Supplementary Material to "Scalable Interpretable Multi-Response Regression via SEED"

Zemin Zheng, M. Taha Bahadori, Yan Liu and Jinchi Lv

Appendix B. Auxiliary Technical Results

We provide some auxiliary results in this section.

B.1 Existence and Identifiability of Decomposition (2) in Section 2.1

$\mathbf{P}$-orthogonality of the left singular vectors arises since the eigenvectors of the generalized eigenvalue problem (3) are $\mathbf{P}$-orthogonal (Parlett, 1998, Chapter 15). Given that our analysis is based on estimating the top-r^* eigenvectors of (3), we need to adapt this formulation to connect the estimate with the ground truth. In practice, we have showed that SEED does not need the $\mathbf{P}$-orthogonality assumption in Section 3.

Now we show the existence of regression coefficient matrix $\mathbf{C}^*$ satisfying $\mathbf{X}\mathbf{C}^* = \mathbf{Y}^*$ and the decomposition in (2) with $\mathbf{u}_k^* \perp \text{Ker}(\mathbf{P})$ and $\text{rank}(\mathbf{C}^*) = \text{rank}(\mathbf{X}\mathbf{C}^*) = r^*$. Based on the singular value decomposition of $\mathbf{Y}^* = \mathbf{Z}\mathbf{S}\mathbf{V}^T$ with $\mathbf{S} \in \mathbb{R}^{r^*\times r^*}$ invertible, we can get r^* solutions $\widetilde{\mathbf{u}}_1, \ldots, \widetilde{\mathbf{u}}_{r^*}$ from the equations $\mathbf{X}\widetilde{\mathbf{u}}_k = \mathbf{z}_k$, where $\mathbf{z}_k$ is the kth column of $\mathbf{Z}$. Recall that $\mathbf{P} = \mathbf{X}^T\mathbf{X}$. The $\mathbf{P}$-orthogonality of $\{\widetilde{\mathbf{u}}_k\}$ follows from the orthogonality of $\{\mathbf{z}_k\}$. By removing the projections of $\{\widetilde{\mathbf{u}}_k\}$ on $\text{Ker}(\mathbf{P})$, we get a matrix $\mathbf{U} = (\mathbf{u}_1, \ldots, \mathbf{u}_{r^*})$ with $\mathbf{P}$-orthogonal columns satisfying $\mathbf{u}_k \perp \text{Ker}(\mathbf{P})$ and $\mathbf{X}\mathbf{u}_k = \mathbf{z}_k$ since removing the $\mathbf{P}$-projection does not change the validity of the equations. Let $\mathbf{C}^* = \mathbf{U}\mathbf{S}\mathbf{V}^T$. It is easy to check that $\mathbf{X}\mathbf{C}^* = \mathbf{Y}^*$ and decomposition (2) follows from the rescaling of $\mathbf{u}_k$ to unit length and multiplying the rescaled diagonal components on the columns of $\mathbf{V}$.

On the other hand, provided that $\mathbf{u}_k^* \perp \text{Ker}(\mathbf{P})$, we have the unique choices of $\mathbf{u}_k^*$ and $\mathbf{v}_k^*$ for $\mathbf{C}^* = \sum_{k=1}^{r^*} \mathbf{u}_k^*\mathbf{v}_k^{*T}$ satisfying decomposition (2) such that $\mathbf{X}\mathbf{C}^* = \mathbf{Y}^*$, up to simultaneous sign changes of $\mathbf{u}_k^*$ and $\mathbf{v}_k^*$. Because when $\mathbf{u}_k^*$ are $\mathbf{P}$-orthogonal and $\mathbf{v}_k^*$ are orthogonal, $\mathbf{v}_k^*$ are the r^* eigenvectors of the matrix $\mathbf{C}^{*T}\mathbf{P}\mathbf{C}^*$ (of rank r^*), thus linearly independent and their directions are unique up to sign changes, which further makes the linear combinations $\mathbf{C}^* = \sum_{k=1}^{r^*} \mathbf{u}_k^*\mathbf{v}_k^{*T}$ unique when $\mathbf{u}_k^*$ are unit length vectors. It guarantees the identifiability of decomposition (2) for $\mathbf{C}^*$.

In fact, given a semi-positive definite matrix $\mathbf{P}$, $\mathbf{P}$-orthogonality corresponds to the inner product $\langle \mathbf{u}_1, \mathbf{u}_2\rangle_{\mathbf{P}} = \mathbf{u}_1^T\mathbf{P}\mathbf{u}_2$. In the subspace $\mathbf{u} \perp \text{Ker}(\mathbf{P})$, this inner product defines a proper norm. To show this, we first note that $\mathbf{u}^T\mathbf{P}\mathbf{u} = 0$ requires $\mathbf{u} = \mathbf{0}$ according to the definition of the subspace. Second, the norm also satisfies absolute homogeneity, that is, for any scalar a, $(a\mathbf{u})^T\mathbf{P}(a\mathbf{u}) = a^2\mathbf{u}^T\mathbf{P}\mathbf{u}$. Third, it also satisfies the triangular inequality. To see this, let $\mathbf{P} = \mathbf{H}^T\mathbf{H}$ such that $\mathbf{u}_1^T\mathbf{P}\mathbf{u}_2 = \widetilde{\mathbf{u}}_1^T\widetilde{\mathbf{u}}_2$ with $\widetilde{\mathbf{u}}_1 = \mathbf{H}\mathbf{u}_1$ and $\widetilde{\mathbf{u}}_2 = \mathbf{H}\mathbf{u}_2$. Then the triangular inequality follows from the triangular inequality in L_2-norm. Thus, replacing L_2-norm with this new norm, the $\mathbf{P}$-orthogonalization for the left singular vectors can be done in a similar way to the regular orthogonalization.

B.2 Proof of Proposition 1

Let $\mathbf{u}_k^{*T}\mathbf{P}\mathbf{u}_k^* = c_k^2$ and $\mathbf{v}_k^{*T}\mathbf{v}_k^* = qa_k^2$. It follows from the definition of σ_k that $\sigma_k = a_kc_k$, which decreases with k for $1 \le k \le r^* - 1$. Since $\mathbf{C}^* = \sum_{k=1}^{r^*} \mathbf{u}_k^*\mathbf{v}_k^{*T}$, for the noise-free data we can write

$$\mathbf{X}^T\mathbf{Y}^*\mathbf{Y}^{*T}\mathbf{X} = \mathbf{X}^T\left(\mathbf{X}\mathbf{C}^*\mathbf{C}^{*T}\mathbf{X}^T\right)\mathbf{X} = \sum_{k=1}^{r^*} qa_k^2\mathbf{X}^T\mathbf{X}\mathbf{u}_k^*\mathbf{u}_k^{*T}\mathbf{X}^T\mathbf{X},$$

where the last step is due to the orthogonality of $\mathbf{v}_k^*$. Therefore, because of the $\mathbf{P}$-orthogonality of $\mathbf{u}_k^*$, right-multiplying the above equation by $\mathbf{u}_j^*$ yields

$$\mathbf{X}^T\mathbf{Y}^*\mathbf{Y}^{*T}\mathbf{X}\mathbf{u}_j^* = nqa_j^2c_j^2\mathbf{X}^T\mathbf{X}\mathbf{u}_j^*,$$

which means $\mathbf{u}_j^*$ is an eigenvector of the generalized eigenvalue problem (3) with respect to the nonzero eigenvalue $\lambda_j = nqa_j^2c_j^2 = nq\sigma_j^2$. Furthermore, based on the $\mathbf{P}$-orthogonality of $\mathbf{u}_k^*$, we obtain

$$\mathbf{Y}^{*T}\mathbf{X}\mathbf{u}_j^* = \mathbf{C}^{*T}\mathbf{X}^T\mathbf{X}\mathbf{u}_j^* = \sum_{k=1}^{r^*}\mathbf{v}_k^*\mathbf{u}_k^{*T}\mathbf{X}^T\mathbf{X}\mathbf{u}_j^* = \mathbf{v}_j^*\mathbf{u}_j^{*T}\mathbf{X}^T\mathbf{X}\mathbf{u}_j^*,$$

which gives the expression of $\mathbf{v}_k^*$ in equation (4).

By the same argument, it is not difficult to see that $\mathbf{X}\mathbf{u}_j^*$ is the eigenvector of $\mathbf{Y}^*\mathbf{Y}^{*T}$ with respect to the same eigenvalue λ_j. Since all $\mathbf{u}_j^*$ belong to the space $\mathbf{u} \perp \mathrm{Ker}(\mathbf{P})$, the rank of $\mathbf{Y}^*\mathbf{Y}^{*T}$ is r^*. Thus, we know that $\lambda_1, \cdots, \lambda_{r^*}$ are all nonzero eigenvalues of $\mathbf{Y}^*\mathbf{Y}^{*T}$. On the other hand, the generalized eigenvalue problem (3) can also be written as

$$\mathbf{X}^T(\mathbf{Y}^*\mathbf{Y}^{*T} - \lambda\mathbf{I})\mathbf{X}\mathbf{u} = \mathbf{0}. \tag{A.1}$$

We claim that constrained on the space $\mathbf{u} \perp \mathrm{Ker}(\mathbf{P})$, the above equation (A.1) would share exactly the same nonzero eigenvalues with $\mathbf{Y}^*\mathbf{Y}^{*T}$. This is due to a key fact that $\mathbf{Y}^*\mathbf{Y}^{*T} - \lambda\mathbf{I}$ is of full rank n when λ is nonzero but not any eigenvalue of $\mathbf{Y}^*\mathbf{Y}^{*T}$, which gives

$$\mathrm{span}\{\mathbf{X}^T(\mathbf{Y}^*\mathbf{Y}^{*T} - \lambda\mathbf{I})\mathbf{X}\} = \mathrm{span}\{\mathbf{X}\}.$$

Here $\mathrm{span}\{\mathbf{A}\}$ denotes the space spanned by the row vectors of a given matrix $\mathbf{A}$. It follows that equation (A.1) holds if and only if $\mathbf{X}\mathbf{u} = \mathbf{0}$. Therefore, constrained on the space $\mathbf{u} \perp \mathrm{Ker}(\mathbf{P})$, the generalized eigenvalue problem (3) shares exactly the same nonzero eigenvalues with $\mathbf{Y}^*\mathbf{Y}^{*T}$.

Based on the above discussion, we conclude that constrained on the space $\mathbf{u} \perp \mathrm{Ker}(\mathbf{P})$, the generalized eigenvalue problem (3) has exactly r^* nonzero eigenvalues $\lambda_1, \cdots, \lambda_{r^*}$ with the corresponding eigenvectors $\mathbf{u}_1^*, \cdots, \mathbf{u}_{r^*}^*$. Thus, we finish the proof of Proposition 1.

B.3 Proof of Theorem 4

We present the proof for an arbitrary k, $k = 1, \ldots, r^*$ and drop the index k for simplicity of the notation. The proof indeed follows from the fact that $\|\mathbf{u}\|_\infty \leq \|\mathbf{u}\|_2$ and the estimation error bounds in Theorem 1, which implies that with probability at least $1 - \delta$ and for sufficiently large n, the following bounds hold,

$$|\widehat{u}_i - u_i^*| < \frac{5}{4}C_u\sqrt{\frac{s}{n}\log\frac{pq}{\delta}}, \quad i = 1, \ldots, p. \tag{A.2}$$

On one hand, thresholding the elements of $\widehat{\mathbf{u}}$ by $\theta \in (\frac{5}{4}T, \frac{7}{4}T)$ with $T = C_u\sqrt{\frac{s}{n}\log\frac{pq}{\delta}}$ will make the magnitudes of all non-zero elements of $T_\theta(\widehat{\mathbf{u}})$ no less than $\frac{5}{4}C_u\sqrt{\frac{s}{n}\log\frac{pq}{\delta}}$. Then it follows immediately from the error bound (A.2) that the corresponding elements in $\mathbf{u}^*$ are also nonzero, which gives $\text{supp}(T_\theta(\widehat{\mathbf{u}})) \subset \text{supp}(\mathbf{u}^*)$. On the other hand, the error bound (A.2) and the lower bound on the magnitudes of the nonzero elements of $\mathbf{u}^*$ in Condition 4 together ensure that $\widehat{u}_i > \frac{7}{4}T$ if $i \in \text{supp}(\mathbf{u}^*)$, which is still nonzero after the aforementioned thresholding procedure. Therefore, we get $\text{supp}(\mathbf{u}^*) \subset \text{supp}(T_\theta(\widehat{\mathbf{u}}))$.

Based on the analysis before, we conclude that $\text{supp}(T_\theta(\widehat{\mathbf{u}})) = \text{supp}(\mathbf{u}^*)$. Moreover, given the estimation error bounds on $\|\widehat{\mathbf{v}} - \mathbf{v}^*\|_2/\sqrt{q}$ in Theorem 1 and the minimum signal strength of $\mathbf{v}^*$ in Condition 4, similarly we have $\text{supp}(T_\theta(\widehat{\mathbf{v}})) = \text{supp}(\mathbf{v}^*)$. It completes the proof of Theorem 4.

B.4 Proof of Lemma 5

It holds that

$$\begin{aligned}
\mathbb{P}\left[\frac{1}{n}\|\mathbf{E}^T\mathbf{X}\|_{2,s} \ge \epsilon\right] &= \mathbb{P}\left[\frac{1}{n^2}\|\mathbf{E}^T\mathbf{X}\|_{2,s}^2 \ge \epsilon^2\right] = \mathbb{P}\left[\sup_{\|\mathbf{u}\|_2=1,\|\mathbf{u}\|_0\le s}\frac{1}{n^2}\|\mathbf{E}^T\mathbf{X}\mathbf{u}\|_2^2 > \epsilon^2\right] \\
&= \mathbb{P}\left[\sup_{\|\mathbf{u}\|_2=1,\|\mathbf{u}\|_0\le s}\sum_{i=1}^q\frac{1}{n^2}(\mathbf{E}_{:i}^T\mathbf{X}\mathbf{u})^2 > \epsilon^2\right] \le \mathbb{P}\left[\sum_{i=1}^q\sup_{\|\mathbf{u}\|_2=1,\|\mathbf{u}\|_0\le s}\frac{1}{n^2}(\mathbf{E}_{:i}^T\mathbf{X}\mathbf{u})^2 > \epsilon^2\right],
\end{aligned}$$

where $\mathbf{E}_{:i}$ denotes the ith column of $\mathbf{E}$. For the last term above, an application of the union bound yields

$$\begin{aligned}
&\mathbb{P}\left[\sum_{i=1}^q\sup_{\|\mathbf{u}\|_2=1,\|\mathbf{u}\|_0\le s}\frac{1}{n^2}(\mathbf{E}_{:i}^T\mathbf{X}\mathbf{u})^2 > \epsilon^2\right] \le \sum_{i=1}^q\mathbb{P}\left[\sup_{\|\mathbf{u}\|_2=1,\|\mathbf{u}\|_0\le s}\frac{1}{n^2}(\mathbf{E}_{:i}^T\mathbf{X}\mathbf{u})^2 > \epsilon^2/q\right] \\
&= \sum_{i=1}^q\mathbb{P}\left[\sup_{\|\mathbf{u}\|_2=1,\|\mathbf{u}\|_0\le s}\frac{1}{n}|\mathbf{E}_{:i}^T\mathbf{X}\mathbf{u}| > \epsilon/\sqrt{q}\right] \le \sum_{i=1}^q\mathbb{P}\left[\max_{1\le j\le p}\frac{1}{n}|\mathbf{E}_{:i}^T\mathbf{X}_{:j}| > \epsilon/\sqrt{qs}\right] \\
&\le \sum_{i=1}^q\sum_{j=1}^p\mathbb{P}\left[\frac{1}{\sqrt{n}}|\mathbf{E}_{:i}^T\mathbf{X}_{:j}| > \epsilon\frac{\sqrt{n}}{\sqrt{qs}}\right] \le 2pq\exp\left(-\frac{n\epsilon^2}{2qs\gamma^2P^2}\right),
\end{aligned}$$

where the Gaussian tail bound is applied in the last inequality. Let $\delta = 2pq\exp\left(-\frac{n\epsilon^2}{2qs\gamma^2P^2}\right)$, we obtain $\epsilon = \sqrt{\frac{2qs(\gamma P)^2}{n}\log\frac{2pq}{\delta}}$, which concludes the proof.

B.5 Proof of Lemma 6

Remark. When the perturbed variant (11) does not have s-sparse solution, we define $\widehat{\lambda}_k$ and $\widehat{\mathbf{u}}_k$ according to equation (A.3). Then the perturbation bounds in (12) and (13) still hold when the perturbation is smaller than the magnitudes of the population eigenvectors. If we remove the sparsity constraint, the perturbed eigenvalue can deviate further from the population one, which can also be seen from (12).

We will prove the results in two steps.

Proof of equation (12). Denote by U some subspace satisfying $U \perp \mathrm{Ker}(\mathbf{P})$. According to the Courant-Fischer min-max theorem for generalized eigenvalues (Stewart and Sun, 1990, Chapter VI, Corollary 1.16), the kth s-sparse eigenvalue of the perturbed variant can be written as

$$\begin{aligned}\widehat{\lambda}_k &= \max_{\dim(U)=k} \min_{\mathbf{u}\in U\setminus\{\mathbf{0}\},\ \|\mathbf{u}\|_0\le s} \frac{\mathbf{u}^T(\mathbf{Q}+\boldsymbol{\delta}_Q)\mathbf{u}}{\mathbf{u}^T\mathbf{P}\mathbf{u}} \\ &= \max_{\dim(U)=k} \min_{\mathbf{u}\in U\setminus\{\mathbf{0}\},\ \|\mathbf{u}\|_0\le s} \left\{\frac{\mathbf{u}^T\mathbf{Q}\mathbf{u}}{\mathbf{u}^T\mathbf{P}\mathbf{u}} + \frac{\mathbf{u}^T\boldsymbol{\delta}_Q\mathbf{u}}{\mathbf{u}^T\mathbf{P}\mathbf{u}}\right\} \\ &\le \max_{\dim(U)=k} \min_{\mathbf{u}\in U\setminus\{\mathbf{0}\},\ \|\mathbf{u}\|_0\le s^*} \left\{\frac{\mathbf{u}^T\mathbf{Q}\mathbf{u}}{\mathbf{u}^T\mathbf{P}\mathbf{u}} + \phi_s^{-1}\|\boldsymbol{\delta}_Q\|_{2,s}\right\} \\ &= \lambda_k + \phi_s^{-1}\|\boldsymbol{\delta}_Q\|_{2,s},\end{aligned} \tag{A.3}$$

where we use the sparsity of $\mathbf{u}$ together with Condition 1 to bound $\mathbf{u}^T\mathbf{P}\mathbf{u}$ from below. Similarly, another application of the same theorem yields $\widehat{\lambda}_k - \lambda_k \ge -\phi_s^{-1}\|\boldsymbol{\delta}_Q\|_{2,s}$. Thus, we obtain the result in equation (12).

Proof of equation (13). For notational clarity, we prove the bound for any fixed k and drop the index k from the vectors. Given every s-sparse eigenvector $\widehat{\mathbf{u}}$, we can write it as $\widehat{\mathbf{u}} = \alpha\mathbf{u} + \beta\mathbf{u}'$, where $\alpha \ge 0$ (otherwise change $\widehat{\mathbf{u}}$ to $-\widehat{\mathbf{u}}$ such that it takes the correct direction), $\mathbf{u}^T\mathbf{P}\mathbf{u}' = 0$, and $\|\mathbf{u}'\|_2 = 1$. According to equation (11), we have

$$(\mathbf{Q}+\boldsymbol{\delta}_Q)(\alpha\mathbf{u}+\beta\mathbf{u}') = (\lambda+\Delta_\lambda)\mathbf{P}(\alpha\mathbf{u}+\beta\mathbf{u}'),$$

where $\Delta_\lambda = \widehat{\lambda} - \lambda$. Since $\mathbf{Q}\mathbf{u} = \lambda\mathbf{P}\mathbf{u}$ and $\mathbf{u}^T\mathbf{P}\mathbf{u}' = 0$, we have $\mathbf{u}^T\mathbf{Q}\mathbf{u}' = 0$. Thus, left-multiplying both sides by $\mathbf{u}'^T$ and reorganizing the above equation yields

$$\beta\mathbf{u}'^T(\mathbf{Q}-\lambda\mathbf{P})\mathbf{u}' - \beta\Delta_\lambda\mathbf{u}'^T\mathbf{P}\mathbf{u}' = -\alpha\mathbf{u}'^T(\boldsymbol{\delta}_Q)\widehat{\mathbf{u}}. \tag{A.4}$$

When $\beta = 0$, it is a trivial case that $\widehat{\mathbf{u}} = \mathbf{u}$ and definitely satisfies the error bound in equation (13). So we focus on the case that $\beta \ne 0$. Since $\mathbf{u}^T\mathbf{P}\mathbf{u}' = 0$ and $\mathbf{Q}\mathbf{u} = \lambda\mathbf{P}\mathbf{u}$, $\mathbf{u}'$ is $\mathbf{P}$-orthogonal to the eigenspace of eigenvalue λ spanned by $\mathbf{u}$. It follows that

$$|\mathbf{u}'^T(\mathbf{Q}-\lambda\mathbf{P})\mathbf{u}'| \ge d_\lambda\mathbf{u}'^T\mathbf{P}\mathbf{u}',$$

where d_λ is the minimum eigengap between non-zero eigenvalues of equation (10). Recall that $\mathbf{u}'$ is a unit length vector. We have $\|\mathbf{u}'\|_0 \le 2s$ since $\widehat{\mathbf{u}} = \alpha\mathbf{u}+\beta\mathbf{u}'$, $\|\mathbf{u}\|_0 \le s^*$, $\|\widehat{\mathbf{u}}\|_0 \le s$ and $s^* < s$. It yields from Condition 1 that

$$\phi_s = \phi_s\|\mathbf{u}'\|_2^2 \le \mathbf{u}'^T\mathbf{P}\mathbf{u}' \le \phi_s^{-1}\|\mathbf{u}'\|_2^2 = \phi_s^{-1}. \tag{A.5}$$

Similarly, we have $\widehat{\mathbf{u}}^T\mathbf{P}\widehat{\mathbf{u}} \le \phi_s^{-1}$ and $\mathbf{u}^T\mathbf{P}\mathbf{u} \le \phi_s^{-1}$. Thus, it follows from equation (A.4) that

$$|\beta| = \left|\frac{\alpha\mathbf{u}'^T(\boldsymbol{\delta}_Q)\widehat{\mathbf{u}}}{\mathbf{u}'^T(\mathbf{Q}-\lambda\mathbf{P})\mathbf{u}' - \Delta_\lambda\mathbf{u}'^T\mathbf{P}\mathbf{u}'}\right| \le \frac{|\mathbf{u}'^T(\boldsymbol{\delta}_Q)\widehat{\mathbf{u}}|}{(d_\lambda - |\Delta_\lambda|)\mathbf{u}'^T\mathbf{P}\mathbf{u}'} \le \frac{\|\boldsymbol{\delta}_Q\|_{2,s}}{\phi_s(d_\lambda - |\Delta_\lambda|)}.$$

On the other hand, due to the facts $\widehat{\mathbf{u}} = \alpha\mathbf{u} + \beta\mathbf{u}'$ and $\|\widehat{\mathbf{u}}\|_2 = \|\mathbf{u}\|_2 = \|\mathbf{u}'\|_2 = 1$, we get

$$\begin{aligned}1 &= \widehat{\mathbf{u}}^T\widehat{\mathbf{u}} = (\alpha\mathbf{u} + \beta\mathbf{u}')^T(\alpha\mathbf{u} + \beta\mathbf{u}') \\ &= \alpha^2\|\mathbf{u}\|_2^2 + \beta^2\|\mathbf{u}'\|_2^2 + 2\alpha\beta\mathbf{u}^T\mathbf{u}' \le \alpha^2 + \beta^2 + 2|\alpha\beta|.\end{aligned}$$

It yields that $\alpha + |\beta| \ge 1$ as $\alpha \ge 0$. By a similar argument, we can obtain $\alpha - |\beta| \le 1$. Therefore, we conclude that $|1 - \alpha| \le |\beta|$. Together with $\mathbf{u}^T\mathbf{P}\mathbf{u}' = 0$, $\mathbf{u}^T\mathbf{P}\mathbf{u} \le \phi_s^{-1}$, and $\mathbf{u}'^T\mathbf{P}\mathbf{u}' \le \phi_s^{-1}$, we get

$$\begin{aligned}(\widehat{\mathbf{u}} - \mathbf{u})^T\mathbf{P}(\widehat{\mathbf{u}} - \mathbf{u}) &= (\alpha\mathbf{u} + \beta\mathbf{u}' - \mathbf{u})^T\mathbf{P}(\alpha\mathbf{u} + \beta\mathbf{u}' - \mathbf{u}) \\ &= \left[(1 - \alpha)^2\mathbf{u}^T\mathbf{P}\mathbf{u} + \beta^2\mathbf{u}'^T\mathbf{P}\mathbf{u}'\right] \le 2\phi_s^{-1}\beta^2.\end{aligned}$$

Since $\|\widehat{\mathbf{u}} - \mathbf{u}\|_0 \le 2s$, by a similar argument as in equation (A.5), we get

$$\|\widehat{\mathbf{u}} - \mathbf{u}\|_2^2 \le \phi_s^{-1}(\widehat{\mathbf{u}} - \mathbf{u})^T\mathbf{P}(\widehat{\mathbf{u}} - \mathbf{u}) \le 2\phi_s^{-2}\beta^2.$$

Substituting $|\beta|$ with its upper bound established before, we obtain

$$\begin{aligned}\|\widehat{\mathbf{u}} - \mathbf{u}\|_2 &\le \sqrt{2}\phi_s^{-2}(d_\lambda - |\Delta_\lambda|)^{-1}\|\boldsymbol{\delta}_Q\|_{2,s} \\ &= \sqrt{2}\phi_s^{-2}d_\lambda^{-1}\|\boldsymbol{\delta}_Q\|_{2,s} + o(d_\lambda^{-1}\|\boldsymbol{\delta}_Q\|_{2,s}),\end{aligned}$$

where the last equality follows from the uniform bound of $|\Delta_\lambda|$ in (12) and the assumption that $\|\boldsymbol{\delta}_Q\|_{2,s} = o(d_\lambda)$. Thus, the proof is completed.