\|\widehat{\boldsymbol{\alpha}}_k\|^2_1 \geq \kappa\frac{A^{(k)}C_0\varpi}{\sqrt{s\mu_1(\boldsymbol{\Xi})}}c^{-1}\|\widehat{\boldsymbol{\alpha}}_k\|^2_1,\end{aligned} \tag{2.29}$$

where the last inequality is due to $\lambda^{(k)} > c^{-1}$ and and $\|\boldsymbol{\alpha}_1\|_1 \leq \kappa^{-1}\sqrt{s}$ by Lemma 2. By the second inequality in Lemma 6 and (1.15), we have

$$\begin{aligned}&(\boldsymbol{\beta}_k^{\mathrm{T}}\widehat{\mathbf{B}}\boldsymbol{\beta}_k)\left(1+\tau^{(k)}\|\widehat{\boldsymbol{\alpha}}_k\|^2_{\lambda^{(k)}}\right) \\ \geq&\left(\mu_k(\boldsymbol{\Xi}) - N_{k,3}C_0\kappa^{-1}\sqrt{\mu_1(\boldsymbol{\Xi})}\varpi\sqrt{s}\right)\left(1+\kappa\frac{A^{(k)}C_0\varpi}{\sqrt{s\mu_1(\boldsymbol{\Xi})}}c^{-1}\|\widehat{\boldsymbol{\alpha}}_k\|^2_1\right) \\ =&\mu_k(\boldsymbol{\Xi}) - N_{k,3}C_0\kappa^{-1}\sqrt{\mu_1(\boldsymbol{\Xi})}\varpi\sqrt{s} \\ &+\left(\mu_k(\boldsymbol{\Xi}) - N_{k,3}\kappa^{-1}\sqrt{\mu_1(\boldsymbol{\Xi})}C_0\varpi\sqrt{s}\right)\kappa\frac{A^{(k)}C_0\varpi}{\sqrt{s\mu_1(\boldsymbol{\Xi})}}c^{-1}\|\widehat{\boldsymbol{\alpha}}_k\|^2_1 \\ \geq&\mu_k(\boldsymbol{\Xi}) - N_{k,3}C_0\kappa^{-1}\sqrt{\mu_1(\boldsymbol{\Xi})}\varpi\sqrt{s} \\ &+\left(c_3^{-1}\mu_1(\boldsymbol{\Xi}) - N_{k,3}\kappa^{-1}\sqrt{\mu_1(\boldsymbol{\Xi})}\hbar^{-1}\kappa\mu_1(\boldsymbol{\Xi})^{1/2}\right)\kappa\frac{A^{(k)}C_0\varpi}{\sqrt{s\mu_1(\boldsymbol{\Xi})}}c^{-1}\|\widehat{\boldsymbol{\alpha}}_k\|^2_1 \\ \geq&\mu_k(\boldsymbol{\Xi}) - N_{k,3}C_0\kappa^{-1}\sqrt{\mu_1(\boldsymbol{\Xi})}\varpi\sqrt{s} \\ &+\left(c_3^{-1} - N_{k,3}\hbar^{-1}\right)\sqrt{\mu_1(\boldsymbol{\Xi})}\kappa\frac{A^{(k)}C_0\varpi}{\sqrt{s}}c^{-1}\|\widehat{\boldsymbol{\alpha}}_k\|^2_1,\end{aligned} \tag{2.30}$$

where the inequality in the fifth line is due to $\mu_k(\boldsymbol{\Xi}) \geq c_3^{-1}\mu_1(\boldsymbol{\Xi})$ by Condition 3. Now we estimate the right hand side of the second inequality in (1.28). By the third inequality in Lemma 6 and the first inequality in the second line of

(1.26),

$$
\begin{aligned}
&(\widehat{\boldsymbol{\alpha}}_k^{\mathrm{T}}\widehat{\mathbf{B}}\widehat{\boldsymbol{\alpha}}_k)(1+\tau^{(k)}\|\boldsymbol{\beta}_k\|^2_{\lambda^{(k)}}) \le (\widehat{\boldsymbol{\alpha}}_k^{\mathrm{T}}\widehat{\mathbf{B}}\widehat{\boldsymbol{\alpha}}_k)(1+\tau^{(k)}\|\boldsymbol{\beta}_k\|_1^2)\\
\le&\left(\mu_k(\boldsymbol{\Xi})+(b_{k-1,3}\hbar^{-1}+1)\sqrt{\mu_1(\boldsymbol{\Xi})}\kappa^{-1}C_0\varpi\sqrt{s}+\frac{(1+\hbar^{-1})}{\sqrt{s}}\kappa\sqrt{\mu_1(\boldsymbol{\Xi})}C_0\varpi\|\widehat{\boldsymbol{\alpha}}_k\|_1^2\right)\\
&\times\left(1+A^{(k)}M_{k,1}^2\kappa^{-1}\frac{C_0\varpi\sqrt{s}}{\sqrt{\mu_1(\boldsymbol{\Xi})}}\right)=\mu_k(\boldsymbol{\Xi})+\mu_k(\boldsymbol{\Xi})A^{(k)}M_{k,1}^2\kappa^{-1}\frac{C_0\varpi\sqrt{s}}{\sqrt{\mu_1(\boldsymbol{\Xi})}}\\
&+\left(1+A^{(k)}M_{k,1}^2\kappa^{-1}\frac{C_0\varpi\sqrt{s}}{\sqrt{\mu_1(\boldsymbol{\Xi})}}\right)\\
&\times\left[(b_{k-1,3}\hbar^{-1}+1)\sqrt{\mu_1(\boldsymbol{\Xi})}\kappa^{-1}C_0\varpi\sqrt{s}+\frac{(1+\hbar^{-1})}{\sqrt{s}}\kappa\sqrt{\mu_1(\boldsymbol{\Xi})}C_0\varpi\|\widehat{\boldsymbol{\alpha}}_k\|_1^2\right]\\
\le&\mu_k(\boldsymbol{\Xi})+\mu_1(\boldsymbol{\Xi})A^{(k)}M_{k,1}^2\kappa^{-1}\frac{C_0\varpi\sqrt{s}}{\sqrt{\mu_1(\boldsymbol{\Xi})}}\\
&+\left(1+A^{(k)}M_{k,1}^2\hbar^{-1}\right)(b_{k-1,3}\hbar^{-1}+1)C_0\sqrt{\mu_1(\boldsymbol{\Xi})}\kappa^{-1}\varpi\sqrt{s}\\
&+\left(1+A^{(k)}M_{k,1}^2\hbar^{-1}\right)\frac{(1+\hbar^{-1})}{\sqrt{s}}\kappa\sqrt{\mu_1(\boldsymbol{\Xi})}C_0\varpi\|\widehat{\boldsymbol{\alpha}}_k\|_1^2\\
=&\mu_k(\boldsymbol{\Xi})+\left[A^{(k)}M_{k,1}^2+\left(1+A^{(k)}M_{k,1}^2\hbar^{-1}\right)(b_{k-1,3}\hbar^{-1}+1)\right]C_0\sqrt{\mu_1(\boldsymbol{\Xi})}\kappa^{-1}\varpi\sqrt{s}\\
&+\left(1+A^{(k)}M_{k,1}^2\hbar^{-1}\right)\frac{(1+\hbar^{-1})}{\sqrt{s}}\kappa\sqrt{\mu_1(\boldsymbol{\Xi})}C_0\varpi\|\widehat{\boldsymbol{\alpha}}_k\|_1^2,
\end{aligned}
$$

where the second inequality is due to the last inequality in the second line of (1.26). The above inequality, together with (1.28) and (2.30), leads to

$$
\begin{aligned}
&\left[\left(c_3^{-1}-N_{k,3}\hbar^{-1}\right)A^{(k)}c^{-1}-\left(1+A^{(k)}M_{k,1}^2\hbar^{-1}\right)(1+\hbar^{-1})\right]\\
&\times\sqrt{\mu_1(\boldsymbol{\Xi})}\kappa\frac{C_0\varpi}{\sqrt{s}}\|\widehat{\boldsymbol{\alpha}}_k\|_1^2\\
\le&\left[N_{k,3}+A^{(k)}M_{k,1}^2+\left(1+A^{(k)}M_{k,1}^2\hbar^{-1}\right)(b_{k-1,3}\hbar^{-1}+1)\right]\\
&\times\sqrt{\mu_1(\boldsymbol{\Xi})}\kappa^{-1}C_0\varpi\sqrt{s},
\end{aligned}
\tag{2.31}
$$

where we have canceled $\mu_k(\boldsymbol{\Xi})$ on both sides. Because $N_{k,3}=2(b_{k-1,3}\hbar^{-1}+M_{k,1})$ and $M_{k,1}$ does not depend on $\hbar$, we have

$$
\begin{aligned}
&\lim_{\substack{A^{(k)}\to\infty\\ \hbar\to\infty}}\sqrt{\frac{N_{k,3}+A^{(k)}M_{k,1}^2+\left(1+A^{(k)}M_{k,1}^2\hbar^{-1}\right)(b_{k-1,3}\hbar^{-1}+1)}{\left(c_3^{-1}-N_{k,3}\hbar^{-1}\right)A^{(k)}c^{-1}-\left(1+A^{(k)}M_{k,1}^2\hbar^{-1}\right)(1+\hbar^{-1})}}\\
=&M_{k,1}\sqrt{c_3c}.
\end{aligned}
$$

Therefore, there exist $(A_k^L)'$ and $(\hbar_0)'$ only depending on δ_0, c and $c_2 \sim c_5$ such that for any $A^{(k)} \geq (A_k^L)'$ and $\hbar \geq (\hbar_0)'$, we have

$$\|\widehat{\boldsymbol{\alpha}}_k\|_1 \leq D_{k,1}\kappa^{-1}\sqrt{s}.$$

where $D_{k,1} = 2M_{k,1}\sqrt{c_3 c}$.

Proof of Lemma 8.

We first estimate the left hand side of (1.30). Note that

$$\begin{aligned}
&\boldsymbol{\beta}_k^{\mathrm{T}}\widehat{\mathbf{B}}\boldsymbol{\beta}_k - \widehat{\boldsymbol{\alpha}}_k^{\mathrm{T}}\widehat{\mathbf{B}}\widehat{\boldsymbol{\alpha}}_k \\
\geq&\boldsymbol{\beta}_k^{\mathrm{T}}\mathbf{B}\boldsymbol{\beta}_k - \widehat{\boldsymbol{\alpha}}_k^{\mathrm{T}}\mathbf{B}\widehat{\boldsymbol{\alpha}}_k - |\boldsymbol{\beta}_k^{\mathrm{T}}(\widehat{\mathbf{B}}-\mathbf{B})\boldsymbol{\beta}_k - \widehat{\boldsymbol{\alpha}}_k^{\mathrm{T}}(\widehat{\mathbf{B}}-\mathbf{B})\widehat{\boldsymbol{\alpha}}_k|
\end{aligned} \tag{2.32}$$

By the first line of (2.28) and $\widehat{\mathbf{P}}_{k-1}\widehat{\boldsymbol{\gamma}}_k = \mathbf{0}$,

$$\widehat{\boldsymbol{\alpha}}_k^{\mathrm{T}}\mathbf{B}\widehat{\boldsymbol{\alpha}}_k = \sum_{i=1}^{K}\mu_i(\boldsymbol{\Xi})(\boldsymbol{\gamma}_i^{\mathrm{T}}\widehat{\boldsymbol{\gamma}}_k)^2 \leq \mu_1(\boldsymbol{\Xi})\sum_{i=1}^{k-1}(\boldsymbol{\gamma}_i^{\mathrm{T}}\widehat{\boldsymbol{\gamma}}_k)^2 + \sum_{i=k}^{K}\mu_i(\boldsymbol{\Xi})(\boldsymbol{\gamma}_i^{\mathrm{T}}\widehat{\boldsymbol{\gamma}}_k)^2 \tag{2.33}$$

$$\begin{aligned}
\leq&\mu_1(\boldsymbol{\Xi})\|\mathbf{P}_{k-1}\widehat{\boldsymbol{\gamma}}_k\|_2^2 + \mu_k(\boldsymbol{\Xi})(\boldsymbol{\gamma}_k^{\mathrm{T}}\widehat{\boldsymbol{\gamma}}_k)^2 + \mu_{k+1}(\boldsymbol{\Xi})\sum_{i=k+1}^{K}(\boldsymbol{\gamma}_i^{\mathrm{T}}\widehat{\boldsymbol{\gamma}}_k)^2 \\
=&\mu_1(\boldsymbol{\Xi})\|\mathbf{P}_{k-1}\widehat{\boldsymbol{\gamma}}_k - \widehat{\mathbf{P}}_{k-1}\widehat{\boldsymbol{\gamma}}_k\|_2^2 + \mu_k(\boldsymbol{\Xi})(\boldsymbol{\gamma}_k^{\mathrm{T}}\widehat{\boldsymbol{\gamma}}_k)^2 + \mu_{k+1}(\boldsymbol{\Xi})\sum_{i=k+1}^{K}(\boldsymbol{\gamma}_i^{\mathrm{T}}\widehat{\boldsymbol{\gamma}}_k)^2 \\
\leq&\mu_1(\boldsymbol{\Xi})\|\mathbf{P}_{k-1} - \widehat{\mathbf{P}}_{k-1}\|^2 + \mu_k(\boldsymbol{\Xi})(\boldsymbol{\gamma}_k^{\mathrm{T}}\widehat{\boldsymbol{\gamma}}_k)^2 + \mu_{k+1}(\boldsymbol{\Xi})\left[\|\widehat{\boldsymbol{\gamma}}_k\|_2^2 - (\boldsymbol{\gamma}_k^{\mathrm{T}}\widehat{\boldsymbol{\gamma}}_k)^2\right] \\
=&\mu_1(\boldsymbol{\Xi})\|\mathbf{P}_{k-1} - \widehat{\mathbf{P}}_{k-1}\|^2 + \mu_k(\boldsymbol{\Xi})(\boldsymbol{\gamma}_k^{\mathrm{T}}\widehat{\boldsymbol{\gamma}}_k)^2 + \mu_{k+1}(\boldsymbol{\Xi})\left[1 - (\boldsymbol{\gamma}_k^{\mathrm{T}}\widehat{\boldsymbol{\gamma}}_k)^2\right] \\
=&\mu_1(\boldsymbol{\Xi})\|\mathbf{P}_{k-1} - \widehat{\mathbf{P}}_{k-1}\|^2 + \mu_k(\boldsymbol{\Xi}) - (\mu_k(\boldsymbol{\Xi}) - \mu_{k+1}(\boldsymbol{\Xi}))\left[1 - (\boldsymbol{\gamma}_k^{\mathrm{T}}\widehat{\boldsymbol{\gamma}}_k)^2\right].
\end{aligned}$$

On the other hand, by $\mathbf{P}_{k-1}\boldsymbol{\gamma}_k = \mathbf{0}$ and the similar arguments as in (2.23),

$$\begin{aligned}
&\boldsymbol{\beta}_k^{\mathrm{T}}\mathbf{B}\boldsymbol{\beta}_k = \boldsymbol{\beta}_k^{\mathrm{T}}\mathbf{Z}^{\mathrm{T}}\boldsymbol{\Xi}\mathbf{Z}\boldsymbol{\beta}_k \geq \mu_k(\boldsymbol{\Xi}) - 2\mu_1(\boldsymbol{\Xi})\|\widehat{\mathbf{P}}_{k-1}\boldsymbol{\gamma}_k\|_2^2 \\
=&\mu_k(\boldsymbol{\Xi}) - 2\mu_k(\boldsymbol{\Xi})\|\widehat{\mathbf{P}}_{k-1}\boldsymbol{\gamma}_k - \mathbf{P}_{k-1}\boldsymbol{\gamma}_k\|_2^2 \\
=&\mu_k(\boldsymbol{\Xi}) - 2\mu_k(\boldsymbol{\Xi})\|\widehat{\mathbf{P}}_{k-1} - \mathbf{P}_{k-1}\|^2.
\end{aligned} \tag{2.34}$$

Combining (2.33) and (2.34) gives

$$
\begin{aligned}
(\boldsymbol{\beta}_k^{\mathrm{T}}\mathbf{B}\boldsymbol{\beta}_k - \widehat{\boldsymbol{\alpha}}_k^{\mathrm{T}}\mathbf{B}\widehat{\boldsymbol{\alpha}}_k) \geq& (\mu_k(\boldsymbol{\Xi}) - \mu_{k+1}(\boldsymbol{\Xi}))\left[1 - (\boldsymbol{\gamma}_k^{\mathrm{T}}\widehat{\boldsymbol{\gamma}}_k)^2\right] \\
&- (2\mu_k(\boldsymbol{\Xi}) + \mu_1(\boldsymbol{\Xi}))\|\widehat{\mathbf{P}}_{k-1} - \mathbf{P}_{k-1}\|^2 \\
\geq&(\mu_k(\boldsymbol{\Xi}) - \mu_{k+1}(\boldsymbol{\Xi}))\left[1 + (\boldsymbol{\gamma}_k^{\mathrm{T}}\widehat{\boldsymbol{\gamma}}_k)\right]\left[1 - (\boldsymbol{\gamma}_k^{\mathrm{T}}\widehat{\boldsymbol{\gamma}}_k)\right] - 3\mu_1(\boldsymbol{\Xi})\|\widehat{\mathbf{P}}_{k-1} - \mathbf{P}_{k-1}\|^2 \\
\geq& c_2\mu_k(\boldsymbol{\Xi})\left[1 - (\boldsymbol{\gamma}_k^{\mathrm{T}}\widehat{\boldsymbol{\gamma}}_k)\right] - 3\mu_1(\boldsymbol{\Xi})b_{k-1,3}C_0^2\kappa^{-2}\mu_1(\boldsymbol{\Xi})^{-1}\varpi^2 s \\
=&\frac{1}{2}c_2\mu_k(\boldsymbol{\Xi})\|\boldsymbol{\gamma}_k - \widehat{\boldsymbol{\gamma}}_k\|_2^2 - 3b_{k-1,3}C_0^2\kappa^{-2}\varpi^2 s,
\end{aligned}
\tag{2.35}
$$

where the inequality in the third line is due to Condition 3, (1.22) and $\boldsymbol{\gamma}_k^{\mathrm{T}}\widehat{\boldsymbol{\gamma}}_k \geq 0$. For the last term on the right hand side of (2.32), by the similar arguments as in (2.11), (2.14) and (2.15) in the proof of Lemma 3, we can obtain

$$
\begin{aligned}
&|\boldsymbol{\beta}_k^{\mathrm{T}}(\widehat{\mathbf{B}} - \mathbf{B})\boldsymbol{\beta}_k - \widehat{\boldsymbol{\alpha}}_k^{\mathrm{T}}(\widehat{\mathbf{B}} - \mathbf{B})\widehat{\boldsymbol{\alpha}}_k| \\
\leq&4\sqrt{\mu_1(\boldsymbol{\Xi})}C_0\varpi\|\boldsymbol{\beta}_k - \widehat{\boldsymbol{\alpha}}_k\|_1 + C_0^2\varpi^2\|\boldsymbol{\beta}_k - \widehat{\boldsymbol{\alpha}}_k\|_1(\|\boldsymbol{\beta}_k\|_1 + \|\widehat{\boldsymbol{\alpha}}_k\|_1) \\
\leq&4\sqrt{\mu_1(\boldsymbol{\Xi})}C_0\varpi\|\boldsymbol{\beta}_k - \widehat{\boldsymbol{\alpha}}_k\|_1 + C_0^2\varpi^2\|\boldsymbol{\beta}_k - \widehat{\boldsymbol{\alpha}}_k\|_1(M_{k,1} + D_{k,1})\kappa^{-1}\sqrt{s} \\
=&4\sqrt{\mu_1(\boldsymbol{\Xi})}C_0\varpi\|\boldsymbol{\beta}_k - \widehat{\boldsymbol{\alpha}}_k\|_1 + (M_{k,1} + D_{k,1})C_0\varpi\|\boldsymbol{\beta}_k - \widehat{\boldsymbol{\alpha}}_k\|_1\kappa^{-1}(C_0\varpi\sqrt{s}) \\
\leq&4\sqrt{\mu_1(\boldsymbol{\Xi})}C_0\varpi\|\boldsymbol{\beta}_k - \widehat{\boldsymbol{\alpha}}_k\|_1 + (M_{k,1} + D_{k,1})C_0\varpi\|\boldsymbol{\beta}_k - \widehat{\boldsymbol{\alpha}}_k\|_1\hbar^{-1}\mu_1(\boldsymbol{\Xi})^{1/2} \\
=&[4 + (M_{k,1} + D_{k,1})\hbar^{-1}]\sqrt{\mu_1(\boldsymbol{\Xi})}C_0\varpi\|\boldsymbol{\beta}_k - \widehat{\boldsymbol{\alpha}}_k\|_1,
\end{aligned}
\tag{2.36}
$$

where the third inequality follows from (1.24) and (1.29), and the fifth one is due to (1.15). By (1.28),

$$
1 + \tau^{(k)}\|\widehat{\boldsymbol{\alpha}}_k\|_{\lambda^{(k)}}^2 \leq \frac{\widehat{\boldsymbol{\alpha}}_k^{\mathrm{T}}\widehat{\mathbf{B}}\widehat{\boldsymbol{\alpha}}_k}{\boldsymbol{\beta}_k^{\mathrm{T}}\widehat{\mathbf{B}}\boldsymbol{\beta}_k}(1 + \tau^{(k)}\|\boldsymbol{\beta}_k\|_{\lambda^{(k)}}^2), \tag{2.37}
$$

By the second inequality in Lemma 6, we have

$$
\begin{aligned}
\boldsymbol{\beta}_k^{\mathrm{T}}\widehat{\mathbf{B}}\boldsymbol{\beta}_k &\geq \mu_k(\boldsymbol{\Xi}) - N_{k,3}\mu_1(\boldsymbol{\Xi})^{1/2}C_0\kappa^{-1}\varpi\sqrt{s} \\
&\geq \mu_k(\boldsymbol{\Xi}) - N_{k,3}\mu_1(\boldsymbol{\Xi})^{1/2}\hbar^{-1}\mu_1(\boldsymbol{\Xi})^{1/2} \\
&\geq c_3\mu_1(\boldsymbol{\Xi}) - N_{k,3}\hbar^{-1}\mu_1(\boldsymbol{\Xi}) = N_{k,7}\mu_1(\boldsymbol{\Xi}),
\end{aligned}
\tag{2.38}
$$

where the second inequality is due to (1.15) and $N_{k,7} = c_3 - N_{k,3}\hbar^{-1} = c_3 - 2(b_{k-1,3}\hbar^{-1} + M_{k,1})\hbar^{-1}$ (see the definition of $N_{k,3}$ in Lemma 6), and by the third inequality in Lemma 6,

$$
\begin{aligned}
\widehat{\boldsymbol{\alpha}}_k^{\mathrm{T}}\widehat{\mathbf{B}}\widehat{\boldsymbol{\alpha}}_k &\le \mu_k(\boldsymbol{\Xi}) + (b_{k-1,3}\hbar^{-1}+1)\sqrt{\mu_1(\boldsymbol{\Xi})}\kappa^{-1}C_0\varpi\sqrt{s} \\
&+ \frac{(1+\hbar^{-1})}{\sqrt{s}}\kappa\sqrt{\mu_1(\boldsymbol{\Xi})}C_0\varpi\|\widehat{\boldsymbol{\alpha}}_k\|_1^2 \\
&\le \mu_1(\boldsymbol{\Xi}) + (b_{k-1,3}\hbar^{-1}+1)\sqrt{\mu_1(\boldsymbol{\Xi})}\kappa^{-1}C_0\varpi\sqrt{s} \\
&+ \frac{(1+\hbar^{-1})}{\sqrt{s}}\kappa\sqrt{\mu_1(\boldsymbol{\Xi})}C_0\varpi D_{k,1}^2\kappa^{-2}s \\
&= \mu_1(\boldsymbol{\Xi}) + [(b_{k-1,3}\hbar^{-1}+1) + (1+\hbar^{-1})D_{k,1}^2]\sqrt{\mu_1(\boldsymbol{\Xi})}\kappa^{-1}C_0\varpi\sqrt{s} \\
&\le \mu_1(\boldsymbol{\Xi}) + [(b_{k-1,3}\hbar^{-1}+1) + (1+\hbar^{-1})D_{k,1}^2]\sqrt{\mu_1(\boldsymbol{\Xi})}\hbar^{-1}\sqrt{\mu_1(\boldsymbol{\Xi})} \\
&= N_{k,6}\mu_1(\boldsymbol{\Xi}), \qquad (2.39)
\end{aligned}
$$

where the second inequality is due to (1.29) and $N_{k,6} = 1+(b_{k-1,3}\hbar^{-1}+1)\hbar^{-1}+(1+\hbar^{-1})D_{k,1}^2\hbar^{-1}$. Now (2.37) − (2.39) and (1.26) give

$$
\begin{aligned}
1+\tau^{(k)}\|\widehat{\boldsymbol{\alpha}}_k\|_{\lambda^{(k)}}^2 &\le \frac{\widehat{\boldsymbol{\alpha}}_k^{\mathrm{T}}\widehat{\mathbf{B}}\widehat{\boldsymbol{\alpha}}_k}{\boldsymbol{\beta}_k^{\mathrm{T}}\widehat{\mathbf{B}}\boldsymbol{\beta}_k}(1+\tau^{(k)}\|\boldsymbol{\beta}_k\|_{\lambda^{(k)}}^2) \\
&\le N_{k,6}N_{k,7}^{-1}(1+A^{(k)}M_{k,1}^2\hbar^{-1}). \qquad (2.40)
\end{aligned}
$$

Therefore, by (2.32)-(2.36) and (2.40),

$$
\begin{aligned}
&(\boldsymbol{\beta}_k^{\mathrm{T}}\widehat{\mathbf{B}}\boldsymbol{\beta}_k - \widehat{\boldsymbol{\alpha}}_k^{\mathrm{T}}\widehat{\mathbf{B}}\widehat{\boldsymbol{\alpha}}_k)(1+\tau^{(k)}\|\widehat{\boldsymbol{\alpha}}_k\|_{\lambda^{(k)}}^2) \qquad (2.41) \\
\ge &\left(\frac{1}{2}c_2\mu_k(\boldsymbol{\Xi})\|\boldsymbol{\gamma}_k-\widehat{\boldsymbol{\gamma}}_k\|_2^2 - 3b_{k-1,3}C_0^2\kappa^{-2}\varpi^2 s\right. \\
&\left.-[4+(M_{k,1}+D_{k,1})\hbar^{-1}]\sqrt{\mu_1(\boldsymbol{\Xi})}C_0\varpi\|\boldsymbol{\beta}_k-\widehat{\boldsymbol{\alpha}}_k\|_1\right)(1+\tau^{(k)}\|\widehat{\boldsymbol{\alpha}}_k\|_{\lambda^{(k)}}^2) \\
\ge &\frac{1}{2}c_2\mu_k(\boldsymbol{\Xi})\|\boldsymbol{\gamma}_k-\widehat{\boldsymbol{\gamma}}_k\|_2^2 - N_{k,4}C_0^2\kappa^{-2}\varpi^2 s - N_{k,5}C_0\sqrt{\mu_1(\boldsymbol{\Xi})}\varpi\|\boldsymbol{\beta}_k-\widehat{\boldsymbol{\alpha}}_k\|_1,
\end{aligned}
$$

where

$$
\begin{aligned}
N_{k,4} &= 3b_{k-1,3}N_{k,6}N_{k,7}^{-1}(1+A^{(k)}M_{k,1}^2\hbar^{-1}), \\
N_{k,5} &= [4+(M_{k,1}+D_{k,1})\hbar^{-1}]N_{k,6}N_{k,7}^{-1}(1+A^{(k)}M_{k,1}^2\hbar^{-1}).
\end{aligned}
$$

Therefore, by (1.30) and (2.41), we obtain

$$
\begin{aligned}
&\frac{1}{2}c_2\mu_k(\boldsymbol{\Xi})\|\boldsymbol{\gamma}_k-\widehat{\boldsymbol{\gamma}}_k\|_2^2 \\
\le &N_{k,4}C_0^2\kappa^{-2}\varpi^2 s + N_{k,5}C_0\sqrt{\mu_1(\boldsymbol{\Xi})}\varpi\|\boldsymbol{\beta}_k-\widehat{\boldsymbol{\alpha}}_k\|_1 \\
&+(\widehat{\boldsymbol{\alpha}}_k^{\mathrm{T}}\widehat{\mathbf{B}}\widehat{\boldsymbol{\alpha}}_k)[\tau^{(k)}\|\boldsymbol{\beta}_k\|_{\lambda^{(k)}}^2 - \tau^{(k)}\|\widehat{\boldsymbol{\alpha}}_k\|_{\lambda^{(k)}}^2]. \qquad (2.42)
\end{aligned}
$$

To estimate the term $(\widehat{\boldsymbol{\alpha}}_k^{\mathrm{T}}\widehat{\mathbf{B}}\widehat{\boldsymbol{\alpha}}_k)[\tau^{(k)}\|\boldsymbol{\beta}_k\|^2_{\lambda^{(k)}} - \tau^{(k)}\|\widehat{\boldsymbol{\alpha}}_k\|^2_{\lambda^{(k)}}]$ in (2.42), we consider the following three possible situations, separately:

- **Situation 1:** $\boldsymbol{\beta}_k^{\mathrm{T}}\widehat{\mathbf{B}}\boldsymbol{\beta}_k \geq \widehat{\boldsymbol{\alpha}}_k^{\mathrm{T}}\widehat{\mathbf{B}}\widehat{\boldsymbol{\alpha}}_k$.

  In this case, by (1.30) and (2.39), we have

$$
\begin{aligned}
0 &\leq (\boldsymbol{\beta}_k^{\mathrm{T}}\widehat{\mathbf{B}}\boldsymbol{\beta}_k - \widehat{\boldsymbol{\alpha}}_k^{\mathrm{T}}\widehat{\mathbf{B}}\widehat{\boldsymbol{\alpha}}_k)(1 + \tau^{(k)}\|\widehat{\boldsymbol{\alpha}}_k\|^2_{\lambda^{(k)}}) \\
&\leq (\widehat{\boldsymbol{\alpha}}_k^{\mathrm{T}}\widehat{\mathbf{B}}\widehat{\boldsymbol{\alpha}}_k)[\tau^{(k)}\|\boldsymbol{\beta}_k\|^2_{\lambda^{(k)}} - \tau^{(k)}\|\widehat{\boldsymbol{\alpha}}_k\|^2_{\lambda^{(k)}}] \\
&\leq N_{k,6}\mu_1(\boldsymbol{\Xi})[\tau^{(k)}\|\boldsymbol{\beta}_k\|^2_{\lambda^{(k)}} - \tau^{(k)}\|\widehat{\boldsymbol{\alpha}}_k\|^2_{\lambda^{(k)}}].
\end{aligned}
$$

- **Situation 2:** $\boldsymbol{\beta}_k^{\mathrm{T}}\widehat{\mathbf{B}}\boldsymbol{\beta}_k < \widehat{\boldsymbol{\alpha}}_k^{\mathrm{T}}\widehat{\mathbf{B}}\widehat{\boldsymbol{\alpha}}_k$ and $\|\boldsymbol{\beta}_k\|^2_{\lambda^{(k)}} < \|\widehat{\boldsymbol{\alpha}}_k\|^2_{\lambda^{(k)}}$.

  In this case, by (2.38), we have

$$
\begin{aligned}
&(\widehat{\boldsymbol{\alpha}}_k^{\mathrm{T}}\widehat{\mathbf{B}}\widehat{\boldsymbol{\alpha}}_k)[\tau^{(k)}\|\boldsymbol{\beta}_k\|^2_{\lambda^{(k)}} - \tau^{(k)}\|\widehat{\boldsymbol{\alpha}}_k\|^2_{\lambda^{(k)}}] \\
<&(\boldsymbol{\beta}_k^{\mathrm{T}}\widehat{\mathbf{B}}\boldsymbol{\beta}_k)[\tau^{(k)}\|\boldsymbol{\beta}_k\|^2_{\lambda^{(k)}} - \tau^{(k)}\|\widehat{\boldsymbol{\alpha}}_k\|^2_{\lambda^{(k)}}] \\
\leq&N_{k,7}\mu_1(\boldsymbol{\Xi})[\tau^{(k)}\|\boldsymbol{\beta}_k\|^2_{\lambda^{(k)}} - \tau^{(k)}\|\widehat{\boldsymbol{\alpha}}_k\|^2_{\lambda^{(k)}}].
\end{aligned}
$$

- **Situation 3:** $\boldsymbol{\beta}_k^{\mathrm{T}}\widehat{\mathbf{B}}\boldsymbol{\beta}_k < \widehat{\boldsymbol{\alpha}}_k^{\mathrm{T}}\widehat{\mathbf{B}}\widehat{\boldsymbol{\alpha}}_k$ and $\|\boldsymbol{\beta}_k\|^2_{\lambda^{(k)}} \geq \|\widehat{\boldsymbol{\alpha}}_k\|^2_{\lambda^{(k)}}$.

  In this case, by (2.39), we have

$$
\begin{aligned}
&(\widehat{\boldsymbol{\alpha}}_k^{\mathrm{T}}\widehat{\mathbf{B}}\widehat{\boldsymbol{\alpha}}_k)[\tau^{(k)}\|\boldsymbol{\beta}_k\|^2_{\lambda^{(k)}} - \tau^{(k)}\|\widehat{\boldsymbol{\alpha}}_k\|^2_{\lambda^{(k)}}] \\
\leq&N_{k,6}\mu_1(\boldsymbol{\Xi})[\tau^{(k)}\|\boldsymbol{\beta}_k\|^2_{\lambda^{(k)}} - \tau^{(k)}\|\widehat{\boldsymbol{\alpha}}_k\|^2_{\lambda^{(k)}}].
\end{aligned}
$$

Hence, in all the three situations, we have either

$$
\begin{aligned}
&(\widehat{\boldsymbol{\alpha}}_k^{\mathrm{T}}\widehat{\mathbf{B}}\widehat{\boldsymbol{\alpha}}_k)[\tau^{(k)}\|\boldsymbol{\beta}_k\|^2_{\lambda^{(k)}} - \tau^{(k)}\|\widehat{\boldsymbol{\alpha}}_k\|^2_{\lambda^{(k)}}] \\
\leq&N_{k,6}\mu_1(\boldsymbol{\Xi})[\tau^{(k)}\|\boldsymbol{\beta}_k\|^2_{\lambda^{(k)}} - \tau^{(k)}\|\widehat{\boldsymbol{\alpha}}_k\|^2_{\lambda^{(k)}}].
\end{aligned}
$$

or

$$
\begin{aligned}
&(\widehat{\boldsymbol{\alpha}}_k^{\mathrm{T}}\widehat{\mathbf{B}}\widehat{\boldsymbol{\alpha}}_k)[\tau^{(k)}\|\boldsymbol{\beta}_k\|^2_{\lambda^{(k)}} - \tau^{(k)}\|\widehat{\boldsymbol{\alpha}}_k\|^2_{\lambda^{(k)}}] \\
\leq&N_{k,7}\mu_1(\boldsymbol{\Xi})[\tau^{(k)}\|\boldsymbol{\beta}_k\|^2_{\lambda^{(k)}} - \tau^{(k)}\|\widehat{\boldsymbol{\alpha}}_k\|^2_{\lambda^{(k)}}],
\end{aligned}
$$

which together with (2.42) lead to the conclusion.

Proof of Lemma 9.

We first consider the second term on the right hand side of (1.31). Recall that $\boldsymbol{\beta}_k = \boldsymbol{\alpha}_k - \boldsymbol{\delta}_k$. By (1.23),

$$\begin{aligned} & N_{k,5} C_0 \sqrt{\mu_1(\boldsymbol{\Xi})} \varpi \|\boldsymbol{\beta}_k - \widehat{\boldsymbol{\alpha}}_k\|_1 = N_{k,5} C_0 \sqrt{\mu_1(\boldsymbol{\Xi})} \varpi \|\boldsymbol{\alpha}_k - \boldsymbol{\delta}_k - \widehat{\boldsymbol{\alpha}}_k\|_1 \\ \le & N_{k,5} C_0 \sqrt{\mu_1(\boldsymbol{\Xi})} \varpi \|\boldsymbol{\alpha}_k - \widehat{\boldsymbol{\alpha}}_k\|_1 + N_{k,5} C_0 \sqrt{\mu_1(\boldsymbol{\Xi})} \varpi \|\boldsymbol{\delta}_k\|_1 \\ \le & N_{k,5} C_0 \sqrt{\mu_1(\boldsymbol{\Xi})} \varpi \|\boldsymbol{\alpha}_k - \widehat{\boldsymbol{\alpha}}_k\|_1 + N_{k,5} C_0 \sqrt{\mu_1(\boldsymbol{\Xi})} \varpi M_{k,0} C_0 \kappa^{-2} \mu_1(\boldsymbol{\Xi})^{-1/2} \varpi s \\ = & N_{k,5} C_0 \sqrt{\mu_1(\boldsymbol{\Xi})} \varpi \|\boldsymbol{\alpha}_k - \widehat{\boldsymbol{\alpha}}_k\|_1 + N_{k,5} M_{k,0} C_0^2 \kappa^{-2} \varpi^2 s. \end{aligned} \tag{2.43}$$

Now we consider the last term on the right hand side of (1.31). By (1.24),

$$\begin{aligned} & \|\boldsymbol{\beta}_k\|_2^2 \le (\|\boldsymbol{\alpha}_k\|_2 + \|\boldsymbol{\delta}_k\|_2)^2 \le (\|\boldsymbol{\alpha}_k\|_2 + \|\boldsymbol{\delta}_k\|_1)^2 \\ \le & \|\boldsymbol{\alpha}_k\|_2^2 + 2\|\boldsymbol{\alpha}_k\|_1 \|\boldsymbol{\delta}_k\|_1 + \|\boldsymbol{\delta}_k\|_1^2 \\ \le & \|\boldsymbol{\alpha}_k\|_2^2 + 2(\|\boldsymbol{\alpha}_k\|_1 + \|\boldsymbol{\delta}_k\|_1)\|\boldsymbol{\delta}_k\|_1 \\ \le & \|\boldsymbol{\alpha}_k\|_2^2 + 2M_{k,1}\|\boldsymbol{\alpha}_1\|_1 M_{k,0} \kappa^{-2} \mu_1(\boldsymbol{\Xi})^{-1/2} C_0 \varpi s \end{aligned} \tag{2.44}$$

Similarly,

$$\begin{aligned} & \|\boldsymbol{\beta}_k\|_1^2 \le (\|\boldsymbol{\alpha}_k\|_1 + \|\boldsymbol{\delta}_k\|_1)^2 = \|\boldsymbol{\alpha}_k\|_1^2 + 2\|\boldsymbol{\alpha}_k\|_1 \|\boldsymbol{\delta}_k\|_1 + \|\boldsymbol{\delta}_k\|_1^2 \\ \le & \|\boldsymbol{\alpha}_k\|_1^2 + 2M_{k,1}\|\boldsymbol{\alpha}_1\|_1 M_{k,0} \kappa^{-2} \mu_1(\boldsymbol{\Xi})^{-1/2} C_0 \varpi s. \end{aligned} \tag{2.45}$$

Combining (2.44) and (2.45) gives

$$\begin{aligned} & \tau^{(k)} \|\boldsymbol{\beta}_k\|_{\lambda^{(k)}}^2 = \tau^{(k)}(1 - \lambda^{(k)}) \|\boldsymbol{\beta}_k\|_2^2 + \tau^{(k)} \lambda^{(k)} \|\boldsymbol{\beta}_k\|_1^2 \\ \le & \tau^{(k)}(1 - \lambda^{(k)}) \left[\|\boldsymbol{\alpha}_k\|_2^2 + 2M_{k,1}\|\boldsymbol{\alpha}_1\|_1 M_{k,0} \kappa^{-2} \mu_1(\boldsymbol{\Xi})^{-1/2} C_0 \varpi s \right] \\ & + \tau^{(k)} \lambda^{(k)} \left[\|\boldsymbol{\alpha}_k\|_1^2 + 2M_{k,1}\|\boldsymbol{\alpha}_1\|_1 M_{k,0} \kappa^{-2} \mu_1(\boldsymbol{\Xi})^{-1/2} C_0 \varpi s \right] \\ = & \tau^{(k)} \|\boldsymbol{\alpha}_k\|_{\lambda^{(k)}}^2 + \tau^{(k)} 2M_{k,1}\|\boldsymbol{\alpha}_1\|_1 M_{k,0} \kappa^{-2} \mu_1(\boldsymbol{\Xi})^{-1/2} C_0 \varpi s \\ = & \tau^{(k)} \|\boldsymbol{\alpha}_k\|_{\lambda^{(k)}}^2 + \frac{A^{(k)} C_0 \varpi}{\|\boldsymbol{\alpha}_1\|_1 \sqrt{\mu_1(\boldsymbol{\Xi})}} 2M_{k,1}\|\boldsymbol{\alpha}_1\|_1 M_{k,0} \kappa^{-2} \mu_1(\boldsymbol{\Xi})^{-1/2} C_0 \varpi s \\ = & \tau^{(k)} \|\boldsymbol{\alpha}_k\|_{\lambda^{(k)}}^2 + 2A^{(k)} M_{k,1} M_{k,0} \kappa^{-2} \mu_1(\boldsymbol{\Xi})^{-1} C_0^2 \varpi^2 s \end{aligned}$$

which together with (2.39) lead to

$$
\begin{aligned}
& N_{k,6}\mu_1(\boldsymbol{\Xi})\tau^{(k)}\|\boldsymbol{\beta}_k\|^2_{\lambda^{(k)}} \qquad (2.46)\\
\le & N_{k,6}\mu_1(\boldsymbol{\Xi})\tau^{(k)}\|\boldsymbol{\alpha}_k\|^2_{\lambda^{(k)}} + N_{k,6}\mu_1(\boldsymbol{\Xi})2A^{(k)}M_{k,1}M_{k,0}\kappa^{-2}\mu_1(\boldsymbol{\Xi})^{-1}C_0^2\varpi^2 s\\
\le & N_{k,6}\mu_1(\boldsymbol{\Xi})\tau^{(k)}\|\boldsymbol{\alpha}_k\|^2_{\lambda^{(k)}} + 2A^{(k)}N_{k,6}M_{k,1}M_{k,0}\kappa^{-2}C_0^2\varpi^2 s
\end{aligned}
$$

By (1.31) and (2.43)-(2.46),

$$
\begin{aligned}
& \frac{1}{2}c_2\mu_k(\boldsymbol{\Xi})\|\boldsymbol{\gamma}_k - \widehat{\boldsymbol{\gamma}}_k\|_2^2 \qquad (2.47)\\
\le & N_{k,4}C_0^2\kappa^{-2}\varpi^2 s + N_{k,5}C_0\sqrt{\mu_1(\boldsymbol{\Xi})}\varpi\|\boldsymbol{\beta}_k - \widehat{\boldsymbol{\alpha}}_k\|_1\\
& + N_{k,6}\mu_1(\boldsymbol{\Xi})[\tau^{(k)}\|\boldsymbol{\beta}_k\|^2_{\lambda^{(k)}} - \tau^{(k)}\|\widehat{\boldsymbol{\alpha}}_k\|^2_{\lambda^{(k)}}]\\
\le & N_{k,4}C_0^2\kappa^{-2}\varpi^2 s + N_{k,5}C_0\sqrt{\mu_1(\boldsymbol{\Xi})}\varpi\|\boldsymbol{\alpha}_k - \widehat{\boldsymbol{\alpha}}_k\|_1 + N_{k,5}M_{k,0}C_0^2\kappa^{-2}\varpi^2 s\\
& + 2A^{(k)}N_{k,6}M_{k,1}M_{k,0}\kappa^{-2}C_0^2\varpi^2 s + N_{k,6}\mu_1(\boldsymbol{\Xi})[\tau^{(k)}\|\boldsymbol{\alpha}_k\|^2_{\lambda^{(k)}} - \tau^{(k)}\|\widehat{\boldsymbol{\alpha}}_k\|^2_{\lambda^{(k)}}]\\
= & N_{k,1}C_0^2\kappa^{-2}\varpi^2 s + N_{k,5}C_0\sqrt{\mu_1(\boldsymbol{\Xi})}\varpi\|\boldsymbol{\alpha}_k - \widehat{\boldsymbol{\alpha}}_k\|_1\\
& + N_{k,6}\mu_1(\boldsymbol{\Xi})[\tau^{(k)}\|\boldsymbol{\alpha}_k\|^2_{\lambda^{(k)}} - \tau^{(k)}\|\widehat{\boldsymbol{\alpha}}_k\|^2_{\lambda^{(k)}}],
\end{aligned}
$$

where $N_{k,1} = N_{k,4} + N_{k,5}M_{k,0} + 2A^{(k)}N_{k,6}M_{k,1}M_{k,0}$. Next, we calculate

$$
\begin{aligned}
& \tau^{(k)}\|\boldsymbol{\alpha}_k\|^2_{\lambda^{(k)}} - \tau^{(k)}\|\widehat{\boldsymbol{\alpha}}_k\|^2_{\lambda^{(k)}}\\
= & \frac{A^{(k)}C_0\varpi}{\|\boldsymbol{\alpha}_1\|_1\mu_1(\boldsymbol{\Xi})^{1/2}}\left[(1-\lambda^{(k)})(\|\boldsymbol{\alpha}_k\|_2^2 - \|\widehat{\boldsymbol{\alpha}}_k\|_2^2) + \lambda^{(k)}(\|\boldsymbol{\alpha}_k\|_1^2 - \|\widehat{\boldsymbol{\alpha}}_k\|_1^2)\right].
\end{aligned}
$$

We have

$$\begin{aligned}
&\frac{A^{(k)}C_0\varpi}{\|\boldsymbol{\alpha}_1\|_1\mu_1(\boldsymbol{\Xi})^{1/2}}(1-\lambda^{(k)})(\|\boldsymbol{\alpha}_k\|_2^2-\|\widehat{\boldsymbol{\alpha}}_k\|_2^2) \qquad (2.48)\\
=&\frac{A^{(k)}C_0\varpi}{\|\boldsymbol{\alpha}_1\|_1\mu_1(\boldsymbol{\Xi})^{1/2}}(1-\lambda^{(k)})(\|(\boldsymbol{\alpha}_k)_{J_k}\|_2^2-\|(\widehat{\boldsymbol{\alpha}}_k)_{J_k}\|_2^2-\|(\widehat{\boldsymbol{\alpha}}_k)_{J_k^c}\|_2^2)\\
\le&\frac{A^{(k)}C_0\varpi}{\|\boldsymbol{\alpha}_1\|_1\mu_1(\boldsymbol{\Xi})^{1/2}}(1-\lambda^{(k)})(\|(\boldsymbol{\alpha}_k)_{J_k}\|_2^2-\|(\widehat{\boldsymbol{\alpha}}_k)_{J_k}\|_2^2)\\
=&\frac{A^{(k)}C_0\varpi}{\|\boldsymbol{\alpha}_1\|_1\mu_1(\boldsymbol{\Xi})^{1/2}}(1-\lambda^{(k)})(\|(\boldsymbol{\alpha}_k)_{J_k}\|_2-\|(\widehat{\boldsymbol{\alpha}}_k)_{J_k}\|_2)(\|(\boldsymbol{\alpha}_k)_{J_k}\|_2+\|(\widehat{\boldsymbol{\alpha}}_k)_{J_k}\|_2)\\
=&\frac{A^{(k)}C_0\varpi}{\|\boldsymbol{\alpha}_1\|_1\mu_1(\boldsymbol{\Xi})^{1/2}}(1-\lambda^{(k)})[(\|(\boldsymbol{\alpha}_k)_{J_k}\|_2-\|(\widehat{\boldsymbol{\alpha}}_k)_{J_k}\|_2)(2\|(\boldsymbol{\alpha}_k)_{J_k}\|_2)\\
&\qquad-(\|(\boldsymbol{\alpha}_k)_{J_k}\|_2-\|(\widehat{\boldsymbol{\alpha}}_k)_{J_k}\|_2)^2]\\
\le&\frac{A^{(k)}C_0\varpi}{\|\boldsymbol{\alpha}_1\|_1\mu_1(\boldsymbol{\Xi})^{1/2}}(1-\lambda^{(k)})\,[(\|(\boldsymbol{\alpha}_k)_{J_k}\|_2-\|(\widehat{\boldsymbol{\alpha}}_k)_{J_k}\|_2)(2\|(\boldsymbol{\alpha}_k)_{J_k}\|_2)]\\
\le&A^{(k)}C_0\mu_1(\boldsymbol{\Xi})^{-1/2}\varpi(1-\lambda^{(k)})\|(\widehat{\boldsymbol{\alpha}}_k-\boldsymbol{\alpha}_k)_{J_k}\|_2(2\|(\boldsymbol{\alpha}_k)_{J_k}\|_2/\|\boldsymbol{\alpha}_1\|_1)\\
\le&A^{(k)}C_0\mu_1(\boldsymbol{\Xi})^{-1/2}\varpi(1-\lambda^{(k)})\|(\widehat{\boldsymbol{\alpha}}_k-\boldsymbol{\alpha}_k)_{J_k}\|_1(2\|\boldsymbol{\alpha}_k\|_1/\|\boldsymbol{\alpha}_1\|_1)\\
=&A^{(k)}C_0\mu_1(\boldsymbol{\Xi})^{-1/2}\varpi(1-\lambda^{(k)})\|(\widehat{\boldsymbol{\alpha}}_k-\boldsymbol{\alpha}_k)_{J_k}\|_1(2\|\boldsymbol{\alpha}_k\|_1/\|\boldsymbol{\alpha}_1\|_1),
\end{aligned}$$

and

$$\begin{aligned}
&\frac{A^{(k)}C_0\varpi}{\|\boldsymbol{\alpha}_1\|_1\mu_1(\boldsymbol{\Xi})^{1/2}}\lambda^{(k)}(\|\boldsymbol{\alpha}_k\|_1^2-\|\widehat{\boldsymbol{\alpha}}_k\|_1^2) \qquad (2.49)\\
=&\frac{A^{(k)}C_0\varpi}{\|\boldsymbol{\alpha}_1\|_1\mu_1(\boldsymbol{\Xi})^{1/2}}\lambda^{(k)}(\|\boldsymbol{\alpha}_k\|_1-\|\widehat{\boldsymbol{\alpha}}_k\|_1)(\|\boldsymbol{\alpha}_k\|_1+\|\widehat{\boldsymbol{\alpha}}_k\|_1)\\
=&\frac{A^{(k)}C_0\varpi}{\|\boldsymbol{\alpha}_1\|_1\mu_1(\boldsymbol{\Xi})^{1/2}}\lambda^{(k)}\left[(\|\boldsymbol{\alpha}_k\|_1-\|\widehat{\boldsymbol{\alpha}}_k\|_1)(2\|\boldsymbol{\alpha}_k\|_1)-(\|\boldsymbol{\alpha}_k\|_1-\|\widehat{\boldsymbol{\alpha}}_k\|_1)^2\right]\\
\le&\frac{A^{(k)}C_0\varpi}{\|\boldsymbol{\alpha}_1\|_1\mu_1(\boldsymbol{\Xi})^{1/2}}\lambda^{(k)}\,[(\|\boldsymbol{\alpha}_k\|_1-\|\widehat{\boldsymbol{\alpha}}_k\|_1)(2\|\boldsymbol{\alpha}_k\|_1)]\\
=&A^{(k)}C_0\mu_1(\boldsymbol{\Xi})^{-1/2}\varpi\lambda^{(k)}(\|(\boldsymbol{\alpha}_k)_{J_k}\|_1-\|(\widehat{\boldsymbol{\alpha}}_k)_{J_k}\|_1-\|(\widehat{\boldsymbol{\alpha}}_k)_{J_k^c}\|_1)(2\|\boldsymbol{\alpha}_k\|_1/\|\boldsymbol{\alpha}_1\|_1)\\
\le&A^{(k)}C_0\mu_1(\boldsymbol{\Xi})^{-1/2}\varpi\lambda^{(k)}(\|(\boldsymbol{\alpha}_k-\widehat{\boldsymbol{\alpha}}_k)_{J_k}\|_1-\|(\widehat{\boldsymbol{\alpha}}_k)_{J_k^c}\|_1)(2\|\boldsymbol{\alpha}_k\|_1/\|\boldsymbol{\alpha}_1\|_1).
\end{aligned}$$

Therefore, combining (2.48) and (2.49), we obtain

$$\begin{aligned}
&\tau^{(k)}\|\boldsymbol{\alpha}_k\|_{\lambda^{(k)}}^2-\tau^{(k)}\|\widehat{\boldsymbol{\alpha}}_k\|_{\lambda^{(k)}}^2\\
\le&A^{(k)}C_0\mu_1(\boldsymbol{\Xi})^{-1/2}\varpi\left[\|(\boldsymbol{\alpha}_k-\widehat{\boldsymbol{\alpha}}_k)_{J_k}\|_1-\lambda^{(k)}\|(\widehat{\boldsymbol{\alpha}}_k)_{J_k^c}\|_1\right](2\|\boldsymbol{\alpha}_k\|_1/\|\boldsymbol{\alpha}_1\|_1),
\end{aligned}$$

which leads to

$$
\begin{aligned}
&N_{k,6}\mu_1(\boldsymbol{\Xi})[\tau^{(k)}\|\boldsymbol{\alpha}_k\|^2_{\lambda^{(k)}} - \tau^{(k)}\|\widehat{\boldsymbol{\alpha}}_k\|^2_{\lambda^{(k)}}] \\
\leq &N_{k,2}\mu_1(\boldsymbol{\Xi})^{1/2}C_0\varpi\left[\|(\boldsymbol{\alpha}_k - \widehat{\boldsymbol{\alpha}}_k)_{J_k}\|_1 - \lambda^{(k)}\|(\widehat{\boldsymbol{\alpha}}_k)_{J_k^c}\|_1\right],
\end{aligned} \tag{2.50}
$$

where $N_{k,2} = A^{(k)}N_{k,6}(2\|\boldsymbol{\alpha}_k\|_1/\|\boldsymbol{\alpha}_1\|_1)$. By (2.47), (2.50) and noting that $\|\boldsymbol{\alpha}_k - \widehat{\boldsymbol{\alpha}}_k\|_1 = \|(\boldsymbol{\alpha}_k - \widehat{\boldsymbol{\alpha}}_k)_{J_k}\|_1 + \|(\widehat{\boldsymbol{\alpha}}_k)_{J_k^c}\|_1$, we have

$$
\begin{aligned}
&\frac{1}{2}c_2\mu_k(\boldsymbol{\Xi})\|\boldsymbol{\gamma}_k - \widehat{\boldsymbol{\gamma}}_k\|_2^2 \\
\leq &N_{k,1}C_0^2\kappa^{-2}\varpi^2 s + (N_{k,2} + N_{k,5})\mu_1(\boldsymbol{\Xi})^{1/2}C_0\varpi\|(\boldsymbol{\alpha}_k - \widehat{\boldsymbol{\alpha}}_k)_{J_k}\|_1 \\
&- (\lambda^{(k)}N_{k,2} - N_{k,5})\mu_1(\boldsymbol{\Xi})^{1/2}C_0\varpi\|(\widehat{\boldsymbol{\alpha}}_k)_{J_k^c}\|_1,
\end{aligned}
$$

3. Appendix C: Additional simulation

This is an additional simulation study when there is only one functional predictor. Both $z(s)$ and $\beta(t,s)$ are generated from general Gaussian processes. We consider three types of Gaussian processes with different smoothness levels and the following different covariance functions:

$$\mathbf{\Sigma}_1(s,s') = e^{-\{10|s-s'|\}^2}, \quad \mathbf{\Sigma}_2(s,s') = \left\{1 + 20|s-s'| + \frac{20}{3}(s-s')^2\right\} e^{-20|s-s'|},$$
$$\mathbf{\Sigma}_3(s,s') = e^{-\{10|s-s'|\}^{1.5}}, \tag{3.1}$$

The first one in (3.1) is the squared exponential covariance function and the corresponding Gaussian process has mean square derivatives of all orders (Chapter 4 in Rasmussen and Williams [2]). The second one belongs to the Matérn class and the corresponding Gaussian process has the second order mean square derivative. The last one is the γ-exponential covariance function with $\gamma = 1.5$ and the Gaussian process is mean square continuous but not mean square differentiable. We plot sample curves for each of the three Gaussian processes in Figure 1.

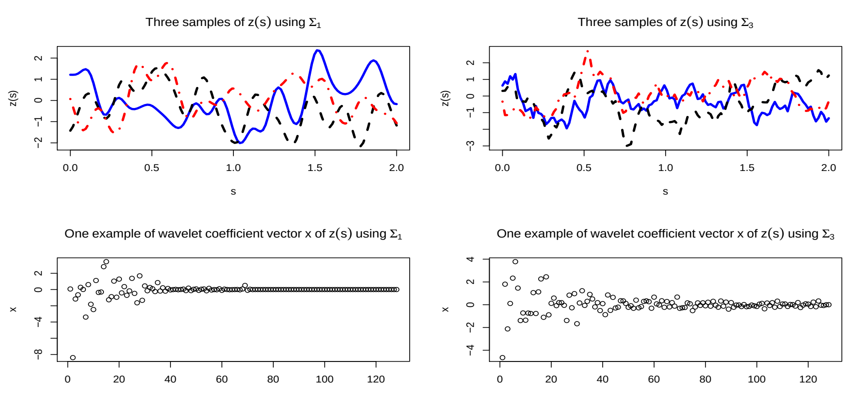

Fig 1. *Three sample curves and the wavelet coefficient vector for the blue curve from the Gaussian processes generated with the covariance function $\mathbf{\Sigma}_1$ or $\mathbf{\Sigma}_3$ specified in (3.1) in Simulation 1.*

We generate $z(s)$ from two Gaussian processes with covariance functions $\mathbf{\Sigma}_1$

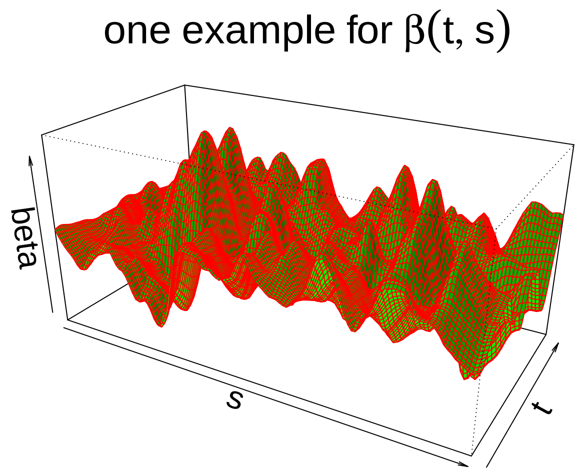

FIG 2. *One example of $\beta(t,s)$ generated by (3.2) with $\mathbf{\Sigma}_2$ in the additional simulation in Appendix C.*

and $\mathbf{\Sigma}_3$, respectively. We generate $\beta_0(t)$ and $\beta(t,s)$ in the following way,

$$\beta_0(t) = \zeta_0(t), \quad \beta(t,s) = 5\zeta_1(t)\xi_1(s) + 10\zeta_2(t)\xi_2(s) + 15\zeta_3(t)\xi_3(s), \tag{3.2}$$

where $\zeta_0(t), \zeta_1(t), \cdots, \zeta_3(t)$ are independently generated from the Gaussian process with covariance function $\mathbf{\Sigma}_2(t,t')$ and $\xi_1(s), \cdots, \xi_3(s)$ from $\mathbf{\Sigma}_2(s,s')$, where $\mathbf{\Sigma}_j(t,t')$ and $\mathbf{\Sigma}_j(s,s')$ have the same expression but differ in domains. One example for $\beta(t,s)$ generated from (3.2) is plotted in Figure 2. The noise $\varepsilon(t)$ is generated from the Gaussian process with covariance function $\mathbf{\Sigma}_\varepsilon(t,t') = \sigma^2\rho^{\{10|t-t'|\}^2}$. We consider two noise levels $\sigma^2 = 0.1, 0.25$, and two correlation levels $\rho = 0, 0.7$, respectively. In Figure 3, we plot three sample curves of $\varepsilon(t)$ for each of the two settings, $(\sigma^2, \rho) = (0.1, 0)$ and $(0.1, 0.7)$, respectively. When $\rho = 0.7$, strong within-function correlation exists in $\varepsilon(t)$ and the sample noise is actually a smooth curve. We consider all the 8 combinations of two types of $z(s)$, two σ^2 and two ρ values. For each combination, we repeat the following procedure 50 times. In each repeat, we first generate one $\beta_0(t)$ and one $\beta(t,s)$ based on (3.2). Then we generate 100 discretely observed random samples

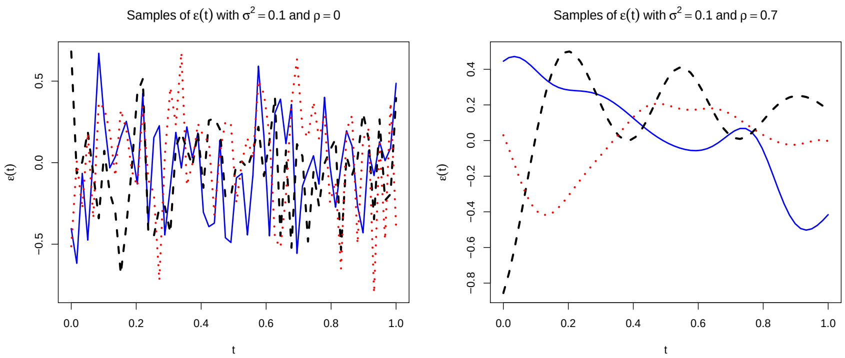

FIG 3. *Three sample curves of* $\varepsilon(t)$ *for* $(\sigma^2, \rho) = (0.1, 0)$ *or* $(0.1, 0.7)$ *in Simulation 1.*

$\{z_i(s_k), \varepsilon_i(t_\ell) | 1 \le i \le 100, 1 \le k \le 128, 1 \le \ell \le 60\}$, where $\{s_k, 1 \le k \le 128\}$ is the set of equally spaced observation points in $[0, 2]$ and $\{t_\ell, 1 \le \ell \le 60\}$ is the set of equally spaced observation points in $[0, 1]$. Then we calculate $y_i(t_l)$ based on the model (1.1) and use $\{z_i(s_k), y_i(t_\ell) | 1 \le i \le 100, 1 \le k \le 128, 1 \le \ell \le 60\}$ as the training data. Examples of $y_i(t)$ for different settings of $x(s)$ and $\beta(t, s)$ while fixing $\sigma^2 = 0.1$ and $\rho = 0$ are drawn in Figure **??**. Similarly the test data set $\{z_i^{\text{test}}(s_k), y_i^{\text{test}}(t_\ell) | 1 \le i \le 500, 1 \le k \le 128, 1 \le \ell \le 60\}$ is generated with size of 500. We use the training data to choose the tuning parameters and fit the model for each method. Then the final model is applied to the test data to calculate the predicted curves $\widehat{y}_i^{\text{predict}}(t)$, $1 \le i \le 500$. We calculate the mean squared prediction error for this repeat by

$$MSPE = \frac{1}{500} \sum_{i=1}^{500} \left\{ \frac{1}{60} \sum_{\ell=1}^{60} \left(\widehat{y}_i^{\text{predict}}(t_\ell) - y_i^{\text{test}}(t_\ell) \right)^2 \right\}. \tag{3.3}$$

We report the averages and standard deviations of the MSPEs of 50 repeats in Table 1 for all 8 settings. In all settings, our method *wSigComp* has the smallest MSPEs. For *wSigComp*, *linmod* and *PACE-reg*, the prediction errors increase when the magnitude σ^2 of noise and the within-function correlation ρ in $\varepsilon(t)$ increase. For *wSigComp*, smoother functional predictors generated from Gaussian process with the covariance function $\mathbf{\Sigma}_1$ tend to improve the prediction accuracy compared to those with $\mathbf{\Sigma}_3$. The number of components chosen by our method is 3 or 4, the number of principal components for $z(s)$ chosen by *pffr.pc* is 15 and more components are chosen by *PACE-reg*. In this simulation study, the FPC based methods, *pffr.pc* and *PACE-reg*, appear to perform worse than other methods, which may be because the regression functions cannot be well approximated by a small or moderate number of eigenfunctions.

TABLE 1
The averages and standard deviations (in parenthesis) of the MSPEs of 50 repeats for Simulation 1.

$z(s)$	σ^2	ρ	*wSigComp*	linmod	*pffr.pc*	*pffr*	*PACE-reg*
$\mathbf{\Sigma}_1$	0.1	0	0.115(0.003)	0.213(0.060)	1.598(0.727)	0.134(0.005)	5.219(5.189)
		0.7	0.130(0.008)	0.275(0.098)	1.649(0.980)	0.172(0.033)	6.165(4.717)
	0.25	0	0.271(0.004)	0.395(0.057)	1.376(0.518)	0.159(0.010)	6.445(4.946)
		0.7	0.307(0.015)	0.447(0.077)	1.514(0.713)	0.205(0.032)	7.080(6.304)
$\mathbf{\Sigma}_3$	0.1	0	0.121(0.008)	0.222(0.072)	1.536(0.945)	0.310(0.007)	5.479(5.099)
		0.7	0.146(0.014)	0.251(0.074)	1.407(0.890)	0.355(0.029)	6.458(4.621)
	0.25	0	0.280(0.008)	0.409(0.055)	1.567(0.851)	0.385(0.027)	6.635(4.806)
		0.7	0.334(0.021)	0.469(0.067)	1.645(0.701)	0.434(0.038)	7.585(6.241)

4. Appendix D: Additional figures

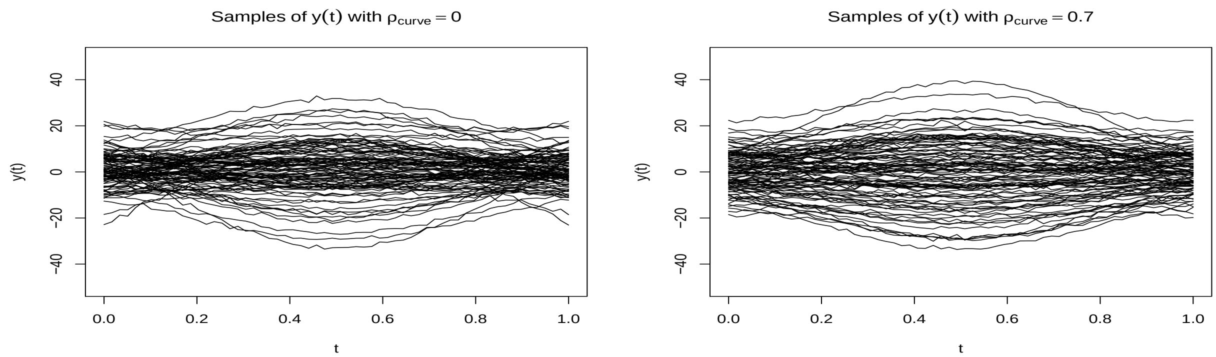

FIG 4. *100 sample curves of $y(t)$ from $z_q(s)$ with covariance function $\mathbf{\Sigma}_1$, $\sigma^2 = 0.25$, and $\rho_{\text{curve}} = 0$ and 0.7, respectively, in Simulation 3.*

References

[1] Ledoux, M. and Talagrand, M. (2011) *Probability in Banach Spaces: Isoperimetry and Processes.* Classics in Mathematics. Springer.

[2] Rasmussen, C. E. and Williams, C. K. I. (2005) *Gaussian Processes for Machine Learning (Adaptive Computation and Machine Learning series)*, chap. 4. The MIT Pres.

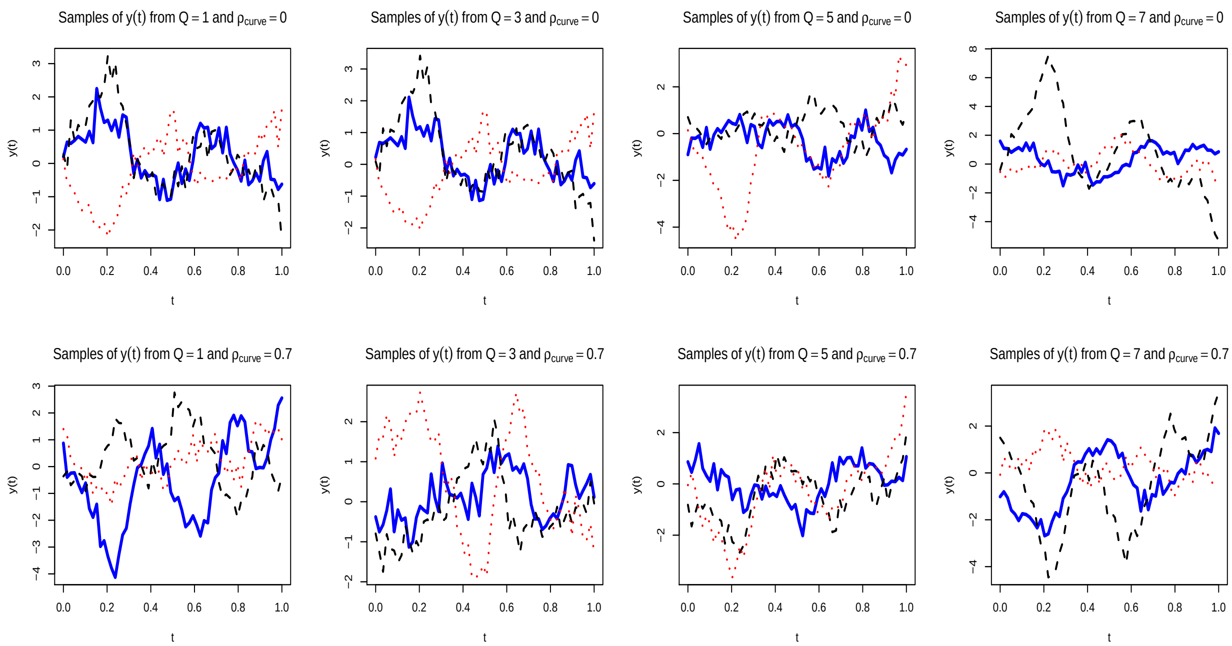

FIG 5. *Three sample curves of $y(t)$ for different Q and ρ_{curve} in Simulation 4.*