Supplementary Material for "Variable Selection and Minimax Prediction in High-Dimensional Functional Linear Models"

Xingche Guo, Yehua Li and Tailen Hsing

This Supplemental Paper is organized as follows. We provide technical details in Section S.1, where Section S.1.1 includes details for deriving the Karush-Kuhn-Tucker conditions in function spaces, Section S.1.2 introduces a special case for high-dimensional functional predictors with partially separable covariance structure. All technical proofs are provided in Section S.2, where Section S.2.1 contains the proofs of the propositions, Section S.2.2 contains the proofs of Theorems 1-4 as well as Corollary 1, Section S.2.3 provides the proofs of the lemmas used in the main proof of Theorem 1, and Section S.2.4 provides some additional technical lemmas. Substantiating examples are provided in Section S.3 to support the technical assumptions made in the paper, and some additional simulation results are provided in Section S.4.

S.1. Technical Details

S.1.1 Karush-Kuhn-Tucker Conditions in Function Spaces

In this section, we introduce the Karush-Kuhn-Tucker (KKT) condition in function spaces and specialize it for (2.4). First, we review the notion of Gateaux differentiability. For convenience, let $\mathscr{J}$ denote a mapping from some Hilbert space $\mathbb{H}$ to $\mathbb{R}$, where $\mathscr{J}$ is not necessarily linear. We note that the Hilbert space assumption in the definition below could be relaxed depending on the context of the application.

Definition 1. *(Gateaux differentiability) For $f, \psi \in \mathbb{H}$, we say that $\mathscr{J}$ is Gateaux differentiable at f in the direction of ψ if $\lim_{\tau\downarrow 0^+} \frac{\mathscr{J}(f+\tau\psi)-\mathscr{J}(f)}{\tau}$ and $\lim_{\tau\uparrow 0^-} \frac{\mathscr{J}(f+\tau\psi)-\mathscr{J}(f)}{\tau}$ exist and are equal. The common limit in this case is denoted by $\mathscr{D}_{\mathscr{J}}(f;\psi)$ and is referred to as the Gateaux derivative of $\mathscr{J}$ at f in the direction of ψ. If $\mathscr{D}_{\mathscr{J}}(f;\psi)$ is defined for all $\psi \in \mathbb{H}$, we say that $\mathscr{J}$ is Gateaux differentiable at f.*

Clearly, if $\mathscr{J}$ is Gateaux differentiable at f then $\mathscr{D}_{\mathscr{J}}(f;\cdot) \in \mathfrak{B}(\mathbb{H},\mathbb{R})$, the space of continuous linear functionals on $\mathbb{H}$. On the other hand, if $\mathscr{J}$ is convex but not necessarily Gateaux differentiable, then the useful notions of sub-derivative and sub-differential can be defined as follows.

Definition 2. *(Sub-derivative and sub-differential) The Gateaux sub-differential* *of a convex functional* $\mathscr{J}$ *at* g *is defined as the collection* $\partial_{\mathscr{J}(g)} = \{\mathscr{A} \in \mathfrak{B}(\mathbb{H}, \mathbb{R}) : \mathscr{J}(f) \geq \mathscr{J}(g) + \mathscr{A}(f-g)$ *for all* $f \in \mathbb{H}\}$ *of linear functionals, where the elements in* $\partial_{\mathscr{J}(g)}$ *are referred to as sub-derivatives.*

Proposition 5. *Any Gateaux differentiable mapping* $\mathscr{J}$ *from* $\mathbb{H}$ *to* $\mathbb{R}$ *is convex if and only if* $\mathscr{J}(f) \geq \mathscr{J}(g) + \mathscr{D}_{\mathscr{J}}(g; f-g)$ *for all* $f, g \in \mathbb{H}$*, in which case* $\mathscr{J}(g)$ *is the global minimum of* $\mathscr{J}(\cdot)$ *if and only if* $\mathscr{D}_{\mathscr{J}}(g; \cdot) \equiv 0$*. Suppose, on the other hand, that* $\mathscr{J}$ *is convex but not Gateaux differentiable. Then* $\mathscr{J}(g)$ *is the global minimum of* $\mathscr{J}$ *if and only if* $0 \in \partial_{\mathscr{J}(g)}$*.*

With $\mathscr{T}_n$ and $\boldsymbol{g}_n$ defined in (2.5), the objective function $\ell(\boldsymbol{f})$ can be expressed as

$$\ell(\boldsymbol{f}) = \sum_{i=1}^{4} \ell_i(\boldsymbol{f}) + \frac{1}{2n}\|\boldsymbol{\varepsilon}_n\|_2^2, \tag{S.1}$$

where $\ell_1(\boldsymbol{f}) = \frac{1}{2}\langle \mathscr{T}_n(\boldsymbol{f}-\boldsymbol{f}_0), \boldsymbol{f}-\boldsymbol{f}_0\rangle_2$, $\ell_2(\boldsymbol{f}) = -\langle \boldsymbol{g}_n, \boldsymbol{f}-\boldsymbol{f}_0\rangle_2$, $\ell_3(\boldsymbol{f}) = \frac{\lambda_2}{2}\|\boldsymbol{f}\|_2^2$, $\ell_4(\boldsymbol{f}) = \lambda_1 \sum_{j=1}^{p} \|\Psi_j f_j\|_2$, $\boldsymbol{f} \in \mathbb{L}_2^p$. The following proposition contains the key elements in minimizing $\ell(\boldsymbol{f})$ based on (S.1).

Proposition 6. *The functionals* $\ell_i, i = 1, 2, 3$*, are Gateaux differentiable at all* $\boldsymbol{f} \in \mathbb{L}_2^p$*, where* $\mathscr{D}_{\ell_1}(\boldsymbol{f}; \boldsymbol{\psi}) = \langle \mathscr{T}_n(\boldsymbol{f}-\boldsymbol{f}_0), \boldsymbol{\psi}\rangle_2$*,* $\mathscr{D}_{\ell_2}(\boldsymbol{f}; \boldsymbol{\psi}) = -\langle \boldsymbol{g}_n, \boldsymbol{\psi}\rangle_2$*, and* $\mathscr{D}_{\ell_3}(\boldsymbol{f}; \boldsymbol{\psi}) = \lambda_2\langle \boldsymbol{f}, \boldsymbol{\psi}\rangle_2$*. The sub-differential of* ℓ_4 *at* $\boldsymbol{f}$ *contains all functionals of the form* $\lambda_1 \langle \boldsymbol{\omega}, \cdot\rangle_2$*,* $\boldsymbol{\omega} \in \mathbb{L}_2^p$*, such that* $\omega_j = \frac{\Psi_j^2 f_j}{\|\Psi_j f_j\|_2}$ *if* $f_j \neq 0$ *and* $\omega_j = \Psi_j \eta_j$

for any arbitrary η_j *with* $\|\eta_j\|_2 \leq 1$ *if* $f_j = 0$.

Note that the KKT condition (2.5) can be easily derived from Propositions 5 and 6. The proofs for Propositions 5 and 6 are given in the Supplementary Material.

S.1.2 Partially Separable Covariance Structure

To gain a deeper understanding of Conditions C.2-C.4, we consider functional predictors with a partially separable covariance structure (Zapata et al., 2021), namely,

$$\mathscr{T}^{(\mathscr{S},\mathscr{S})} = \sum_{k=1}^{\infty} \boldsymbol{A}_k \psi_k \otimes \psi_k, \tag{S.2}$$

where $\{\psi_k, k \geq 1\}$ are orthonormal functions in $\mathbb{L}_2[0,1]$ and $\{\boldsymbol{A}_k, k \geq 1\}$ are a sequence of $q \times q$ covariance matrices. Further, consider $\boldsymbol{A}_k = \nu_k \boldsymbol{R}$, with $\nu_1 \geq \nu_2 \geq \cdots > 0$ a sequence of eigenvalues and $\boldsymbol{R}$ a $q \times q$ correlation matrix. In this setting, $\{X_j, j \in \mathscr{S}\}$ share the same eigenvalues and eigenfunctions, and their principal component scores have the same correlation structure across different order k. To satisfy Condition C.2, we must have $\nu_1 = 1$ and $\sum_{k \geq 1} \nu_k \leq \tau < \infty$. To find the upper bound for $\varkappa(\lambda_2)$, first note that

$$\mathscr{T}^{(\mathscr{S},\mathscr{S})}(\mathscr{T}_{\lambda_2}^{(\mathscr{S},\mathscr{S})})^{-1} = \sum_{k=1}^{\infty} \boldsymbol{A}_k(\boldsymbol{A}_k + \lambda_2 \boldsymbol{I})^{-1} \psi_k \otimes \psi_k =: \sum_{k=1}^{\infty} \boldsymbol{B}_k \psi_k \otimes \psi_k,$$

where $\boldsymbol{B}_k = \boldsymbol{R}(\boldsymbol{R} + \vartheta_k \boldsymbol{I})^{-1}$ and $\vartheta_k = \lambda_2/\nu_k \to \infty$ as $k \to \infty$. Writing $\boldsymbol{B}_k = \{B_{k,jj'}\}_{j,j'=1}^{q}$, it follows that

$$\left|\left|\left| \mathscr{T}^{(\mathscr{S},\mathscr{S})}(\mathscr{T}_{\lambda_2}^{(\mathscr{S},\mathscr{S})})^{-1} \right|\right|\right|_{\infty,\infty} \leq \max_{1\leq j\leq q} \sum_{j'=1}^{q} \max_k |B_{k,jj'}|. \tag{S.3}$$

In Section S.3 of the supplementary material, we examine two specific scenarios where $\boldsymbol{R}$ is either a $MA(1)$ or $AR(1)$ correlation matrix. We find that the upper bound of $\varkappa(\lambda_2)$ is equal to some constant independent of λ_2 and the true signal size q. Furthermore, we find that Condition C.4 holds for all legitimate $MA(1)$ correlation matrices and for $AR(1)$ correlation matrices characterized by an autoregressive coefficient less than 1/3.

S.2. Technical Proofs

S.2.1 Proof of Propositions

Proof of Proposition 1

Proof. Rewrite the minimization function (2.4),

$$\ell(\boldsymbol{f}) := \frac{1}{2n}\sum_{i=1}^{n}\left(Y_i - \sum_{j=1}^{p}\langle \widetilde{X}_{ij}, f_j\rangle_2\right)^2 + \lambda_1 \sum_{j=1}^{p}\|\Psi_j f_j\|_2 + \frac{\lambda_2}{2}\sum_{j=1}^{p}\|f_j\|_2^2.$$

The minimizer $\widetilde{f}_j(t)$ can always be represented in the form

$$\widetilde{f}_j(t) = \widehat{f}_j(t) + \eta_j(t),$$

where $\widehat{f}_j(\cdot) = \sum_{i=1}^n c_{ij}\widetilde{X}_{ij}(\cdot) \in \mathbb{M}_{nj}$, and $\eta_j(\cdot) \in \mathbb{M}_{nj}^{\perp}$. Therefore, we have $\langle \widetilde{X}_{ij}, \widetilde{f}_j\rangle_2 = \langle \widetilde{X}_{ij}, \widehat{f}_j\rangle_2$, $\|\widetilde{f}_j\|_2^2 = \|\widehat{f}_j\|_2^2 + \|\eta_j\|_2^2$, and $\|\Psi_j\widetilde{f}_j\|_2^2 = \|\Psi_j\widehat{f}_j\|_2^2 + \|\Psi_j\eta_j\|_2^2$. The last equation holds by Condition (C.1). Therefore, $\widetilde{f}_j(t)$ is the minimizer when $\eta_j \equiv 0$. □

Proof of Proposition 2

Proof. The KKT condition (2.5) follows readily from Propositions 5 and 6. We can show the existence of functional KKT solution by showing that the minimizer of (S.1) exists. Note that (S.1) can be reformulated as a constrained quadratic programming problem:

$$\min_{\boldsymbol{f}} \{\ell_1(\boldsymbol{f}) + \ell_2(\boldsymbol{f})\} \text{ such that } \ell_3(\boldsymbol{f}) \leq C_1 \text{ and } \ell_4(\boldsymbol{f}) \leq C_2.$$

where (C_1, C_2) here have a one-to-one correspondence with the regularization parameters (λ_1, λ_2) via the Lagrangian duality. It follows from Proposition 1 that the solution can be found in a finite-dimensional subspace. Therefore, the above minimization problem involves a continuous finite-dimensional quadratic objective function over a compact set. By Weierstrass' extreme value theorem, the minimum is always achieved. To show uniqueness, first note that there is either a unique solution or an (uncountably) infinite number of solutions. This is because if $\boldsymbol{f}_1$ and $\boldsymbol{f}_2$ are two minimizers, then by convexity $\ell(\alpha \boldsymbol{f}_1 + (1-\alpha)\boldsymbol{f}_2) \leq \alpha\ell(\boldsymbol{f}_1) + (1-\alpha)\ell(\boldsymbol{f}_2)$, and hence

$$\ell(\alpha \boldsymbol{f}_1 + (1-\alpha)\boldsymbol{f}_2) = \ell(\boldsymbol{f}_1) = \ell(\boldsymbol{f}_2) \text{ for all } \alpha \in (0,1). \tag{S.4}$$

If $\boldsymbol{f}_1 \neq \boldsymbol{f}_2$, then by the strict convexity of ℓ_3 we have $\ell_3(\alpha \boldsymbol{f}_1 + (1-\alpha)\boldsymbol{f}_2) < \alpha\ell_3(\boldsymbol{f}_1) + (1-\alpha)\ell_3(\boldsymbol{f}_2)$. Since ℓ_1, ℓ_2 and ℓ_4 are all convex and in view of (S.1), the relationsip (S.4) cannot hold. Thus, $\boldsymbol{f}_1 = \boldsymbol{f}_2$.

□

Proof of Proposition 3

Proof. Write the spectral decomposition of $\mathscr{T}^{(j,j)}$ as $\mathscr{T}^{(j,j)} = \sum_{k\geq 1} \nu_{jk}\eta_{jk}\otimes\eta_{jk}$ where $\{\nu_{jk}\}_{k\geq 1}$ are the eigenvalues of $\mathscr{T}^{(j,j)}$ in decreasing order, and $\{\eta_{jk}\}_{k\geq 1}$ are the corresponding eigenfunctions. Define

$$\mathscr{T}_m^{(j,j)} = \Pi_{j,m}\mathscr{T}^{(j,j)}\Pi_{j,m} = \sum_{k=1}^{m} \nu_{jk}\eta_{jk}\otimes\eta_{jk}$$

where $\Pi_{j,m} = \sum_{k=1}^{m} \eta_{jk}\otimes\eta_{jk}$ is the projection operator onto the m-dimensional principal components of $\mathscr{T}^{(j,j)}$. Recall that $\boldsymbol{\mathscr{Q}}^{(\mathscr{S},\mathscr{S})} = \mathrm{diag}(\mathscr{T}^{(j,j)})_{1\leq j\leq q}$. It is straightforward that

$$\boldsymbol{\mathscr{Q}}_{\alpha,m}^{(\mathscr{S},\mathscr{S})} = \mathbf{\Pi}_m \boldsymbol{\mathscr{Q}}_{\alpha}^{(\mathscr{S},\mathscr{S})} \mathbf{\Pi}_m = \mathbf{\Pi}_m \left(\boldsymbol{\mathscr{Q}}^{(\mathscr{S},\mathscr{S})} + \alpha \boldsymbol{\mathscr{I}} \right) \mathbf{\Pi}_m,$$

where $\mathbf{\Pi}_m = \text{diag}(\Pi_{j,m})_{1 \le j \le q}$. We know that $\boldsymbol{\mathscr{Q}}_{\alpha,m}^{(\mathscr{S},\mathscr{S})} \to \boldsymbol{\mathscr{Q}}_{\alpha}^{(\mathscr{S},\mathscr{S})}$ as $m \to \infty$. Define

$$\boldsymbol{\mathscr{T}}_{\alpha,m}^{(\mathscr{S},\mathscr{S})} = \mathbf{\Pi}_m \boldsymbol{\mathscr{T}}_{\alpha}^{(\mathscr{S},\mathscr{S})} \mathbf{\Pi}_m = \mathbf{\Pi}_m \left(\boldsymbol{\mathscr{T}}^{(\mathscr{S},\mathscr{S})} + \alpha \boldsymbol{\mathscr{I}} \right) \mathbf{\Pi}_m.$$

Note that

$$\begin{aligned} \mathscr{T}^{(j_1,j_2)} = \mathscr{T}_m^{(j_1,j_2)} &+ \mathbb{E}\left(\Pi_{j_1,m} \widetilde{X}_{j_1} \otimes \Pi_{j_2,m}^c \widetilde{X}_{j_2} \right) + \mathbb{E}\left(\Pi_{j_1,m}^c \widetilde{X}_{j_1} \otimes \Pi_{j_2,m} \widetilde{X}_{j_2} \right) \\ &+ \mathbb{E}\left(\Pi_{j_1,m}^c \widetilde{X}_{j_1} \otimes \Pi_{j_2,m}^c \widetilde{X}_{j_2} \right), \end{aligned} \tag{S.5}$$

where $\Pi_{j,m}^c = \sum_{k>m} \eta_{jk} \otimes \eta_{jk}$. By Cauchy–Schwarz inequality, for any $f_1, f_2 \in \mathbb{L}_2$,

$$\mathbb{E}\left| \left\langle \Pi_{j_1,m} \widetilde{X}_{j_1}, f_1 \right\rangle_2 \left\langle \Pi_{j_2,m}^c \widetilde{X}_{j_2}, f_2 \right\rangle_2 \right| \le \left\| \mathscr{T}_m^{(j_1,j_1)} f_1 \right\|_2 \left\| \left(\mathscr{T}^{(j_2,j_2)} - \mathscr{T}_m^{(j_2,j_2)} \right) f_2 \right\|_2.$$

As m approaches infinity, the second term on the right-hand side of (S.5) converges to 0. Similarly, the third and fourth terms also converge to 0. As a result, we show that $\mathscr{T}_{\alpha,m}^{(\mathscr{S},\mathscr{S})} \to \mathscr{T}_{\alpha}^{(\mathscr{S},\mathscr{S})}$ as $m \to \infty$.

Note that $\mathscr{T}_{\alpha,m}^{(\mathscr{S},\mathscr{S})}$ and $\mathscr{Q}_{\alpha,m}^{(\mathscr{S},\mathscr{S})}$ have one-to-one mapping to a vector space of at most mq dimensions. According to Lu and Pearce (2000), there exists a relationship between the eigenvalues of $\mathscr{T}_{\alpha,m}^{(\mathscr{S},\mathscr{S})}$ and $\mathscr{Q}_{\alpha,m}^{(\mathscr{S},\mathscr{S})}$ as follows:

$$\Lambda_k\left(\mathscr{T}_{\alpha,m}^{(\mathscr{S},\mathscr{S})}\right) \leq \Lambda_k\left(\mathscr{Q}_{\alpha,m}^{(\mathscr{S},\mathscr{S})}\right) \left|\!\left|\!\left| \left(\mathscr{Q}_{\alpha,m}^{(\mathscr{S},\mathscr{S})}\right)^{-1/2} \mathscr{T}_{\alpha,m}^{(\mathscr{S},\mathscr{S})} \left(\mathscr{Q}_{\alpha,m}^{(\mathscr{S},\mathscr{S})}\right)^{-1/2} \right|\!\right|\!\right|_{2,2}.$$

By the definition of operator norm,

$$\begin{aligned}
&\left|\!\left|\!\left| \left(\mathscr{Q}_{\alpha,m}^{(\mathscr{S},\mathscr{S})}\right)^{-1/2} \mathscr{T}_{\alpha,m}^{(\mathscr{S},\mathscr{S})} \left(\mathscr{Q}_{\alpha,m}^{(\mathscr{S},\mathscr{S})}\right)^{-1/2} \right|\!\right|\!\right|_{2,2} \\
&= \left|\!\left|\!\left| \mathbf{\Pi}_m \left(\mathscr{Q}_{\alpha}^{(\mathscr{S},\mathscr{S})}\right)^{-1/2} \mathbf{\Pi}_m \mathscr{T}_{\alpha}^{(\mathscr{S},\mathscr{S})} \mathbf{\Pi}_m \left(\mathscr{Q}_{\alpha}^{(\mathscr{S},\mathscr{S})}\right)^{-1/2} \mathbf{\Pi}_m \right|\!\right|\!\right|_{2,2} \\
&\leq \left|\!\left|\!\left| \left(\mathscr{Q}_{\alpha}^{(\mathscr{S},\mathscr{S})}\right)^{-1/2} \mathscr{T}_{\alpha}^{(\mathscr{S},\mathscr{S})} \left(\mathscr{Q}_{\alpha}^{(\mathscr{S},\mathscr{S})}\right)^{-1/2} \right|\!\right|\!\right|_{2,2} \\
&\leq b.
\end{aligned}$$

The last inequality holds due to Condition C.5. Finally, let $m \to \infty$ and $\alpha \to 0$, we have

$$\Lambda_k\left(\mathscr{T}_{\alpha,m}^{(\mathscr{S},\mathscr{S})}\right) \to \Lambda_k\left(\mathscr{T}^{(\mathscr{S},\mathscr{S})}\right), \quad \Lambda_k\left(\mathscr{Q}_{\alpha,m}^{(\mathscr{S},\mathscr{S})}\right) \to \Lambda_k\left(\mathscr{Q}^{(\mathscr{S},\mathscr{S})}\right).$$

□

Proof of Proposition 4

Proof. The convex program (4.10) can be reformulated as a constrained quadratic program

$$\min_{\boldsymbol{d}_j \in \mathbb{R}^{M_j}} \left\{ \frac{1}{2} \boldsymbol{d}_j^\top \boldsymbol{\Omega}_j \boldsymbol{d}_j - \boldsymbol{\varrho}_j^\top \boldsymbol{d}_j \right\}, \quad \text{such that} \quad \|\boldsymbol{d}_j\|_2 \leq C_1,$$

where the regularization parameter λ_1 and constraint level C_1 are in one-to-one correspondence via Lagrangian duality. As a result, the above minimization problem involves a continuous finite-dimensional quadratic objective function over a compact set. The Weierstrass' extreme value theorem guarantees that the minimum is always achieved. According to the Karush-Kuhn-Tucker (KKT) condition to (4.10)

$$\boldsymbol{\Omega}_j \boldsymbol{d}_j - \boldsymbol{\varrho}_j + \lambda_1 \boldsymbol{r}_j = \mathbf{0}, \tag{S.6}$$

where $\boldsymbol{r}_j$ denotes the sub-gradient of $\|\boldsymbol{d}_j\|_2$ such that $\|\boldsymbol{r}_j\|_2 \leq 1$ and $\boldsymbol{r}_j = \|\boldsymbol{d}_j\|_2^{-1} \boldsymbol{d}_j$ holds for $\boldsymbol{d}_j \neq 0$. When $\|\boldsymbol{\varrho}_j\|_2 \leq \lambda_1$, suppose $\boldsymbol{d}_j \neq 0$, according to (S.6), we have

$$\lambda_1 + \Lambda_{\min}(\boldsymbol{\Omega}_j) \|\boldsymbol{d}_j\|_2 \leq \|\boldsymbol{\varrho}_j\|_2 \leq \lambda_1 + \Lambda_{\max}(\boldsymbol{\Omega}_j) \|\boldsymbol{d}_j\|_2,$$

where $\Lambda_{\min}(\boldsymbol{\Omega}_j)$ and $\Lambda_{\max}(\boldsymbol{\Omega}_j)$ represent the smallest and largest eigenvalues of the $\boldsymbol{\Omega}_j$, respectively. In order words, when $\|\boldsymbol{\varrho}_j\|_2 \leq \lambda_1$, we must have $\boldsymbol{d}_j = 0$. On the other hand, when $\|\boldsymbol{\varrho}_j\|_2 > \lambda_1$, suppose $\boldsymbol{d}_j = 0$, according to (S.6), we have $\boldsymbol{\varrho} = \lambda_1 \boldsymbol{r}_j$, and hence $\|\boldsymbol{\varrho}_j\|_2 \leq \lambda_1$. This statement presents a contradiction, therefore,

$$\begin{cases} \boldsymbol{d}_j = 0, & \text{if } \|\boldsymbol{\varrho}_j\|_2 \leq \lambda_1, \\ \boldsymbol{d}_j \neq 0, & \text{if } \|\boldsymbol{\varrho}_j\|_2 > \lambda_1. \end{cases}$$

□

Proof of Proposition 5

Proof. To begin with, assume $\mathscr{J}$ is convex and Gateaux differentiable. Suppose $\mathscr{J}(f) \geq \mathscr{J}(g) + \mathscr{D}_{\mathscr{J}}(g; f-g)$ for all $f, g \in \mathbb{H}$. Define $h = \lambda f + (1-\lambda)g$, then $\mathscr{J}(f) \geq \mathscr{J}(h) + \mathscr{D}_{\mathscr{J}}(h; f-h)$ and $\mathscr{J}(g) \geq \mathscr{J}(h) + \mathscr{D}_{\mathscr{J}}(h; g-h)$, by the linear combination of the two inequalities, we have:

$$\lambda \mathscr{J}(f) + (1-\lambda)\mathscr{J}(g) \geq \mathscr{J}(h) + \mathscr{D}_{\mathscr{J}}(h; 0) = \mathscr{J}(\lambda f + (1-\lambda)g),$$

which shows convexity. On the other hand, by convexity, for all $f, g \in \mathbb{H}$,

$\lambda \in (0,1)$, we have

$$\mathscr{J}(f) - \mathscr{J}(g) \geq \frac{\mathscr{J}(g + \lambda(f-g)) - \mathscr{J}(g)}{\lambda},$$

let $\lambda \downarrow 0^+$, then the right-hand side will go to $\mathscr{D}_{\mathscr{J}}(g; f-g)$.

To find the global minimum of $\mathscr{J}(\cdot)$, suppose $\mathscr{D}_{\mathscr{J}}(g;\psi) = 0$ for all $\psi \in \mathbb{H}$, then $\mathscr{J}(g) \leq \mathscr{J}(f)$ for all $f \in \mathbb{H}$. On the other hand, by setting $f_1 = g + \tau\psi$, $f_2 = g - \tau\psi$, we have

$$\frac{\mathscr{J}(g) - \mathscr{J}(g - \tau\psi)}{\tau} \leq \mathscr{D}_{\mathscr{J}}(g;\psi) \leq \frac{\mathscr{J}(g+\tau\psi) - \mathscr{J}(g)}{\tau}.$$

suppose $\mathscr{J}(g)$ is the global minimum, the left side is smaller than 0 and the right side is greater than 0. By the definition of Gateaux differentiability, the limits on both sides exist and are equal when $\tau \to 0$. Therefore, $\mathscr{D}_{\mathscr{J}}(g;\psi) = 0$ for all ψ.

Now assume $\mathscr{J}$ is convex but not Gateaux differentiable. Then we can easily show $\mathscr{J}(g)$ is the global minimum of $\mathscr{J}$ if and only if $0 \in \partial_{\mathscr{J}(g)}$ using a similar derivation as above.

$\square$

S.2.2 Proofs of Theorems and Corollary

Proof for Theorem 1

Recall that $\widehat{\boldsymbol{f}}$ is the solution of KKT condition (2.5) and $\widehat{\mathscr{S}} = \{i \in \{1,\ldots,p\} : \widehat{f}_i \neq 0\}$. Write $\widetilde{\boldsymbol{X}}_n = (\widetilde{\boldsymbol{X}}_{\mathscr{S}}, \widetilde{\boldsymbol{X}}_{\mathscr{S}^c})$ by grouping the columns in $\mathscr{S}$ and $\mathscr{S}^c$. For $j \in \mathscr{S}^c$, in the scenario where $\mathscr{T}^{(j,j)}$ possesses finitely many nonzero eigenvalues, there exist infinitely many f_j such that $\langle f_j, \mathscr{T}^{(j,j)} f_j \rangle_2 = 0$, and those f_j do not make contributions to the response. Without loss of generality, we assume that $\boldsymbol{f}_{0\mathscr{S}^c} = \mathbf{0}$, and we have $\boldsymbol{f}_0 = (\boldsymbol{f}_{0\mathscr{S}}^\top, \mathbf{0}^\top)^\top$. Similarly, partition $\widehat{\boldsymbol{f}} = (\widehat{\boldsymbol{f}}_{\mathscr{S}}^\top, \widehat{\boldsymbol{f}}_{\mathscr{S}^c}^\top)^\top$, $\boldsymbol{g}_n = (\boldsymbol{g}_{\mathscr{S}}^\top, \boldsymbol{g}_{\mathscr{S}^c}^\top)^\top$ and $\boldsymbol{\omega} = (\boldsymbol{\omega}_{\mathscr{S}}^\top, \boldsymbol{\omega}_{\mathscr{S}^c}^\top)^\top$. With the partitions defined above and those in Section 3, the KKT condition in (2.5) can be rewritten as

$$\begin{pmatrix} \boldsymbol{\mathscr{T}}_n^{(\mathscr{S},\mathscr{S})} & \boldsymbol{\mathscr{T}}_n^{(\mathscr{S},\mathscr{S}^c)} \\ \boldsymbol{\mathscr{T}}_n^{(\mathscr{S}^c,\mathscr{S})} & \boldsymbol{\mathscr{T}}_n^{(\mathscr{S}^c,\mathscr{S}^c)} \end{pmatrix} \begin{pmatrix} \widehat{\boldsymbol{f}}_{\mathscr{S}} - \boldsymbol{f}_{0\mathscr{S}} \\ \widehat{\boldsymbol{f}}_{\mathscr{S}^c} \end{pmatrix} - \begin{pmatrix} \boldsymbol{g}_{\mathscr{S}} \\ \boldsymbol{g}_{\mathscr{S}^c} \end{pmatrix} + \lambda_2 \begin{pmatrix} \widehat{\boldsymbol{f}}_{\mathscr{S}} \\ \widehat{\boldsymbol{f}}_{\mathscr{S}^c} \end{pmatrix} + \lambda_1 \begin{pmatrix} \boldsymbol{\omega}_{\mathscr{S}} \\ \boldsymbol{\omega}_{\mathscr{S}^c} \end{pmatrix} = \mathbf{0}. \tag{S.7}$$

Proof of (i) of Theorem 1

To utilize the Primal-Dual Witness argument in Wainwright (2009), let $\check{\boldsymbol{f}}_{\mathscr{S}}$ be the solution of the functional elastic-net problem knowing the true signal set $\mathscr{S}$. In other words, $\check{\boldsymbol{f}}_{\mathscr{S}}$ is the value of $\boldsymbol{f}_{\mathscr{S}}$ that minimizes

$$\frac{1}{2}\left\langle \mathscr{T}_n^{(\mathscr{S},\mathscr{S})}(\boldsymbol{f}_{\mathscr{S}} - \boldsymbol{f}_{0\mathscr{S}}), \boldsymbol{f}_{\mathscr{S}} - \boldsymbol{f}_{0\mathscr{S}}\right\rangle_2 - \langle \boldsymbol{g}_{\mathscr{S}}, \boldsymbol{f}_{\mathscr{S}} - \boldsymbol{f}_{0\mathscr{S}}\rangle_2 + \sum_{j\in\mathscr{S}} \mathrm{Pen}(f_j; \lambda_1, \lambda_2).$$

Using similar arguments as for Proposition 2 ,

$$\mathscr{T}_n^{(\mathscr{S},\mathscr{S})}(\check{\boldsymbol{f}}_{\mathscr{S}} - \boldsymbol{f}_{0\mathscr{S}}) - \boldsymbol{g}_{\mathscr{S}} + \lambda_2 \check{\boldsymbol{f}}_{\mathscr{S}} + \lambda_1 \boldsymbol{\omega}_{\mathscr{S}} = 0, \tag{S.8}$$

where $\boldsymbol{\omega}_{\mathscr{S}} = (\Psi_j \eta_j, j \in \mathscr{S})$ is the functional subgradient of ℓ_4 for this problem described in Proposition 2 and 6. For convenience, let $\boldsymbol{\eta}_{\mathscr{W}} = (\eta_j, j \in \mathscr{W})$ for any set $\mathscr{W}$. By Proposition 2, the solution to the functional elastic-net problem is unique and satisfies the KKT equation (S.7). If we can show that $\left(\check{\boldsymbol{f}}_{\mathscr{S}}^\top, \mathbf{0}^\top\right)^\top$ solves (S.7), then $\widehat{\boldsymbol{f}} = \left(\check{\boldsymbol{f}}_{\mathscr{S}}^\top, \mathbf{0}^\top\right)^\top$ and $\widehat{\mathscr{S}} \subset \mathscr{S}$. It remains to show

$$\mathscr{T}_n^{(\mathscr{S}^c,\mathscr{S})}(\check{\boldsymbol{f}}_{\mathscr{S}} - \boldsymbol{f}_{0\mathscr{S}}) - \boldsymbol{g}_{\mathscr{S}^c} + \lambda_1 \boldsymbol{\omega}_{\mathscr{S}^c} = 0, \tag{S.9}$$

for some $\boldsymbol{\omega}_{\mathscr{S}^c}$ satisfying $\boldsymbol{\omega}_{\mathscr{S}^c} = (\Psi_j \eta_j, j \in \mathscr{S}^c)$ where $\|\boldsymbol{\eta}_{\mathscr{S}^c}\|_\infty \le 1$. However, by (S.8),

$$\check{\boldsymbol{f}}_{\mathscr{S}} - \boldsymbol{f}_{0\mathscr{S}} = \left(\mathscr{T}_{n,\lambda_2}^{(\mathscr{S},\mathscr{S})}\right)^{-1} \left(\boldsymbol{g}_{\mathscr{S}} - \lambda_2 \boldsymbol{f}_{0\mathscr{S}} - \lambda_1 \boldsymbol{\omega}_{\mathscr{S}}\right), \tag{S.10}$$

and hence, upon combining (S.9) and (S.10), any $\boldsymbol{\omega}_{\mathscr{S}^c}$ that solves (S.9) must satisfy

$$\begin{aligned} \boldsymbol{\omega}_{\mathscr{S}^c} := \frac{1}{\lambda_1} \Bigg\{ & \boldsymbol{g}_{\mathscr{S}^c} - \mathscr{T}_n^{(\mathscr{S}^c,\mathscr{S})} \left(\mathscr{T}_{n,\lambda_2}^{(\mathscr{S},\mathscr{S})} \right)^{-1} \boldsymbol{g}_{\mathscr{S}} \Bigg\} \\ & + \mathscr{T}_n^{(\mathscr{S}^c,\mathscr{S})} \left(\mathscr{T}_{n,\lambda_2}^{(\mathscr{S},\mathscr{S})} \right)^{-1} \left(\frac{\lambda_2}{\lambda_1} \boldsymbol{f}_{0\mathscr{S}} + \boldsymbol{\omega}_{\mathscr{S}} \right). \end{aligned} \tag{S.11}$$

By Condition C.1, the existence of $\boldsymbol{\omega}_{\mathscr{S}^c}$ satisfying (S.9) is guaranteed by

$$\|\boldsymbol{\omega}_{\mathscr{S}^c}\|_\infty \leq C_{\min}. \tag{S.12}$$

The rest of this subsection will be focusing on (S.12).

It is easy to see that, for any $\boldsymbol{f} \in \mathbb{L}_2^q[0,1]$, $(\mathscr{T}_n^{(\mathscr{S},\mathscr{S})} \boldsymbol{f})(t) = \frac{1}{n} \int \widetilde{\boldsymbol{X}}_{\mathscr{S}}^\top(t) \widetilde{\boldsymbol{X}}_{\mathscr{S}}(u) \boldsymbol{f}(u) du$. The first term on the right-hand side of (S.11) can be rewritten as $(\lambda_1 n)^{-1} \widetilde{\boldsymbol{X}}_{\mathscr{S}^c}^\top (\mathrm{I} - \boldsymbol{\Delta}_n) \boldsymbol{\varepsilon}_n$, where

$$\boldsymbol{\Delta}_n = \frac{1}{n} \int \widetilde{\boldsymbol{X}}_{\mathscr{S}}(u) \left\{ \left(\mathscr{T}_{n,\lambda_2}^{(\mathscr{S},\mathscr{S})} \right)^{-1} \widetilde{\boldsymbol{X}}_{\mathscr{S}}^\top \right\} (u) du. \tag{S.13}$$

Thus, for all $j \in \mathscr{S}^c$,

$$\begin{aligned}\|\omega_j\|_2 &= \left\| \frac{\sigma}{\lambda_1 n} \widetilde{\boldsymbol{X}}_{\bullet j}^{\top} \left(\mathrm{I} - \boldsymbol{\Delta}_n \right) \boldsymbol{z}_n + \mathscr{T}_n^{(j,\mathscr{S})} \left(\mathscr{T}_{n,\lambda_2}^{(\mathscr{S},\mathscr{S})} \right)^{-1} \left(\frac{\lambda_2}{\lambda_1} \boldsymbol{f}_{0\mathscr{S}} + \boldsymbol{\omega}_{\mathscr{S}} \right) \right\|_2 \\ &\leq \left\| \frac{\sigma}{\lambda_1 n} \widetilde{\boldsymbol{X}}_{\bullet j}^{\top} \left(\mathrm{I} - \boldsymbol{\Delta}_n \right) \boldsymbol{z}_n \right\|_2 + \left\| \mathscr{T}_n^{(j,\mathscr{S})} \left(\mathscr{T}_{n,\lambda_2}^{(\mathscr{S},\mathscr{S})} \right)^{-1} \left(\frac{\lambda_2}{\lambda_1} \boldsymbol{f}_{0\mathscr{S}} + \boldsymbol{\omega}_{\mathscr{S}} \right) \right\|_2,\end{aligned} \tag{S.14}$$

where $\boldsymbol{z}_n = \sigma^{-1} \boldsymbol{\varepsilon}_n$ has covariance matrix equal to an identity matrix. If $\widehat{\mathscr{S}} \not\subset \mathscr{S}$ then (S.12) fails, and, by Lemmas S.2 and S.3 below,

$$\begin{aligned}\mathbb{P}\left(\widehat{\mathscr{S}} \not\subset \mathscr{S} \right) &\leq \mathbb{P}\left(\|\boldsymbol{\omega}_{\mathscr{S}^c}\|_\infty > \left(1 - \frac{\gamma}{9}\right) C_{\min} \right) \\ &\leq \exp\left(-D^{(1)} \lambda_1^2 n \right) + \exp\left(-D^{(2)} \frac{\lambda_2^2 n}{q} \right).\end{aligned} \tag{S.15}$$

Note that $\exp\left(-D^{(1)}\lambda_1^2 n\right) \leq \exp\left(-D^{(1)} C_{\max}^{-2} q \frac{\lambda_2^2 n}{q}\right)$ since $\lambda_1 > C_{\max}^{-1} \lambda_2$. Applying Lemma S.1 with $\epsilon = 1/2$, we can bound the rhs of (S.15) by the probability in (3.7), provided $\lambda_2^2 n/q > (2 \log 2) D^{-1}$, which is guaranteed by Condition (3.6) for sufficiently large $D_{2,2}^*$ in the condition.

To conclude the proof of part (i) of Theorem 1, it remains to establish the following three lemmas, the proofs of which are in the Supplemental Material.

Lemma S.1. *For $a_k, b_k > 0$, $k = 1, \ldots, K$,*

$$\sum_{k=1}^{K} a_k \exp(-b_k x) \leq \exp\left\{ -(1 - \epsilon) b x \right\}$$

for $x > (\epsilon b)^{-1}\log(Ka)$*, where* $\epsilon \in (0,1)$*,* $a = \max_k a_k$ *and* $b = \min_k b_k$*.*

Lemma S.2. *Let* γ *be as in Condition C.3. Suppose* $\lambda_1 > D_1^*(\sigma+1)\tau^{1/2}(C_{\min}\gamma)^{-1}\sqrt{\frac{\log(p-q)}{n}}$ *for some constant* D_1^**. We have*

$$\mathbb{P}\left(\max_{j\in\mathscr{S}^c}\left\|\frac{\sigma}{\lambda_1 n}\widetilde{\boldsymbol{X}}_{\bullet j}^{\top}\left(\mathrm{I}-\boldsymbol{\Delta}_n\right)\boldsymbol{z}_n\right\|_2 \geq \frac{\gamma C_{\min}}{9}\right) \leq \exp\left(-D^{(1)}\lambda_1^2 n\right)$$

where $D^{(1)} = D_2^* C_{\min}^2\gamma^2(\sigma+1)^{-2}\tau^{-1}$ *and* D_2^* *is a universal constant.*

Lemma S.3. *Let* γ *be as in Condition C.3. Suppose, for some constant* D_1^**,*

$$\lambda_2 > D_1^*\frac{\tau(\rho_1+1)}{(C_{\min}/C_{\max})^2\gamma^2}\max\left(\frac{q\log(p-q)}{n},\sqrt{\frac{q^2}{n}}\right) \quad \text{and} \quad \frac{\lambda_1}{\lambda_2} > \left(\frac{3}{\gamma}-2\right)C_{\max}^{-1}.$$

Then

$$\mathbb{P}\left\{\max_{j\in\mathscr{S}^c}\left\|\boldsymbol{\mathscr{T}}_n^{(j,\mathscr{S})}\left(\boldsymbol{\mathscr{T}}_{n,\lambda_2}^{(\mathscr{S},\mathscr{S})}\right)^{-1}\left(\frac{\lambda_2}{\lambda_1}\boldsymbol{f}_{0\mathscr{S}}+\boldsymbol{\omega}_{\mathscr{S}}\right)\right\|_2 \geq \left(1-\frac{2\gamma}{9}\right)C_{\min}\right\} \leq \exp\left(-D^{(2)}\frac{\lambda_2^2 n}{q}\right)$$

where $D^{(2)} = D_2^*(C_{\min}/C_{\max})^2\gamma^2(\rho_1+1)^{-2}\tau^{-1}$ *and* D_2^* *is a universal constant.*

Proof of (ii) of Theorem 1

We need to show that $\|\widehat{f}_j\|_2 > 0$ for all $j \in \mathscr{S}_G$ with the probability lower bound stated in the theorem. For simplicity, assume that $\mathscr{S}_G = \mathscr{S}$. The same arguments hold if $\mathscr{S}$ is replaced by $\mathscr{S}_G$ below.

Note that $\mathbb{P}(\widehat{\mathscr{S}} \supset \mathscr{S}) = \mathbb{P}(\min_{j\in\mathscr{S}}\|\widehat{f}_j\|_2 > 0) \geq \mathbb{P}(\min_{j\in\mathscr{S}}\|(\mathscr{T}^{(j,j)})^{1/2}\widehat{f}_j\|_2 >$

0). By the triangle inequality,

$$\min_{j\in\mathscr{S}}\left\|(\mathscr{T}^{(j,j)})^{1/2}\widehat{f}_j\right\|_2 \geq \min_{j\in\mathscr{S}}\left\|(\mathscr{T}^{(j,j)})^{1/2}f_{0j}\right\|_2 - \max_{j\in\mathscr{S}}\left\|(\mathscr{T}^{(j,j)})^{1/2}(\widehat{f}_j - f_{0j})\right\|_2$$
$$\geq G - \max_{j\in\mathscr{S}}\left\|(\mathscr{T}^{(j,j)})^{1/2}(\widehat{f}_j - f_{0j})\right\|_2.$$

Thus, it suffices to provide an upper bound for $\mathbb{P}\left(\max_{j\in\mathscr{S}}\|(\mathscr{T}^{(j,j)})^{1/2}(\widehat{f}_j - f_{0j})\|_2 > G\right)$.

By (S.10), we have

$$\begin{aligned}\check{\boldsymbol{f}}_{\mathscr{S}} - \boldsymbol{f}_{0\mathscr{S}} =& (\boldsymbol{\mathscr{T}}_{\lambda_2}^{(\mathscr{S},\mathscr{S})})^{-1}\left(\boldsymbol{g}_{\mathscr{S}} - \lambda_2\boldsymbol{f}_{0\mathscr{S}} - \lambda_1\boldsymbol{\omega}_{\mathscr{S}}\right)\\ &+ \left\{(\boldsymbol{\mathscr{T}}_{n,\lambda_2}^{(\mathscr{S},\mathscr{S})})^{-1} - (\boldsymbol{\mathscr{T}}_{\lambda_2}^{(\mathscr{S},\mathscr{S})})^{-1}\right\}\left(\boldsymbol{g}_{\mathscr{S}} - \lambda_2\boldsymbol{f}_{0\mathscr{S}} - \lambda_1\boldsymbol{\omega}_{\mathscr{S}}\right).\end{aligned}$$

Since $(\boldsymbol{\mathscr{T}}_{n,\lambda_2}^{(\mathscr{S},\mathscr{S})})^{-1} - (\boldsymbol{\mathscr{T}}_{\lambda_2}^{(\mathscr{S},\mathscr{S})})^{-1} = (\boldsymbol{\mathscr{T}}_{\lambda_2}^{(\mathscr{S},\mathscr{S})})^{-1}\left(\boldsymbol{\mathscr{T}}^{(\mathscr{S},\mathscr{S})} - \boldsymbol{\mathscr{T}}_n^{(\mathscr{S},\mathscr{S})}\right)(\boldsymbol{\mathscr{T}}_{n,\lambda_2}^{(\mathscr{S},\mathscr{S})})^{-1}$,

$$\begin{aligned}&\max_{j\in\mathscr{S}}\left\|(\mathscr{T}^{(j,j)})^{1/2}(\check{f}_j - f_{0j})\right\|_2 = \left\|(\boldsymbol{\mathscr{Q}}^{(\mathscr{S},\mathscr{S})})^{1/2}(\check{\boldsymbol{f}}_{\mathscr{S}} - \boldsymbol{f}_{0\mathscr{S}})\right\|_\infty\\ &\leq \left|\!\left|\!\left|(\boldsymbol{\mathscr{Q}}^{(\mathscr{S},\mathscr{S})})^{1/2}(\boldsymbol{\mathscr{T}}_{\lambda_2}^{(\mathscr{S},\mathscr{S})})^{-1}\right|\!\right|\!\right|_{\infty,\infty}\left\{\|\boldsymbol{g}_{\mathscr{S}}\|_\infty + \lambda_2\left(\|\boldsymbol{f}_{0\mathscr{S}}\|_\infty + \frac{\lambda_1}{\lambda_2}C_{\max}\right)\right\}\\ &\quad\times\left(1 + \frac{\sqrt{q}}{\lambda_2}|\!|\!|\boldsymbol{\mathscr{T}}^{(\mathscr{S},\mathscr{S})} - \boldsymbol{\mathscr{T}}_n^{(\mathscr{S},\mathscr{S})}|\!|\!|_{2,2}\right),\end{aligned}$$

where we applied the inequality

$$\left|\!\left|\!\left|(\boldsymbol{\mathscr{T}}^{(\mathscr{S},\mathscr{S})} - \boldsymbol{\mathscr{T}}_n^{(\mathscr{S},\mathscr{S})})(\boldsymbol{\mathscr{T}}_{n,\lambda_2}^{(\mathscr{S},\mathscr{S})})^{-1}\right|\!\right|\!\right|_{\infty,\infty} \leq \frac{\sqrt{q}}{\lambda_2}|\!|\!|\boldsymbol{\mathscr{T}}^{(\mathscr{S},\mathscr{S})} - \boldsymbol{\mathscr{T}}_n^{(\mathscr{S},\mathscr{S})}|\!|\!|_{2,2}.$$

By Lemma S.4 with $\|\boldsymbol{f}_{0\mathscr{S}}\|_\infty = 1$,

$$\max_{j\in\mathscr{S}} \left\|(\mathscr{T}^{(j,j)})^{1/2}(\widehat{f}_j - f_{0j})\right\|_2 \\ \leq \frac{6-4\aleph(\lambda_2)}{(1-\aleph(\lambda_2))\sqrt{\lambda_2}}\left(\|\boldsymbol{g}_{\mathscr{S}}\|_\infty + \lambda_2 + C_{\max}\lambda_1\right)\left(1+\frac{\sqrt{q}}{\lambda_2}\left|\!\left|\!\left|\boldsymbol{\mathscr{T}}^{(\mathscr{S},\mathscr{S})} - \boldsymbol{\mathscr{T}}_n^{(\mathscr{S},\mathscr{S})}\right|\!\right|\!\right|_{2,2}\right). \tag{S.16}$$

Thus, with G as given in the theorem,

$$\mathbb{P}\left(\max_{j\in\mathscr{S}} \left\|(\mathscr{T}^{(j,j)})^{1/2}(\widehat{f}_j - f_{0j})\right\|_2 > G\right) \\ \leq \mathbb{P}\left(\|\boldsymbol{g}_{\mathscr{S}}\|_\infty > \lambda_2\right) + \mathbb{P}\left(\frac{\sqrt{q}}{\lambda_2}\left|\!\left|\!\left|\boldsymbol{\mathscr{T}}^{(\mathscr{S},\mathscr{S})} - \boldsymbol{\mathscr{T}}_n^{(\mathscr{S},\mathscr{S})}\right|\!\right|\!\right|_{2,2} > 1\right). \tag{S.17}$$

Finally, bound the rhs of (S.17) using Lemmas S.5 and S.6 and note that it is dominated by the expression in (3.7) under Condition (3.6).

Lemma S.4. *Under Condition C.4, for any $\lambda_2 > 0$*

$$\left|\!\left|\!\left|(\boldsymbol{\mathscr{Q}}^{(\mathscr{S},\mathscr{S})})^{1/2}(\boldsymbol{\mathscr{T}}_{\lambda_2}^{(\mathscr{S},\mathscr{S})})^{-1}\right|\!\right|\!\right|_{\infty,\infty} < \frac{6-4\aleph(\lambda_2)}{1-\aleph(\lambda_2)}\frac{1}{\sqrt{\lambda_2}} \tag{S.18}$$

Lemma S.5. *Suppose $\lambda_2 > D_1^*(\sigma+1)\tau^{1/2}\sqrt{\frac{\log q}{n}}$, we have*

$$\mathbb{P}\left(\|\boldsymbol{g}_{\mathscr{S}}\|_\infty \geq \lambda_2\right) \leq \exp\left(-D^{(3)}\lambda_2^2 n\right) \tag{S.19}$$

holds for some $D^{(3)} < D_2^*\left((\sigma+1)^2\tau\right)^{-1}$ where D_1^* and D_2^* are universal con-

stants.

Lemma S.6. *Suppose ρ_1 is the largest eigenvalue of $\mathscr{T}^{(\mathscr{S},\mathscr{S})}$, then*

$$\mathbb{P}\left(\sqrt{q}|||\mathscr{T}^{(\mathscr{S},\mathscr{S})} - \mathscr{T}_n^{(\mathscr{S},\mathscr{S})}|||_{2,2} > u\right) \leq \exp\left\{-\frac{u^2 n}{C^2\rho_1^2 q}\right\}$$

holds for some constant $C > 0$, as long as C and q satisfy

$$\frac{u^2}{C^2\rho_1^2} < q \leq \sqrt{\frac{u^2 n}{\tau C^2 \rho_1}}. \tag{S.20}$$

The proofs for Lemmas S.4 - S.6 are in the Supplementary Material.

Proof of Theorem 2

Proof. When $\widehat{\mathscr{S}} \subset \mathscr{S}$, the excess risk has the form

$$\mathscr{R}(\widehat{\boldsymbol{\beta}}) = \mathbb{E}^*\left[\sum_{j\in\mathscr{S}}\langle X_j^*, \beta_{0j} - \widehat{\beta}_j\rangle_2\right]^2 = \left\|\left(\mathscr{T}^{(\mathscr{S},\mathscr{S})}\right)^{1/2}(\boldsymbol{f}_{0\mathscr{S}} - \widehat{\boldsymbol{f}}_{\mathscr{S}})\right\|_2^2.$$

Following a similar derivation as in (S.16),

$$\begin{aligned}
&\left\|\left(\mathscr{T}^{(\mathscr{S},\mathscr{S})}\right)^{1/2}(\boldsymbol{f}_{0\mathscr{S}} - \widehat{\boldsymbol{f}}_{\mathscr{S}})\right\|_2 \\
&\leq \sqrt{q}\left|\left|\left|(\mathscr{T}^{(\mathscr{S},\mathscr{S})})^{1/2}(\mathscr{T}_{\lambda_2}^{(\mathscr{S},\mathscr{S})})^{-1}\right|\right|\right|_{2,2}\left\{\|\boldsymbol{g}_{\mathscr{S}}\|_\infty + \lambda_2\left(\|\boldsymbol{f}_{0\mathscr{S}}\|_\infty + \frac{\lambda_1}{\lambda_2}C_{\max}\right)\right\} \\
&\quad\cdot\left(1 + \frac{1}{\lambda_2}|||\mathscr{T}^{(\mathscr{S},\mathscr{S})} - \mathscr{T}_n^{(\mathscr{S},\mathscr{S})}|||_{2,2}\right).
\end{aligned}$$

Similar to (S.57),

$$\left\|\left\|\left\|(\mathscr{T}^{(\mathscr{S},\mathscr{S})})^{1/2}(\mathscr{T}_{\lambda_2}^{(\mathscr{S},\mathscr{S})})^{-1}\right\|\right\|\right\|_{2,2} \leq \frac{1}{2\sqrt{\lambda_2}},$$

together with a similar derivation as for (S.16) with $\|\boldsymbol{f}_{0\mathscr{S}}\|_\infty = 1$, we have

$$\mathscr{R}(\widehat{\boldsymbol{f}}) \leq q(2 + C_{\max}\lambda_1/\lambda_2)^2\lambda_2 = q\left(4C_{\max}\lambda_1 + 4\lambda_2 + C_{\max}^2\lambda_1^2/\lambda_2\right)$$

with probability greater than (3.7). $\square$

Proof of Corollary 1

Proof. Recall

$$\alpha(p, q, n) = \max\left(q, \sqrt{\log(p-q)}, \sqrt{q\log n}\right),$$

and define

$$\ell_n = Cq^{-1}\alpha^2(p, q, n)$$

where C is a large enough constant. Let

$$\lambda_2 = (\ell_n q/n)^{1/2} = C^{1/2}\frac{1}{\sqrt{n}}\alpha(p, q, n)$$

and $\lambda_1 = b\lambda_2$ for some suitable constant b. If $q^2\log(p-q) \leq n$ for large n, which is guaranteed by the assumption $q\alpha(p,q,n) = o(n^{1/2})$, we have for all large n,

$$\alpha(p,q,n) = \max\left(q, \frac{q\log(p-q)}{\sqrt{n}}, \sqrt{\log(p-q)}, \sqrt{q\log n}\right),$$

from which it is easy to see that (3.6) holds for b, C sufficiently large. Note that $\ell_n \geq C\log n$. By Theorem 2, the excess risk is bounded by a constant multiple of

$$\lambda_2 q = C^{1/2}\frac{q}{\sqrt{n}}\alpha(p,q,n)$$

where probability at least $1 - n^{-D}$ for some constant D. The claim of the corollary follows from the Borel-Cantelli Lemma by choosing a large enough C and hence $D > 1$. □

Proof of Theorem 3

Proof. Recall that the excess risk for an estimator $\widetilde{\boldsymbol{f}}_{\mathscr{S}}$ has the form

$$\mathscr{R}(\widetilde{\boldsymbol{f}}_{\mathscr{S}}) = \left\|\left(\boldsymbol{\mathscr{T}}^{(\mathscr{S},\mathscr{S})}\right)^{1/2}(\boldsymbol{f}_{0\mathscr{S}} - \widetilde{\boldsymbol{f}}_{\mathscr{S}})\right\|_2^2.$$

For any $\boldsymbol{f}_1, \boldsymbol{f}_2 \in \mathbb{L}_2^q$, define

$$\mathscr{D}(\boldsymbol{f}_1, \boldsymbol{f}_2) = \left\| \left(\mathscr{T}^{(\mathscr{S},\mathscr{S})}\right)^{1/2} (\boldsymbol{f}_1 - \boldsymbol{f}_2) \right\|_2,$$

which is a proper metric in $\mathbb{L}_2^q$. Write the spectral decomposition of $\mathscr{T}^{(\mathscr{S},\mathscr{S})}$ as $\mathscr{T}^{(\mathscr{S},\mathscr{S})} = \sum_{k\geq 1} \rho_k \boldsymbol{\phi}_k \otimes \boldsymbol{\phi}_k$, where $\rho_1 \geq \rho_2 \geq \cdots \geq 0$. By Corollary 2, for any covariance operator $\mathscr{T}^{(\mathscr{S},\mathscr{S})} \in \mathscr{P}(r)$, its eigenvalues satisfy $\rho_{q(k-1)+j} \leq Ck^{-2r}$ for some constant $C > 0$. Consider sub-class of covariance operators, denoted as $\mathscr{P}(r, C, C')$ for some $0 < C' < C < \infty$, which include all $\mathscr{T}^{(\mathscr{S},\mathscr{S})}$ with $C'k^{-2r} \leq \rho_{q(k-1)+j} \leq Ck^{-2r}$. It is straightforward to show that for $k > q$, $c_1(k/q)^{-2r} \leq \rho_k \leq c_2(k/q)^{-2r}$ for some $0 < c_1 < c_2 < \infty$.

As noted in Cai and Yuan (2012) in the proof of their Theorem 1, any lower bound derived under a specific case yields a lower bound for the general case. For the rest of the proof, we will consider a special case where $\mathscr{T}^{(\mathscr{S},\mathscr{S})} \in \mathscr{P}(r, C, C')$ and the functional coefficient in the oracle model has the form

$$\boldsymbol{\beta}_\theta = \mathscr{L}_{\boldsymbol{K}_{\mathscr{S}}^{1/2}} \boldsymbol{f}_\theta, \quad \boldsymbol{f}_\theta = M^{-1/2} \sum_{k=M+1}^{2M} \theta_k \boldsymbol{\phi}_k. \tag{S.21}$$

where $\boldsymbol{\theta} = (\theta_{M+1}, \ldots, \theta_{2M}) \in \{0, 1\}^M$ for some large integer M. The Varshamov–Gilbert bound (Lemma 2.9, Tsybakov (2009)) shows that for any $M \geq 8$, there exists a subset $\boldsymbol{\Theta}_0 = \{\boldsymbol{\theta}^{(0)}, \boldsymbol{\theta}^{(1)}, \ldots, \boldsymbol{\theta}^{(N)}\} \in \{0, 1\}^M$ such that

(a) $\boldsymbol{\theta}^{(0)} = (0, \dots, 0)^\top$; (b) $H(\boldsymbol{\theta}^{(j)}, \boldsymbol{\theta}^{(k)}) \geq M/8$ for any $0 \leq j < k \leq N$, $H(\cdot, \cdot)$ is the Hamming distance; and (c) $N \geq 2^{M/8}$. Because $\{\boldsymbol{f}_\theta : \boldsymbol{\theta} \in \boldsymbol{\Theta}_0\} \subset \mathbb{L}_2^q$, it is clear that $\forall B > 0$

$$\begin{aligned} &\sup_{\mathscr{T}^{(\mathscr{S},\mathscr{S})} \in \mathscr{P}(r)} \sup_{\boldsymbol{f}_{0\mathscr{S}} \in \mathbb{L}_2^q} \mathbb{P}\left(\mathscr{D}(\widetilde{\boldsymbol{f}}_{\mathscr{S}}, \boldsymbol{f}_{0\mathscr{S}}) \geq B\right) \\ &\geq \sup_{\mathscr{T}^{(\mathscr{S},\mathscr{S})} \in \mathscr{P}(r,C,C')} \max_{\boldsymbol{\theta} \in \boldsymbol{\Theta}_0} \mathbb{P}_\theta \left(\mathscr{D}(\widetilde{\boldsymbol{f}}_{\mathscr{S}}, \boldsymbol{f}_\theta) \geq B\right). \end{aligned} \tag{S.22}$$

Here, $\mathbb{P}_\theta$ is the probability measure when the function coefficient has the form given in (S.21).

Next, we proceed to establish the lower bound under the special case using results in Theorem 2.5 of Tsybakov (2009). To that end, for any $\boldsymbol{\theta}, \boldsymbol{\theta}' \in \boldsymbol{\Theta}_0$ such that $\boldsymbol{\theta} \neq \boldsymbol{\theta}'$, the Kullback–Leibler distance between $\mathbb{P}_\theta$ and $\mathbb{P}_{\theta'}$ is given by

$$\begin{aligned} \mathscr{K}\left(\mathbb{P}_\theta \| \mathbb{P}_{\theta'}\right) &= \frac{n}{2\sigma^2} \mathscr{D}^2(\boldsymbol{f}_\theta, \boldsymbol{f}_{\theta'}) = \frac{n}{2\sigma^2 M} \sum_{k=M+1}^{2M} (\theta_k - \theta_k')^2 \rho_k \\ &\leq \frac{n\rho_M}{2\sigma^2 M} H(\boldsymbol{\theta}, \boldsymbol{\theta}') \leq \frac{n\rho_M}{2\sigma^2}. \end{aligned}$$

For any $0 < \alpha < 1/8$, let $M = \lceil c_0 n^{1/(2r+1)} q^{2r/(2r+1)} \rceil$ and $c_0 = D\alpha^{-1/(2r+1)}$ for some large enough $D > 0$, then

$$\frac{1}{N}\sum_{k=1}^{N}\mathscr{K}\left(\mathbb{P}_{\theta^{(k)}}\|\mathbb{P}_{\theta^{(0)}}\right) \leq \frac{c_2}{2\sigma^2}n\left(\frac{M}{q}\right)^{-2r} \leq \frac{c_2 c_0^{-(2r+1)}}{2\sigma^2}M \leq \alpha\log N.$$

On the other hand,

$$\mathscr{D}^2(\boldsymbol{f}_\theta, \boldsymbol{f}_{\theta'}) \geq \frac{\rho_{2M}}{M}H(\boldsymbol{\theta}, \boldsymbol{\theta}') \geq \frac{\rho_{2M}}{8} \geq \frac{c_1}{8}\left(\frac{2M}{q}\right)^{-2r} \geq 4d\alpha^{\frac{2r}{2r+1}}\left(n/q\right)^{-\frac{2r}{2r+1}},$$

for some small enough $d > 0$. By Theorem 2.5 in Tsybakov (2009) we have

$$\begin{aligned}&\inf_{\widetilde{\boldsymbol{f}}_{\mathscr{S}}} \sup_{\mathscr{T}^{(\mathscr{S},\mathscr{S})}\in\mathscr{P}(r,C,C')} \max_{\boldsymbol{\theta}\in\boldsymbol{\Theta}_0} \mathbb{P}_\theta\left(\mathscr{D}^2(\widetilde{\boldsymbol{f}}_{\mathscr{S}}, \boldsymbol{f}_\theta) \geq d\alpha^{\frac{2r}{2r+1}}(n/q)^{-\frac{2r}{2r+1}}\right)\\ &\quad\geq \frac{\sqrt{N}}{1+\sqrt{N}}\left(1 - 2\alpha - \sqrt{\frac{2\alpha}{\log N}}\right).\end{aligned}$$

Letting $a = d\alpha^{2r/(2r+1)}$, we have

$$\lim_{a\to 0}\lim_{n\to\infty}\inf_{\widetilde{\boldsymbol{f}}_{\mathscr{S}}} \sup_{\mathscr{T}^{(\mathscr{S},\mathscr{S})}\in\mathscr{P}(r,C,C')} \max_{\boldsymbol{\theta}\in\boldsymbol{\Theta}_0} \mathbb{P}_\theta\left(\mathscr{D}^2(\widetilde{\boldsymbol{f}}_{\mathscr{S}}, \boldsymbol{f}_\theta) \geq a(n/q)^{-\frac{2r}{2r+1}}\right) = 1. \tag{S.23}$$

The minimax lower bound result in the theorem is derived by combining (S.22) and (S.23). $\square$

Proof of Theorem 4

Proof. We first note that

$$\begin{aligned}
&\mathbb{P}\left(\mathscr{R}(\widehat{\boldsymbol{f}}_{\widehat{\mathscr{S}}}) \geq B\right) - \mathbb{P}\left(\mathscr{R}(\widehat{\boldsymbol{f}}_{\mathscr{S}}) \geq B\right) \\
&= \mathbb{P}\left(\mathscr{R}(\widehat{\boldsymbol{f}}_{\widehat{\mathscr{S}}}) \geq B, \widehat{\mathscr{S}} \neq \mathscr{S}\right) - \mathbb{P}\left(\mathscr{R}(\widehat{\boldsymbol{f}}_{\mathscr{S}}) \geq B, \widehat{\mathscr{S}} \neq \mathscr{S}\right).
\end{aligned}$$

Thus, as long as $\mathbb{P}\left(\widehat{\mathscr{S}} \neq \mathscr{S}\right) \to 0$, we have

$$\lim_{n\to\infty} \sup_{\boldsymbol{\mathscr{T}}^{(\mathscr{S},\mathscr{S})}} \sup_{\boldsymbol{f}_{0\mathscr{S}} \in \mathbb{L}_2^q} \mathbb{P}\left(\mathscr{R}(\widehat{\boldsymbol{f}}_{\widehat{\mathscr{S}}}) \geq B\right) = \lim_{n\to\infty} \sup_{\boldsymbol{\mathscr{T}}^{(\mathscr{S},\mathscr{S})}} \sup_{\boldsymbol{f}_{0\mathscr{S}} \in \mathbb{L}_2^q} \mathbb{P}\left(\mathscr{R}(\widehat{\boldsymbol{f}}_{\mathscr{S}}) \geq B\right).$$

From (3.9), we can easily derive that

$$\widehat{\boldsymbol{f}}_{\mathscr{S}} = \left(\boldsymbol{\mathscr{T}}_{n,\lambda_3}^{(\mathscr{S},\mathscr{S})}\right)^{-1} \left\{\boldsymbol{\mathscr{T}}_n^{(\mathscr{S},\mathscr{S})} \boldsymbol{f}_{0\mathscr{S}} + \boldsymbol{g}_{\mathscr{S}}\right\},$$

where $\boldsymbol{f}_{0\mathscr{S}}$ and $\boldsymbol{g}_{\mathscr{S}}$ are defined in (S.7). Define $\widetilde{\boldsymbol{f}}_{\mathscr{S}} = \left(\boldsymbol{\mathscr{T}}_{\lambda_3}^{(\mathscr{S},\mathscr{S})}\right)^{-1} \boldsymbol{\mathscr{T}}^{(\mathscr{S},\mathscr{S})} \boldsymbol{f}_{0\mathscr{S}}$, then

$$\begin{aligned}
\mathscr{R}^{1/2}(\widehat{\boldsymbol{f}}_{\mathscr{S}}) &= \left\|\left(\boldsymbol{\mathscr{T}}^{(\mathscr{S},\mathscr{S})}\right)^{1/2} (\boldsymbol{f}_{0\mathscr{S}} - \widehat{\boldsymbol{f}}_{\mathscr{S}})\right\|_2 \\
&\leq \left\|\left(\boldsymbol{\mathscr{T}}^{(\mathscr{S},\mathscr{S})}\right)^{1/2} (\boldsymbol{f}_{0\mathscr{S}} - \widetilde{\boldsymbol{f}}_{\mathscr{S}})\right\|_2 + \left\|\left(\boldsymbol{\mathscr{T}}^{(\mathscr{S},\mathscr{S})}\right)^{1/2} (\widetilde{\boldsymbol{f}}_{\mathscr{S}} - \widehat{\boldsymbol{f}}_{\mathscr{S}})\right\|_2.
\end{aligned} \tag{S.24}$$

By Lemma S.7, the first term in (S.24) can be bounded by $\frac{1}{2}\lambda_3^{1/2}\left\|\boldsymbol{f}_{0\mathscr{S}}\right\|_2$. In order to bound the second term, note that

$$
\begin{aligned}
\widetilde{\boldsymbol{f}}_{\mathscr{S}} - \widehat{\boldsymbol{f}}_{\mathscr{S}} &= \left(\mathscr{T}_{\lambda_3}^{(\mathscr{S},\mathscr{S})}\right)^{-1} \mathscr{T}_{n,\lambda_3}^{(\mathscr{S},\mathscr{S})} \left(\widetilde{\boldsymbol{f}}_{\mathscr{S}} - \widehat{\boldsymbol{f}}_{\mathscr{S}}\right) \\
&\quad + \left(\mathscr{T}_{\lambda_3}^{(\mathscr{S},\mathscr{S})}\right)^{-1} \left(\mathscr{T}^{(\mathscr{S},\mathscr{S})} - \mathscr{T}_{n}^{(\mathscr{S},\mathscr{S})}\right) \left(\widetilde{\boldsymbol{f}}_{\mathscr{S}} - \widehat{\boldsymbol{f}}_{\mathscr{S}}\right) \\
&= \left(\mathscr{T}_{\lambda_3}^{(\mathscr{S},\mathscr{S})}\right)^{-1} \mathscr{T}_{n}^{(\mathscr{S},\mathscr{S})} \left(\widetilde{\boldsymbol{f}}_{\mathscr{S}} - \boldsymbol{f}_{0\mathscr{S}}\right) + \lambda_3 \left(\mathscr{T}_{\lambda_3}^{(\mathscr{S},\mathscr{S})}\right)^{-1} \widetilde{\boldsymbol{f}}_{\mathscr{S}} \\
&\quad - \left(\mathscr{T}_{\lambda_3}^{(\mathscr{S},\mathscr{S})}\right)^{-1} \boldsymbol{g}_{\mathscr{S}} + \left(\mathscr{T}_{\lambda_3}^{(\mathscr{S},\mathscr{S})}\right)^{-1} \left(\mathscr{T}^{(\mathscr{S},\mathscr{S})} - \mathscr{T}_{n}^{(\mathscr{S},\mathscr{S})}\right) \left(\widetilde{\boldsymbol{f}}_{\mathscr{S}} - \widehat{\boldsymbol{f}}_{\mathscr{S}}\right) \\
&= \left(\mathscr{T}_{\lambda_3}^{(\mathscr{S},\mathscr{S})}\right)^{-1} \mathscr{T}^{(\mathscr{S},\mathscr{S})} \left(\widetilde{\boldsymbol{f}}_{\mathscr{S}} - \boldsymbol{f}_{0\mathscr{S}}\right) \\
&\quad + \left(\mathscr{T}_{\lambda_3}^{(\mathscr{S},\mathscr{S})}\right)^{-1} \left(\mathscr{T}_{n}^{(\mathscr{S},\mathscr{S})} - \mathscr{T}^{(\mathscr{S},\mathscr{S})}\right) \left(\widetilde{\boldsymbol{f}}_{\mathscr{S}} - \boldsymbol{f}_{0\mathscr{S}}\right) \\
&\quad + \lambda_3 \left(\mathscr{T}_{\lambda_3}^{(\mathscr{S},\mathscr{S})}\right)^{-2} \mathscr{T}^{(\mathscr{S},\mathscr{S})} \boldsymbol{f}_{0\mathscr{S}} - \left(\mathscr{T}_{\lambda_3}^{(\mathscr{S},\mathscr{S})}\right)^{-1} \boldsymbol{g}_{\mathscr{S}} \\
&\quad + \left(\mathscr{T}_{\lambda_3}^{(\mathscr{S},\mathscr{S})}\right)^{-1} \left(\mathscr{T}^{(\mathscr{S},\mathscr{S})} - \mathscr{T}_{n}^{(\mathscr{S},\mathscr{S})}\right) \left(\widetilde{\boldsymbol{f}}_{\mathscr{S}} - \widehat{\boldsymbol{f}}_{\mathscr{S}}\right).
\end{aligned}
$$

Therefore, by the triangular inequality,

$$
\begin{aligned}
&\left\| \left(\mathscr{T}^{(\mathscr{S},\mathscr{S})}\right)^{\nu_1} \left(\widetilde{\boldsymbol{f}}_{\mathscr{S}} - \widehat{\boldsymbol{f}}_{\mathscr{S}}\right) \right\|_2 \\
&\leq \left\| \left(\mathscr{T}^{(\mathscr{S},\mathscr{S})}\right)^{\nu_1} \left(\mathscr{T}_{\lambda_3}^{(\mathscr{S},\mathscr{S})}\right)^{-1} \mathscr{T}^{(\mathscr{S},\mathscr{S})} \left(\widetilde{\boldsymbol{f}}_{\mathscr{S}} - \boldsymbol{f}_{0\mathscr{S}}\right) \right\|_2 \\
&\quad + \left\| \left(\mathscr{T}^{(\mathscr{S},\mathscr{S})}\right)^{\nu_1} \left(\mathscr{T}_{\lambda_3}^{(\mathscr{S},\mathscr{S})}\right)^{-1} \left(\mathscr{T}_{n}^{(\mathscr{S},\mathscr{S})} - \mathscr{T}^{(\mathscr{S},\mathscr{S})}\right) \left(\widetilde{\boldsymbol{f}}_{\mathscr{S}} - \boldsymbol{f}_{0\mathscr{S}}\right) \right\|_2 \\
&\quad + \left\| \lambda_3 \left(\mathscr{T}^{(\mathscr{S},\mathscr{S})}\right)^{\nu_1} \left(\mathscr{T}_{\lambda_3}^{(\mathscr{S},\mathscr{S})}\right)^{-2} \mathscr{T}^{(\mathscr{S},\mathscr{S})} \boldsymbol{f}_{0\mathscr{S}} \right\|_2 \\
&\quad + \left\| \left(\mathscr{T}^{(\mathscr{S},\mathscr{S})}\right)^{\nu_1} \left(\mathscr{T}_{\lambda_3}^{(\mathscr{S},\mathscr{S})}\right)^{-1} \boldsymbol{g}_{\mathscr{S}} \right\|_2 \\
&\quad + \left\| \left(\mathscr{T}^{(\mathscr{S},\mathscr{S})}\right)^{\nu_1} \left(\mathscr{T}_{\lambda_3}^{(\mathscr{S},\mathscr{S})}\right)^{-1} \left(\mathscr{T}^{(\mathscr{S},\mathscr{S})} - \mathscr{T}_{n}^{(\mathscr{S},\mathscr{S})}\right) \left(\widetilde{\boldsymbol{f}}_{\mathscr{S}} - \widehat{\boldsymbol{f}}_{\mathscr{S}}\right) \right\|_2 .
\end{aligned}
\tag{S.25}
$$

For convenience, define

$$
\begin{aligned}
A(\nu) &= \left\| \left(\mathscr{T}^{(\mathscr{S},\mathscr{S})}\right)^{\nu} (\widetilde{\boldsymbol{f}}_{\mathscr{S}} - \widehat{\boldsymbol{f}}_{\mathscr{S}}) \right\|_2, \\
B_1(\nu) &= \left\| \left(\mathscr{T}^{(\mathscr{S},\mathscr{S})}\right)^{\nu} (\widetilde{\boldsymbol{f}}_{\mathscr{S}} - \boldsymbol{f}_{0\mathscr{S}}) \right\|_2, \\
B_2(\nu_1,\nu_2) &= \left\|\left\| \left(\mathscr{T}^{(\mathscr{S},\mathscr{S})}\right)^{\nu_1} \left(\mathscr{T}_{\lambda_3}^{(\mathscr{S},\mathscr{S})}\right)^{-1} \left(\mathscr{T}_{n}^{(\mathscr{S},\mathscr{S})} - \mathscr{T}^{(\mathscr{S},\mathscr{S})}\right) \left(\mathscr{T}^{(\mathscr{S},\mathscr{S})}\right)^{-\nu_2} \right\|\right\|_{2,2}, \\
B_3(\nu) &= \left\|\left\| \lambda_3 \left(\mathscr{T}^{(\mathscr{S},\mathscr{S})}\right)^{\nu} \left(\mathscr{T}_{\lambda_3}^{(\mathscr{S},\mathscr{S})}\right)^{-1} \right\|\right\|_{2,2} \|\boldsymbol{f}_{0\mathscr{S}}\|_2, \\
B_4(\nu) &= \left\| \left(\mathscr{T}^{(\mathscr{S},\mathscr{S})}\right)^{\nu} \left(\mathscr{T}_{\lambda_3}^{(\mathscr{S},\mathscr{S})}\right)^{-1} \boldsymbol{g}_{\mathscr{S}} \right\|_2.
\end{aligned}
$$

Then, (S.25) may be further developed as

$$
A(\nu_1) \leq B_1(\nu_1) + B_2(\nu_1,\nu_2)B_1(\nu_2) + B_3(\nu_1) + B_4(\nu_1) + B_2(\nu_1,\nu_2)A(\nu_2). \tag{S.26}
$$

According to Lemma S.7-S.9, for $0 < \nu \leq 1/2$,

$$
B_1(\nu),\ B_3(\nu),\ B_4(\nu) = O_p(\lambda_3^{\nu}),\ B_4(\nu) = O_p\left(\left(\frac{n}{q}\lambda_3^{1-2\nu+\frac{1}{2r}}\right)^{-\frac{1}{2}}\right).
$$

First, let $\nu_1 = \nu_2 = \nu$ in (S.26), where $0 < \nu < 1/2 - 1/(4r)$. According to

Lemma S.10,

$$B_2(\nu,\nu) = O_p\left(q^{\frac{1}{2}}\left(\frac{n}{q}\lambda_3^{1-2\nu+\frac{1}{2r}}\right)^{-\frac{1}{2}}\right) = O_p\left(q^{\frac{1}{2}}\lambda_3^{\nu}\right) = o_p(1),$$
$$B_4(\nu) = O_p(\lambda_3^{\nu}), \tag{S.27}$$

provided that $\lambda_3 \asymp (n/q)^{-2r/(2r+1)}$ and $q = o\left(n^{\frac{4r\nu}{2r+1+4r\nu}}\right)$. In this case, combining the last term on the rhs of (S.26) with the lhs, we obtain

$$\left\|\left(\mathscr{T}^{(\mathscr{S},\mathscr{S})}\right)^{\nu}\left(\widetilde{\boldsymbol{f}}_{\mathscr{S}} - \widehat{\boldsymbol{f}}_{\mathscr{S}}\right)\right\|_2 = O_p(\lambda_3^{\nu}) \tag{S.28}$$

provided that $\lambda_3 \asymp (n/q)^{-2r/(2r+1)}$.

Next, we let $\nu_1 = 1/2$, $\nu_2 = \nu \in (0, 1/2 - 1/(4r))$ in (S.26). According to Lemma S.10,

$$B_2(1/2,\nu) = O_p\left(q^{\frac{1}{2}}\left(\frac{n}{q}\lambda_3^{\frac{1}{2r}}\right)^{-\frac{1}{2}}\right) = O_p\left(q^{\frac{1}{2}}\lambda_3^{\frac{1}{2}}\right) = o_p\left(q^{\frac{1}{2}}\lambda_3^{\nu}\right)$$

provided that $\lambda_3 \asymp (n/q)^{-2r/(2r+1)}$.

$$\begin{aligned}
&\left|\left|\left|\left(\mathscr{T}^{(\mathscr{S},\mathscr{S})}\right)^{1/2}\left(\mathscr{T}_{\lambda_3}^{(\mathscr{S},\mathscr{S})}\right)^{-1}\left(\mathscr{T}_n^{(\mathscr{S},\mathscr{S})} - \mathscr{T}^{(\mathscr{S},\mathscr{S})}\right)\left(\mathscr{T}^{(\mathscr{S},\mathscr{S})}\right)^{-\nu}\right|\right|\right|_{2,2} \\
&= O_p\left(q^{\frac{1}{2}}\left(\frac{n}{q}\lambda_3^{\frac{1}{2r}}\right)^{-\frac{1}{2}}\right) = O_p\left(q^{\frac{1}{2}}\lambda_3^{\frac{1}{2}}\right) = o_p\left(q^{\frac{1}{2}}\lambda_3^{\nu}\right)
\end{aligned}$$

provided that $\lambda_3 \asymp (n/q)^{-2r/(2r+1)}$. When $q = o\left(n^{\frac{4r\nu}{2r+1+4r\nu}}\right)$, the above expression has an order of $o_p(1)$. In this case, again

$$\left\| \left(\mathscr{T}^{(\mathscr{S},\mathscr{S})}\right)^{1/2} (\widetilde{\boldsymbol{f}}_{\mathscr{S}} - \widehat{\boldsymbol{f}}_{\mathscr{S}}) \right\|_2 = O_p\left(\lambda_3^{\frac{1}{2}} + \left(\frac{n}{q}\lambda_3^{\frac{1}{2r}}\right)^{-\frac{1}{2}} + q^{\frac{1}{2}}\lambda_3^{\frac{1}{2}+\nu} \right).$$

Thus, $\left\| \left(\mathscr{T}^{(\mathscr{S},\mathscr{S})}\right)^{1/2} (\widetilde{\boldsymbol{f}}_{\mathscr{S}} - \widehat{\boldsymbol{f}}_{\mathscr{S}}) \right\|_2 = O_p\left(\lambda_3^{\frac{1}{2}}\right)$. As a result, $\mathscr{R}(\widehat{\boldsymbol{f}}_{\mathscr{S}}) = O_p(\lambda_3)$ provided that $\lambda_3 \asymp (n/q)^{-2r/(2r+1)}$. Finally, let $\nu \to 1/2 - 1/(4r)$, we have $q = o\left(n^{\frac{2r-1}{4r}}\right)$. □

S.2.3 Proofs of Lemmas

Proof of Lemma S.1

Proof. Note that

$$\sum_{k=1}^{K} a_k \exp(-b_k x) \;\leq\; Ka\exp(-bx) \;=\; \exp\left[-\{b - x^{-1}\log(Ka)\}x\right].$$

We established the Lemma by noting that $b - x^{-1}\log(Ka) > (1-\epsilon)b$. □

Proof of Lemma S.2

Proof. First of all, we claim that, for any $\xi \in (0,1)$,

$$
\begin{aligned}
&\mathbb{P}\left(\max_{j\in\mathscr{S}^c}\left\|\frac{\sigma}{\lambda_1 n}\widetilde{\boldsymbol{X}}_{\bullet j}^{\top}\left(\mathrm{I}-\boldsymbol{\Delta}_n\right)\boldsymbol{z}_n\right\|_2 \geq \frac{\xi C_{\min}}{3}\right)\\
&\leq 2(p-q)\exp\left\{-\frac{\lambda_1^2 C_{\min}^2 \xi^2 n}{48\sigma^2\tau}\right\} + (p-q)\exp\left(-\frac{n}{32}\right).
\end{aligned} \tag{S.29}
$$

To show (S.29), first apply the union bound to get

$$
\begin{aligned}
&\mathbb{P}\left(\max_{j\in\mathscr{S}^c}\left\|\frac{\sigma}{\lambda_1 n}\widetilde{\boldsymbol{X}}_{\bullet j}^{\top}\left(\mathrm{I}-\boldsymbol{\Delta}_n\right)\boldsymbol{z}_n\right\|_2 \geq \frac{\xi C_{\min}}{3}\right)\\
&= \mathbb{P}\left(\bigcup_{j\in\mathscr{S}^c}\left\{\left\|\frac{\sigma}{\lambda_1 n}\widetilde{\boldsymbol{X}}_{\bullet j}^{\top}\left(\mathrm{I}-\boldsymbol{\Delta}_n\right)\boldsymbol{z}_n\right\|_2 \geq \frac{\xi C_{\min}}{3}\right\}\right)\\
&\leq \sum_{j\in\mathscr{S}^c}\mathbb{P}\left(\frac{1}{\sqrt{n}}\left\|\widetilde{\boldsymbol{X}}_{\bullet j}^{\top}\left(\mathrm{I}-\boldsymbol{\Delta}_n\right)\boldsymbol{z}_n\right\|_2 \geq \frac{\lambda_1 \xi C_{\min}\sqrt{n}}{3\sigma}\right).
\end{aligned}
$$

Write

$$
\mathscr{T}_n^{(\mathscr{S},\mathscr{S})} = \frac{1}{n}\widetilde{\boldsymbol{X}}_{\mathscr{S}}^{\top}\otimes\widetilde{\boldsymbol{X}}_{\mathscr{S}} = \sum_{k=1}^{\infty}\widehat{\rho}_k\widehat{\boldsymbol{\phi}}_k\otimes\widehat{\boldsymbol{\phi}}_k^{\top}, \tag{S.30}
$$

where the $\widehat{\rho}_k$ are the (nonnegative) eigenvalues of $\mathscr{T}_n^{(\mathscr{S},\mathscr{S})}$ arranged in descending order and $\widehat{\boldsymbol{\phi}}_k$ are the corresponding eigenfunctions. Assume without loss of generality that $\{\widehat{\boldsymbol{\phi}}_k, k\geq 1\}$ is a CONS of $\mathbb{L}_2^q[0,1]$. It follows that

$$
\widetilde{\boldsymbol{X}}_{\mathscr{S}}(u) = \sum_{k=1}^{\infty}\widehat{\boldsymbol{\zeta}}_k\widehat{\boldsymbol{\phi}}_k^{\top}(u), \tag{S.31}
$$

where $\widehat{\boldsymbol{\zeta}}_k := \int \widetilde{\boldsymbol{X}}_{\mathscr{S}}(t)\widehat{\boldsymbol{\phi}}_k(t)dt$ satisfies

$$\frac{1}{n}\widehat{\boldsymbol{\zeta}}_j^\top \widehat{\boldsymbol{\zeta}}_k = \langle \widehat{\boldsymbol{\phi}}_j, \mathscr{T}_n^{(\mathscr{S},\mathscr{S})}\widehat{\boldsymbol{\phi}}_k\rangle_2 = \widehat{\rho}_k \delta_{j,k}. \tag{S.32}$$

Since there are at most n linearly independent $\widehat{\boldsymbol{\zeta}}_k$, $\widehat{\rho}_k = 0, k > n$ and higher order eigenfunctions $\widehat{\boldsymbol{\phi}}_k$, for $k > n$ can be obtained by the Gram-Schmidt orthogonalization. Thus, we can re-express $\boldsymbol{\Delta}_n$ as

$$\begin{aligned}
\boldsymbol{\Delta}_n &= \frac{1}{n}\int \sum_{i=1}^n \widehat{\boldsymbol{\zeta}}_i \widehat{\boldsymbol{\phi}}_i^\top(u)\left\{\sum_{k=1}^\infty (\widehat{\rho}_k + \lambda_2)^{-1}(\widehat{\boldsymbol{\phi}}_k \otimes \widehat{\boldsymbol{\phi}}_k^\top)\widetilde{\boldsymbol{X}}_{\mathscr{S}}\right\}(u)du \\
&= \frac{1}{n}\iint \sum_{i=1}^n \widehat{\boldsymbol{\zeta}}_i \widehat{\boldsymbol{\phi}}_i^\top(u)\left\{\sum_{k=1}^\infty \widehat{\boldsymbol{\phi}}_k(u)(\widehat{\rho}_k + \lambda_2)^{-1}\widehat{\boldsymbol{\phi}}_k^\top(v)\right\}\sum_{j=1}^n \widehat{\boldsymbol{\phi}}_j(v)\widehat{\boldsymbol{\zeta}}_j^\top dudv \\
&= \frac{1}{n}\sum_{k=1}^n \frac{1}{\widehat{\rho}_k + \lambda_2}\widehat{\boldsymbol{\zeta}}_k \widehat{\boldsymbol{\zeta}}_k^\top = \sum_{k=1}^n \frac{\widehat{\rho}_k}{\widehat{\rho}_k + \lambda_2}\widehat{\boldsymbol{\zeta}}_k^* \widehat{\boldsymbol{\zeta}}_k^{*\top},
\end{aligned}$$

where $\widehat{\boldsymbol{\zeta}}_k^* = (n\widehat{\rho}_k)^{-1/2}\widehat{\boldsymbol{\zeta}}_k$ are n-dim orthonormal vectors. Clearly, $\mathrm{I} - \boldsymbol{\Delta}_n$ is a positive-definite matrix with all eigenvalues less or equal to 1.

Conditional on $\boldsymbol{X}_n$, $Q_j(t) := n^{-1/2}\widetilde{\boldsymbol{X}}_{\bullet j}^\top(t)(\mathrm{I} - \boldsymbol{\Delta}_n)\boldsymbol{z}_n$ is a rank n Gaussian process with

$$\mathbb{E}[Q_j(t)|\boldsymbol{X}_n] = 0 \quad \text{and} \quad \mathbb{C}\mathrm{ov}[Q_j(s), Q_j(t)|\boldsymbol{X}_n] = n^{-1}\widetilde{\boldsymbol{X}}_{\bullet j}^\top(s)(\mathrm{I} - \boldsymbol{\Delta}_n)^2\widetilde{\boldsymbol{X}}_{\bullet j}(t).$$

Also, note that

$$n^{-1}\int \widetilde{\boldsymbol{X}}_{\bullet j}^{\top}(s)(\mathrm{I}-\boldsymbol{\Delta}_n)^2\widetilde{\boldsymbol{X}}_{\bullet j}(s)ds \le n^{-1}\int \widetilde{\boldsymbol{X}}_{\bullet j}^{\top}(s)\widetilde{\boldsymbol{X}}_{\bullet j}(s)ds = \operatorname{tr}\left(\mathscr{T}_n^{(j,j)}\right). \tag{S.33}$$

Define the event $\mathscr{D}_j(c_0) = \left\{\operatorname{tr}\left(\mathscr{T}_n^{(j,j)}\right) < c_0\right\}$. It follows that

$$\begin{aligned}
&\mathbb{P}\left(\|Q_j\|_2 \ge \frac{\lambda_1\xi C_{\min}\sqrt{n}}{3\sigma}\right)\\
&= \mathbb{E}\left[\mathbb{P}\left(\|Q_j\|_2^2 \ge \frac{\lambda_1^2\xi^2 C_{\min}^2 n}{9\sigma^2}\middle|\boldsymbol{X}_n\right)\right]\\
&\le \mathbb{E}\left[\mathbb{P}\left(\|Q_j\|_2^2 \ge \frac{\lambda_1^2\xi^2 C_{\min}^2 n}{9\sigma^2}\middle|\boldsymbol{X}_n, \mathscr{D}_j(c_0)\right)\mathrm{I}\left(\mathscr{D}_j(c_0)\right)\right] + \mathbb{P}\left(\mathscr{D}_j^c(c_0)\right)\\
&\le \mathbb{E}\left[\mathbb{P}\left(\|Q_j\|_2^2 \ge \frac{\lambda_1^2\xi^2 C_{\min}^2 n}{9\sigma^2}\middle|\boldsymbol{X}_n, \mathscr{D}_j(c_0)\right)\right] + \mathbb{P}\left(\mathscr{D}_j^c(c_0)\right).
\end{aligned} \tag{S.34}$$

By Lemma S.13 (i) with $L = 1, K = n, s = 4/3$,

$$\mathbb{P}\left(\|Q_j\|_2^2 \ge \frac{\lambda_1^2\xi^2 C_{\min}^2 n}{9\sigma^2}\middle|\boldsymbol{X}_n, \mathscr{D}_j(c_0)\right) \le 2\exp\left\{-\frac{\lambda_1^2\xi^2 C_{\min}^2 n}{24\sigma^2 c_0}\right\}. \tag{S.35}$$

Recall that $\mathscr{T}_n^{(j,j)} = \frac{1}{n}\sum_{i=1}^n \widetilde{X}_{ij}\otimes\widetilde{X}_{ij}$, and $\widetilde{X}_{ij} \overset{indep}{\sim} \mathscr{GP}\left(0, \mathscr{T}^{(j,j)}\right)$. Thus, $\operatorname{tr}\left(\mathscr{T}_n^{(j,j)}\right) = \frac{1}{n}\sum_{i=1}^n \|\widetilde{X}_{ij}\|_2^2$ and

$$\mathbb{P}\left(\mathscr{D}_j^c(c_0)\right) = \mathbb{P}\left(\operatorname{tr}\left(\mathscr{T}_n^{(j,j)}\right) > c_0\right) = \mathbb{P}\left(\sum_{i=1}^n \|\widetilde{X}_{ij}\|_2^2 > nc_0\right).$$

Thus, by Lemma S.13 (ii) with $L = n$, $s = 16/9$ and the assumption (C.2) we obtain

$$\mathbb{P}\left(\mathscr{D}_j^c(c_0)\right) \leq \exp\left\{-\frac{c_0 - \mathrm{tr}(\mathscr{T}^{(j,j)})}{32} n\right\} \leq \exp\left\{-\frac{c_0 - \tau}{32} n\right\}, \tag{S.36}$$

for any $c_0 > (1 + s/2)\tau$. It follows from (S.34)-(S.36), with $c_0 = 2\tau$, $\tau > 1$ and $\lambda_1 < D^*_{1,1}$, we obtain

$$\begin{aligned}
&\mathbb{P}\left(\max_{j \in \mathscr{S}^c} \left\| \frac{\sigma}{\lambda_1 \sqrt{n}} Q_j \right\|_2 \geq \frac{\xi C_{\min}}{3}\right) \\
&\leq 2(p-q)\exp\left\{-\frac{\lambda_1^2 C_{\min}^2 \xi^2 n}{48\sigma^2 \tau}\right\} + (p-q)\exp\left(-\frac{\lambda_1^2 n}{32(D^*_{1,1})^2}\right).
\end{aligned}$$

This proves (S.29). Suppose for $d \in (0,1)$, we have

$$\lambda_1 > \max\left(\sqrt{\frac{48}{d}} \cdot \frac{\sigma \tau^{1/2}}{C_{\min}\xi}, \quad \sqrt{\frac{32}{d}} \cdot D^*_{1,1}\right) \cdot \sqrt{\frac{\log(p-q)}{n}},$$

which is equivalent to

$$\frac{C_{\min}^2 \xi^2}{48\sigma^2\tau} \cdot \lambda_1^2 n - \log(p-q) > (1-d)\frac{C_{\min}^2 \xi^2}{48\sigma^2\tau} \cdot \lambda_1^2 n,$$

and

$$\frac{\lambda_1^2 n}{32(D_{1,1}^*)^2} - \log(p-q) > (1-d)\frac{\lambda_1^2 n}{32(D_{1,1}^*)^2}.$$

By Lemma S.1 with $\xi = \gamma/3$ and $d = 1/2$, we have

$$\mathbb{P}\left(\max_{j\in\mathscr{S}^c}\left\|\frac{\sigma}{\lambda_1 n}\widetilde{\boldsymbol{X}}_{\bullet j}^{\top}\left(\mathrm{I}-\boldsymbol{\Delta}_n\right)\boldsymbol{z}_n\right\|_2 \geq \frac{\gamma C_{\min}}{9}\right) \leq \exp\left(-D\lambda_1^2 n\right)$$

holds for any D and λ_1 such that

$$\lambda_1 > D_1^* \cdot \frac{(\sigma+1)\tau^{1/2}}{C_{\min}\gamma} \cdot \sqrt{\frac{\log(p-q)}{n}},$$

and

$$D < D_2^* \frac{C_{\min}^2\gamma^2}{(\sigma+1)^2\tau} < \min\left\{\frac{C_{\min}^2\gamma^2}{864\sigma^2\tau}, \frac{1}{32(D_{1,1}^*)^2}\right\},$$

where D_1^* and D_2^* are universal constants. □

Proof of Lemma S.3

Proof. Let constants ξ, δ, and $\mu = \lambda_1/\lambda_2$ satisfy

$$\xi \in (0, \gamma/2) \quad \text{and} \quad \delta \in (0, (\gamma - 2\xi)/(1-\gamma)) \quad \text{and} \quad \mu C_{\max} > (1-2\xi)/\xi. \tag{S.37}$$

We claim that

$$\begin{aligned} &\mathbb{P}\left(\max_{j \in \mathscr{S}^c} \left\| \boldsymbol{\mathscr{T}}_n^{(j,\mathscr{S})} \left(\boldsymbol{\mathscr{T}}_{n,\lambda_2}^{(\mathscr{S},\mathscr{S})}\right)^{-1} \left(\frac{\lambda_2}{\lambda_1} \boldsymbol{f}_{0\mathscr{S}} + \boldsymbol{\omega}_{\mathscr{S}}\right)\right\|_2 \geq \left(1 - \frac{2\xi}{3}\right) C_{\min}\right) \\ &\leq \exp\left\{-\frac{\lambda_2^2 \varkappa^2 \delta^2 n}{4C^2 \rho_1^2 q}\right\} + 2(p-q)\exp\left\{-\frac{\lambda_2 (C_{\min}/C_{\max})^2 \xi^2 n}{24(1+\mu^{-1}C_{\max}^{-1})^2 \tau q}\right\}, \end{aligned} \tag{S.38}$$

for some constant $C > 0$ and q that satisfy

$$\frac{\lambda_2^2 \varkappa^2 \delta^2}{4C^2 \rho_1^2} < q \leq \sqrt{\frac{\lambda_2^2 \varkappa^2 \delta^2 n}{4C^2 \tau \rho_1}}. \tag{S.39}$$

To show (S.38), by Lemma S.12, for any $j \in \mathscr{S}^c$,

$$\widetilde{\boldsymbol{X}}_{\bullet j}^\top \stackrel{d}{=} \boldsymbol{\mathscr{T}}^{(j,\mathscr{S})} (\boldsymbol{\mathscr{T}}^{(\mathscr{S},\mathscr{S})})^- \widetilde{\boldsymbol{X}}_{\mathscr{S}}^\top + \boldsymbol{E}_j^\top, \tag{S.40}$$

where $\boldsymbol{E}_j = (E_{1j}, \ldots, E_{nj})^\top$ is a vector of iid zero-mean Gaussian processes independent of $\widetilde{\boldsymbol{X}}_{\mathscr{S}}$ with a covariance operator

$$\boldsymbol{\mathscr{T}}^{(j|\mathscr{S})} := \boldsymbol{\mathscr{T}}^{(j,j)} - \boldsymbol{\mathscr{T}}^{(j,\mathscr{S})} (\boldsymbol{\mathscr{T}}^{(\mathscr{S},\mathscr{S})})^- \boldsymbol{\mathscr{T}}^{(\mathscr{S},j)}. \tag{S.41}$$

With (S.40) and Condition 1,

$$
\begin{aligned}
&\left\|\mathscr{T}_n^{(j,\mathscr{S})}\left(\mathscr{T}_{n,\lambda_2}^{(\mathscr{S},\mathscr{S})}\right)^{-1}\left(\frac{\lambda_2}{\lambda_1}\boldsymbol{f}_{0\mathscr{S}}+\boldsymbol{\omega}_{\mathscr{S}}\right)\right\|_2 \qquad \text{(S.42)}\\
&=\left\|\int\frac{1}{n}\widetilde{\boldsymbol{X}}_{\bullet j}^{\top}(\cdot)\widetilde{\boldsymbol{X}}_{\mathscr{S}}(s)\left(\mathscr{T}_{n,\lambda_2}^{(\mathscr{S},\mathscr{S})}\right)^{-1}\left(\frac{\lambda_2}{\lambda_1}\boldsymbol{f}_{0\mathscr{S}}+\boldsymbol{\omega}_{\mathscr{S}}\right)(s)ds\right\|_2\\
&=\left\|\int\frac{1}{n}\{\boldsymbol{\mathscr{T}}^{(j,\mathscr{S})}(\boldsymbol{\mathscr{T}}^{(\mathscr{S},\mathscr{S})})^{-}\widetilde{\boldsymbol{X}}_{\mathscr{S}}^{\top}+\boldsymbol{E}_j^{\top}\}(\cdot)\widetilde{\boldsymbol{X}}_{\mathscr{S}}(s)\left(\mathscr{T}_{n,\lambda_2}^{(\mathscr{S},\mathscr{S})}\right)^{-1}\left(\frac{\lambda_2}{\lambda_1}\boldsymbol{f}_{0\mathscr{S}}+\boldsymbol{\omega}_{\mathscr{S}}\right)(s)ds\right\|_2\\
&\le\left|\!\left|\!\left|\boldsymbol{\mathscr{T}}^{(j,\mathscr{S})}(\boldsymbol{\mathscr{T}}^{(\mathscr{S},\mathscr{S})})^{-}\right|\!\right|\!\right|_{\infty,2}\left|\!\left|\!\left|\mathscr{T}_n^{(\mathscr{S},\mathscr{S})}(\mathscr{T}_{n,\lambda_2}^{(\mathscr{S},\mathscr{S})})^{-1}\right|\!\right|\!\right|_{\infty,\infty}\left(\frac{\lambda_2}{\lambda_1}\|\boldsymbol{f}_{0\mathscr{S}}\|_\infty+C_{\max}\right)+\left\|\boldsymbol{E}_j^{\top}(\cdot)\boldsymbol{Z}\right\|_2,
\end{aligned}
$$

where

$$
\boldsymbol{Z}_{n\times 1}:=\frac{1}{n}\int\widetilde{\boldsymbol{X}}_{\mathscr{S}}(s)\left(\mathscr{T}_{n,\lambda_2}^{(\mathscr{S},\mathscr{S})}\right)^{-1}\left(\frac{\lambda_2}{\lambda_1}\boldsymbol{f}_{0\mathscr{S}}+\boldsymbol{\omega}_{\mathscr{S}}\right)(s)ds. \qquad \text{(S.43)}
$$

Note that if

$$
\left|\!\left|\!\left|\mathscr{T}_n^{(\mathscr{S},\mathscr{S})}(\mathscr{T}_{n,\lambda_2}^{(\mathscr{S},\mathscr{S})})^{-1}\right|\!\right|\!\right|_{\infty,\infty}<\varkappa(1+\delta)\quad\text{and}\quad\max_{j\in\mathscr{S}^c}\left\|\boldsymbol{E}_j^{\top}(\cdot)\boldsymbol{Z}\right\|_2<\frac{\xi C_{\min}}{3},
$$

then (S.37), (S.42) along with Condition C.3 and $\|\boldsymbol{f}_{0\mathscr{S}}\|_\infty=1$ give

$$
\begin{aligned}
&\max_{j\in\mathscr{S}^c}\left\|\mathscr{T}_n^{(j,\mathscr{S})}\left(\mathscr{T}_{n,\lambda_2}^{(\mathscr{S},\mathscr{S})}\right)^{-1}\left(\frac{\lambda_2}{\lambda_1}\boldsymbol{f}_{0\mathscr{S}}+\boldsymbol{\omega}_{\mathscr{S}}\right)\right\|_2\\
&<(1-\gamma)(1+\delta)\left(1+C_{\max}^{-1}\frac{\lambda_2}{\lambda_1}\right)C_{\min}+\frac{\xi C_{\min}}{3}<\left(1-\frac{2\xi}{3}\right)C_{\min},
\end{aligned}
$$

where the last inequality follows from the fact

$$(1-\gamma)(1+\delta)\left(1+C_{\max}^{-1}\frac{\lambda_2}{\lambda_1}\right)<1-\xi$$

by (S.37). In the following, we establish

$$\mathbb{P}\left(\left\vert\kern-0.25ex\left\vert\kern-0.25ex\left\vert \boldsymbol{\mathscr{T}}_n^{(\mathscr{S},\mathscr{S})}(\boldsymbol{\mathscr{T}}_{n,\lambda_2}^{(\mathscr{S},\mathscr{S})})^{-1}\right\vert\kern-0.25ex\right\vert\kern-0.25ex\right\vert_{\infty,\infty}\geq \varkappa(1+\delta)\right)\leq \exp\left\{-\frac{\lambda_2^2\varkappa^2\delta^2 n}{4C^2\rho_1^2 q}\right\} \tag{S.44}$$

and

$$\mathbb{P}\left(\max_{j\in\mathscr{S}^c}\left\|\boldsymbol{E}_j^\top(\cdot)\boldsymbol{Z}\right\|_2\geq\frac{\xi C_{\min}}{3}\right)\leq 2(p-q)\exp\left\{-\frac{\lambda_2(C_{\min}/C_{\max})^2\xi^2 n}{24(1+\mu^{-1}C_{\max}^{-1})^2\tau q}\right\}. \tag{S.45}$$

To show (S.44), first apply the triangle inequality to obtain

$$\begin{aligned}
&\left\vert\kern-0.25ex\left\vert\kern-0.25ex\left\vert \boldsymbol{\mathscr{T}}_n^{(\mathscr{S},\mathscr{S})}(\boldsymbol{\mathscr{T}}_{n,\lambda_2}^{(\mathscr{S},\mathscr{S})})^{-1}\right\vert\kern-0.25ex\right\vert\kern-0.25ex\right\vert_{\infty,\infty}\\
&\leq \left\vert\kern-0.25ex\left\vert\kern-0.25ex\left\vert \boldsymbol{\mathscr{T}}^{(\mathscr{S},\mathscr{S})}(\boldsymbol{\mathscr{T}}_{\lambda_2}^{(\mathscr{S},\mathscr{S})})^{-1}\right\vert\kern-0.25ex\right\vert\kern-0.25ex\right\vert_{\infty,\infty}+\left\vert\kern-0.25ex\left\vert\kern-0.25ex\left\vert \boldsymbol{\mathscr{T}}_n^{(\mathscr{S},\mathscr{S})}\left\{(\boldsymbol{\mathscr{T}}_{n,\lambda_2}^{(\mathscr{S},\mathscr{S})})^{-1}-(\boldsymbol{\mathscr{T}}_{\lambda_2}^{(\mathscr{S},\mathscr{S})})^{-1}\right\}\right\vert\kern-0.25ex\right\vert\kern-0.25ex\right\vert_{\infty,\infty}\\
&\quad+\left\vert\kern-0.25ex\left\vert\kern-0.25ex\left\vert (\boldsymbol{\mathscr{T}}_n^{(\mathscr{S},\mathscr{S})}-\boldsymbol{\mathscr{T}}^{(\mathscr{S},\mathscr{S})})(\boldsymbol{\mathscr{T}}_{\lambda_2}^{(\mathscr{S},\mathscr{S})})^{-1}\right\vert\kern-0.25ex\right\vert\kern-0.25ex\right\vert_{\infty,\infty}.
\end{aligned} \tag{S.46}$$

Then, by Lemma S.15,

$$
\begin{aligned}
&\left|\!\left|\!\left|\mathscr{T}_n^{(\mathscr{S},\mathscr{S})}\left\{(\mathscr{T}_{n,\lambda_2}^{(\mathscr{S},\mathscr{S})})^{-1}-(\mathscr{T}_{\lambda_2}^{(\mathscr{S},\mathscr{S})})^{-1}\right\}\right|\!\right|\!\right|_{\infty,\infty}\\
&\le\sqrt{q}\left|\!\left|\!\left|\mathscr{T}_n^{(\mathscr{S},\mathscr{S})}(\mathscr{T}_{n,\lambda_2}^{(\mathscr{S},\mathscr{S})})^{-1}(\mathscr{T}^{(\mathscr{S},\mathscr{S})}-\mathscr{T}_n^{(\mathscr{S},\mathscr{S})})(\mathscr{T}_{\lambda_2}^{(\mathscr{S},\mathscr{S})})^{-1}\right|\!\right|\!\right|_{2,2}\\
&\le\sqrt{q}\left|\!\left|\!\left|\mathscr{T}_n^{(\mathscr{S},\mathscr{S})}(\mathscr{T}_{n,\lambda_2}^{(\mathscr{S},\mathscr{S})})^{-1}\right|\!\right|\!\right|_{2,2}|\!|\!|\mathscr{T}^{(\mathscr{S},\mathscr{S})}-\mathscr{T}_n^{(\mathscr{S},\mathscr{S})}|\!|\!|_{2,2}\left|\!\left|\!\left|(\mathscr{T}_{\lambda_2}^{(\mathscr{S},\mathscr{S})})^{-1}\right|\!\right|\!\right|_{2,2}
\end{aligned}
\tag{S.47}
$$

and

$$
\begin{aligned}
&\left|\!\left|\!\left|(\mathscr{T}_n^{(\mathscr{S},\mathscr{S})}-\mathscr{T}^{(\mathscr{S},\mathscr{S})})(\mathscr{T}_{\lambda_2}^{(\mathscr{S},\mathscr{S})})^{-1}\right|\!\right|\!\right|_{\infty,\infty}\\
&\le\sqrt{q}\left|\!\left|\!\left|(\mathscr{T}_n^{(\mathscr{S},\mathscr{S})}-\mathscr{T}^{(\mathscr{S},\mathscr{S})})(\mathscr{T}_{\lambda_2}^{(\mathscr{S},\mathscr{S})})^{-1}\right|\!\right|\!\right|_{2,2}\\
&\le\sqrt{q}|\!|\!|\mathscr{T}_n^{(\mathscr{S},\mathscr{S})}-\mathscr{T}^{(\mathscr{S},\mathscr{S})}|\!|\!|_{2,2}\left|\!\left|\!\left|(\mathscr{T}_{\lambda_2}^{(\mathscr{S},\mathscr{S})})^{-1}\right|\!\right|\!\right|_{2,2}.
\end{aligned}
\tag{S.48}
$$

Since

$$
\left|\!\left|\!\left|(\mathscr{T}_{\lambda_2}^{(\mathscr{S},\mathscr{S})})^{-1}\right|\!\right|\!\right|_{2,2}\le\frac{1}{\lambda_2}\quad\text{and}\quad\left|\!\left|\!\left|\mathscr{T}_n^{(\mathscr{S},\mathscr{S})}(\mathscr{T}_{n,\lambda_2}^{(\mathscr{S},\mathscr{S})})^{-1}\right|\!\right|\!\right|_{2,2}\le 1,
\tag{S.49}
$$

(S.46)-(S.49) together with the Condition C.3 give

$$
\left|\!\left|\!\left|\mathscr{T}_n^{(\mathscr{S},\mathscr{S})}(\mathscr{T}_{n,\lambda_2}^{(\mathscr{S},\mathscr{S})})^{-1}\right|\!\right|\!\right|_{\infty,\infty}\le\varkappa+\frac{2\sqrt{q}}{\lambda_2}|\!|\!|\mathscr{T}^{(\mathscr{S},\mathscr{S})}-\mathscr{T}_n^{(\mathscr{S},\mathscr{S})}|\!|\!|_{2,2}.
$$

Thus, for $\delta > 0$, by Lemma S.6 we have

$$\mathbb{P}\left(\left|\left|\left|\mathscr{T}_n^{(\mathscr{S},\mathscr{S})}(\mathscr{T}_{n,\lambda_2}^{(\mathscr{S},\mathscr{S})})^{-1}\right|\right|\right|_{\infty,\infty} \geq \varkappa(1+\delta)\right) \leq \mathbb{P}\left(|||\mathscr{T}^{(\mathscr{S},\mathscr{S})} - \mathscr{T}_n^{(\mathscr{S},\mathscr{S})}|||_{2,2} \geq \frac{\lambda_2\varkappa\delta}{2\sqrt{q}}\right)$$
$$\leq \exp\left\{-\frac{\lambda_2^2\varkappa^2\delta^2 n}{4C^2\rho_1^2 q}\right\}$$

for some constant $C > 0$ and q satisfies (S.39). This proves (S.44).

To prove (S.45), recall the definitions of $\boldsymbol{Z}$ in (S.43) and let $U_j(\cdot) = \boldsymbol{E}_j^\top(\cdot)\boldsymbol{Z}$. We have

$$\mathbb{P}\left(\max_{j\in\mathscr{S}^c}\|U_j(\cdot)\|_2 \geq \frac{\xi C_{\min}}{3}\right) \leq \sum_{j\in\mathscr{S}^c}\mathbb{P}\left(\|U_j(\cdot)\|_2 \geq \frac{\xi C_{\min}}{3}\right)$$
$$= \sum_{j\in\mathscr{S}^c}\mathbb{E}\left\{\mathbb{P}\left(\|U_j(\cdot)\|_2 \geq \frac{\xi C_{\min}}{3}|\widetilde{\boldsymbol{X}}_{\mathscr{S}}\right)\right\}.$$

Also, conditional on $\widetilde{\boldsymbol{X}}_{\mathscr{S}}$, U_j is a zero-mean Gaussian process with covariance operator $\mathscr{H}_j$ with trace

$$\mathrm{tr}(\mathscr{H}_j) = \|\boldsymbol{Z}\|^2\mathrm{tr}(\mathscr{T}^{(j|\mathscr{S})}), \tag{S.50}$$

where $\mathscr{T}^{(j|\mathscr{S})}$ is defined in (S.41). It remains to bound $\|\boldsymbol{Z}\|^2$ and $\mathrm{tr}(\mathscr{T}^{(j|\mathscr{S})})$. First, by (S.41) and (C.2),

$$\mathrm{tr}(\mathscr{T}^{(j|\mathscr{S})}) \leq \mathrm{tr}(\mathscr{T}^{(j,j)}) \leq \tau. \tag{S.51}$$

By the decompositions in (S.30) and (S.31),

$$\left(\mathscr{T}_{n,\lambda_2}^{(\mathscr{S},\mathscr{S})}\right)^{-1}\left(\frac{\lambda_2}{\lambda_1}\boldsymbol{f}_{0\mathscr{S}}+\boldsymbol{\omega}_{\mathscr{S}}\right)=\sum_{k\geq 1}\frac{1}{\widehat{\rho}_k+\lambda_2}\left\langle\widehat{\boldsymbol{\phi}}_k,\frac{\lambda_2}{\lambda_1}\boldsymbol{f}_{0\mathscr{S}}+\boldsymbol{\omega}_{\mathscr{S}}\right\rangle_2\widehat{\boldsymbol{\phi}}_k,$$

and therefore

$$\begin{aligned}
\|\boldsymbol{Z}\|^2&=\frac{1}{n^2}\left\|\sum_{k\geq 1}\frac{1}{\widehat{\rho}_k+\lambda_2}\left\langle\widehat{\boldsymbol{\phi}}_k,\frac{\lambda_2}{\lambda_1}\boldsymbol{f}_{0\mathscr{S}}+\boldsymbol{\omega}_{\mathscr{S}}\right\rangle_2\widehat{\boldsymbol{\zeta}}_k\right\|^2\\
&=\frac{1}{n}\sum_{k\geq 1}\frac{\widehat{\rho}_k}{(\widehat{\rho}_k+\lambda_2)^2}\left\langle\widehat{\boldsymbol{\phi}}_k,\frac{\lambda_2}{\lambda_1}\boldsymbol{f}_{0\mathscr{S}}+\boldsymbol{\omega}_{\mathscr{S}}\right\rangle_2^2\quad\text{(by (S.32))}\\
&\leq\frac{1}{n\lambda_2}\left\|\frac{\lambda_2}{\lambda_1}\boldsymbol{f}_{0\mathscr{S}}+\boldsymbol{\omega}_{\mathscr{S}}\right\|_2^2\\
&\leq\frac{q}{n\lambda_2}\left(\lambda_2/\lambda_1+C_{\max}\right)^2.
\end{aligned}\tag{S.52}$$

By (S.50), (S.51), (S.52), and an application of Lemma S.14 (i) with $s=4/3$,

$$\sum_{j\in\mathscr{S}^c}\mathbb{E}\left\{\mathbb{P}\left(\|U_j\|_2\geq\frac{\xi C_{\min}}{3}\middle|\widetilde{\boldsymbol{X}}_{\mathscr{S}}\right)\right\}\leq 2(p-q)\exp\left\{-\frac{\lambda_2(C_{\min}/C_{\max})^2\xi^2 n}{24(1+\mu^{-1}C_{\max}^{-1})^2\tau q}\right\}.$$

This concludes the proof of (S.45). According to Lemma S.16, we have $\varkappa\geq\rho_1(\rho_1+\lambda_2)^{-1}$. Let $\xi=\delta=\gamma/3$, we find (S.38) is bounded by

$$\exp\left\{-\frac{\lambda_2^2\gamma^2 n}{36C^2(\rho_1+\lambda_2)^2 q}\right\}+2(p-q)\exp\left\{-\frac{\lambda_2(C_{\min}/C_{\max})^2\gamma^2 n}{864\tau q}\right\}.\tag{S.53}$$

Note that ρ_1 must be bounded from below by a universal constant, denoted as D_0^*. Without this lower bound, the model will only contain noise and no meaningful signals. Below, we will use D^* to denote a universal constant in $(0,\infty)$ whose value changes from line to line. Suppose λ_2 satisfies

$$\lambda_2 > \frac{6\max(1, D_{2,1}^*)}{(D_0^*)^{1/2}} \frac{C\tau^{1/2}(\rho_1+1)}{\gamma} \cdot \sqrt{\frac{q^2}{n}} > \frac{6C\tau^{1/2}(\rho_1+\lambda_2)}{\sqrt{\rho_1}\gamma} \cdot \sqrt{\frac{q^2}{n}}, \quad \text{(S.54)}$$

which meets Condition (S.39). It can be shown that the first term of (S.53) is bounded by $\exp\left(-D_a^{(2)}\frac{\lambda_2^2 n}{q}\right)$ for any $D_a^{(2)} \leq D^*\gamma^2(\rho_1+1)^{-2}$. Suppose for $d \in (0,1)$, λ_2 also satisfies

$$\lambda_2 > \frac{864\tau}{d(C_{\min}/C_{\max})^2\gamma^2} \cdot \frac{q\log(p-q)}{n}, \quad \text{(S.55)}$$

which is equivalent to

$$\frac{(C_{\min}/C_{\max})^2\gamma^2}{864\tau} \cdot \frac{\lambda_2 n}{q} - \log(p-q) > (1-d) \cdot \frac{(C_{\min}/C_{\max})^2\gamma^2}{864\tau} \cdot \frac{\lambda_2 n}{q}.$$

Then, the second term of (S.53) is bounded by

$$\begin{aligned}
&(p-q)\exp\left\{-\frac{\lambda_2(C_{\min}/C_{\max})^2\gamma^2 n}{864\tau q}\right\} \\
\leq& \exp\left\{-\frac{(1-d)(C_{\min}/C_{\max})^2\gamma^2}{864\tau} \cdot \frac{\lambda_2 n}{q}\right\} \\
\leq& \exp\left\{-\frac{(1-d)(C_{\min}/C_{\max})^2\gamma^2}{864 D_{2,1}^*\tau} \cdot \frac{\lambda_2^2 n}{q}\right\}
\end{aligned}$$

$$\leq \exp\left(-D_b^{(2)}\frac{\lambda_2^2 n}{q}\right),$$

where $D_b^{(2)} \leq D^*(1-d)(C_{\min}/C_{\max})^2\gamma^2\tau^{-1}$. The second inequality uses the fact $\lambda_2 < D_{2,1}^*$. It follows from Lemma S.1 with $d = 1/2$

$$\mathbb{P}\left(\max_{j\in\mathscr{S}^c}\left\|\mathscr{T}_n^{(j,\mathscr{S})}\left(\mathscr{T}_{n,\lambda_2}^{(\mathscr{S},\mathscr{S})}\right)^{-1}\left(\frac{\lambda_2}{\lambda_1}\boldsymbol{f}_{0\mathscr{S}}+\boldsymbol{\omega}_{\mathscr{S}}\right)\right\|_2 \geq \left(1-\frac{2\gamma}{9}\right)C_{\min}\right) \leq \exp\left(-D^{(2)}\frac{\lambda_2^2 n}{q}\right)$$

holds for any $D^{(2)}$ and λ_2 such that

$$D^{(2)} \leq D^*\frac{(C_{\min}/C_{\max})^2\gamma^2}{(\rho_1+1)^2\tau} \leq \min\left\{D_a^{(2)}, D_b^{(2)}\right\},$$

and

$$\lambda_2 > D^*\frac{\tau(\rho_1+1)}{(C_{\min}/C_{\max})^2\gamma^2}\max\left(\frac{q\log(p-q)}{n}, \sqrt{\frac{q^2}{n}}\right).$$

□

Proof of Lemma S.4

Proof. Define $\mathscr{E}^{(\mathscr{S},\mathscr{S})}$ to be the operator that only contains the off-diagonal elements of $\mathscr{T}^{(\mathscr{S},\mathscr{S})}$, i.e. $\mathscr{E}^{(\mathscr{S},\mathscr{S})} = \mathscr{T}^{(\mathscr{S},\mathscr{S})} - \mathscr{Q}^{(\mathscr{S},\mathscr{S})} = \mathscr{T}_{\lambda_2}^{(\mathscr{S},\mathscr{S})} - \mathscr{Q}_{\lambda_2}^{(\mathscr{S},\mathscr{S})}$.

Then

$$\begin{aligned}
&\left\|\left|(\mathscr{Q}^{(\mathscr{S},\mathscr{S})})^{1/2}(\mathscr{T}_{\lambda_2}^{(\mathscr{S},\mathscr{S})})^{-1}\right|\right\|_{\infty,\infty}\\
&=\left\|\left|(\mathscr{Q}^{(\mathscr{S},\mathscr{S})})^{1/2}(\mathscr{Q}_{\lambda_2}^{(\mathscr{S},\mathscr{S})})^{-1}+(\mathscr{Q}^{(\mathscr{S},\mathscr{S})})^{1/2}\left\{(\mathscr{T}_{\lambda_2}^{(\mathscr{S},\mathscr{S})})^{-1}-(\mathscr{Q}_{\lambda_2}^{(\mathscr{S},\mathscr{S})})^{-1}\right\}\right|\right\|_{\infty,\infty}\\
&=\left\|\left|(\mathscr{Q}^{(\mathscr{S},\mathscr{S})})^{1/2}(\mathscr{Q}_{\lambda_2}^{(\mathscr{S},\mathscr{S})})^{-1}-(\mathscr{Q}^{(\mathscr{S},\mathscr{S})})^{1/2}(\mathscr{Q}_{\lambda_2}^{(\mathscr{S},\mathscr{S})})^{-1}\mathscr{E}^{(\mathscr{S},\mathscr{S})}(\mathscr{T}_{\lambda_2}^{(\mathscr{S},\mathscr{S})})^{-1}\right|\right\|_{\infty,\infty}\\
&\leq\left\|\left|(\mathscr{Q}^{(\mathscr{S},\mathscr{S})})^{1/2}(\mathscr{Q}_{\lambda_2}^{(\mathscr{S},\mathscr{S})})^{-1}\right|\right\|_{\infty,\infty}\left(1+\left\|\left|\mathscr{E}^{(\mathscr{S},\mathscr{S})}(\mathscr{T}_{\lambda_2}^{(\mathscr{S},\mathscr{S})})^{-1}\right|\right\|_{\infty,\infty}\right).
\end{aligned}\tag{S.56}$$

Note that

$$\begin{aligned}
&\left\|\left|(\mathscr{Q}^{(\mathscr{S},\mathscr{S})})^{1/2}(\mathscr{Q}_{\lambda_2}^{(\mathscr{S},\mathscr{S})})^{-1}\right|\right\|_{\infty,\infty}\\
&=\max_{j\in\mathscr{S}}\left\|\left|(\mathscr{T}^{(j,j)})^{1/2}\left(\mathscr{T}^{(j,j)}+\lambda_2\mathscr{I}\right)^{-1}\right|\right\|_{2,2}\\
&=\max_{j\in\mathscr{S}}\sup_{\|f_j\|_2\leq 1}\left[\sum_{k\geq 1}\frac{\nu_{jk}}{(\nu_{jk}+\lambda_2)^2}\langle f_j,\eta_{jk}\rangle_2^2\right]^{1/2}\leq\frac{1}{2\sqrt{\lambda_2}}.
\end{aligned}\tag{S.57}$$

The last inequality holds by observing that the maximum value of function

$h(x) = x(x+\rho)^{-2}$ is $h(\rho) = (4\rho)^{-1}$. Meanwhile

$$\begin{aligned}
&\left|\!\left|\!\left|\mathscr{E}^{(\mathscr{S},\mathscr{S})}(\mathscr{T}_{\lambda_2}^{(\mathscr{S},\mathscr{S})})^{-1}\right|\!\right|\!\right|_{\infty,\infty} \\
=&\left|\!\left|\!\left|\mathscr{T}_{\lambda_2}^{(\mathscr{S},\mathscr{S})}(\mathscr{T}_{\lambda_2}^{(\mathscr{S},\mathscr{S})})^{-1} - \mathscr{Q}_{\lambda_2}^{(\mathscr{S},\mathscr{S})}(\mathscr{T}_{\lambda_2}^{(\mathscr{S},\mathscr{S})})^{-1}\right|\!\right|\!\right|_{\infty,\infty} \\
\leq&1 + \left|\!\left|\!\left|\mathscr{Q}_{\lambda_2}^{(\mathscr{S},\mathscr{S})}(\mathscr{Q}_{\lambda_2}^{(\mathscr{S},\mathscr{S})} + \mathscr{E}^{(\mathscr{S},\mathscr{S})})^{-1}\right|\!\right|\!\right|_{\infty,\infty} \\
=&1 + \left|\!\left|\!\left|\left\{\mathscr{I} + \mathscr{E}^{(\mathscr{S},\mathscr{S})}(\mathscr{Q}_{\lambda_2}^{(\mathscr{S},\mathscr{S})})^{-1}\right\}^{-1}\right|\!\right|\!\right|_{\infty,\infty}.
\end{aligned} \tag{S.58}$$

By Theorem 3.5.5, Hsing and Eubank (2015), $\mathscr{I} + \mathscr{E}^{(\mathscr{S},\mathscr{S})}(\mathscr{Q}_{\lambda_2}^{(\mathscr{S},\mathscr{S})})^{-1}$ is invertible if

$$\aleph(\lambda_2) = \left|\!\left|\!\left|\mathscr{E}^{(\mathscr{S},\mathscr{S})}(\mathscr{Q}_{\lambda_2}^{(\mathscr{S},\mathscr{S})})^{-1}\right|\!\right|\!\right|_{\infty,\infty} < 1,$$

which is warranted by Condition C.4. In this case,

$$\left|\!\left|\!\left|\left\{\mathscr{I} + \mathscr{E}^{(\mathscr{S},\mathscr{S})}(\mathscr{Q}_{\lambda_2}^{(\mathscr{S},\mathscr{S})})^{-1}\right\}^{-1}\right|\!\right|\!\right|_{\infty,\infty} < \frac{1}{1-\aleph(\lambda_2)}. \tag{S.59}$$

Therefore, (S.18) holds by (S.56)-(S.59). □

Proof of Lemma S.5

Proof. Recall that $g_j = n^{-1}\widetilde{\boldsymbol{X}}_{\bullet j}^{\top}\boldsymbol{\epsilon}_n$. Conditional on $\widetilde{\boldsymbol{X}}_{\bullet j}$, g_j is a rank n Gaussian process with mean zero and covariance operator $\mathscr{R}_j = n^{-1}\sigma^2\mathscr{T}_n^{(j,j)}$, and

$\mathrm{tr}(\mathscr{R}_j) = n^{-1}\mathrm{tr}(\mathscr{T}_n^{(j,j)})\sigma^2$. Define the event $\mathscr{D}_j(c_1) = \left\{\mathrm{tr}\left(\mathscr{T}_n^{(j,j)}\right) < c_1\right\}$, it follows that

$$\begin{aligned}\mathbb{P}\left(\|\boldsymbol{g}_{\mathscr{S}}\|_\infty > \lambda_2\right) &\leq \sum_{j\in\mathscr{S}} \mathbb{P}\Bigg(\|g_j\|_2 \geq \lambda_2\Bigg) \\ &\leq \sum_{j\in\mathscr{S}} \mathbb{E}\left[\mathbb{P}\left(\|g_j\|_2 \geq \lambda_2 \middle| \widetilde{\boldsymbol{X}}_{\bullet j}, \mathscr{D}_j(c_1)\right)\right] + \sum_{j\in\mathscr{S}} \mathbb{P}\left\{\mathscr{D}_j^c(c_1)\right\}.\end{aligned}$$

Setting $c_1 = 2\tau$ and applying Lemma S.13 (i) with $s = 4/3$, we get

$$\mathbb{P}\left(\|g_j\|_2 \geq \lambda_2 \middle| \widetilde{\boldsymbol{X}}_{\bullet j}, \mathscr{D}_j(c_1)\right) \leq 2\exp\left(-\frac{3n\lambda_2^2}{16\sigma^2\tau}\right).$$

Given that $\widetilde{X}_{ij} \overset{indep}{\sim} \mathscr{G}\mathscr{P}\left(0, \mathscr{T}^{(j,j)}\right)$ and $\left\|\left|\mathscr{T}^{(j,j)}\right|\right\|_{2,2} = 1$ together with the facts $\tau > 1$ and $\lambda_2 < D_{2,1}^*$,

$$\begin{aligned}\mathbb{P}\left(\mathscr{D}_j^c(c_1)\right) = \mathbb{P}\left(\sum_{i=1}^n \|\widetilde{X}_{ij}\|^2 > 2n\sigma_0^2\right) &\leq \exp\left(-\frac{\tau}{32}n\right) \\ \leq \exp\left(-\frac{n}{32}\right) &\leq \exp\left(-\frac{\lambda_2^2 n}{32(D_{2,1}^*)^2}\right)\end{aligned}$$

by Lemma S.13 (ii) with $s = 16/9$. Combining the two bounds entails

$$\mathbb{P}\left(\|\boldsymbol{g}_{\mathscr{S}}\|_\infty > \lambda_2\right) \leq 2q\exp\left(-\frac{3n\lambda_2^2}{16\sigma^2\tau}\right) + q\exp\left(-\frac{\lambda_2^2 n}{32(D_{2,1}^*)^2}\right).$$

Suppose

$$\lambda_2 > D_1^*(\sigma+1)\tau^{1/2}\cdot\sqrt{\frac{\log q}{n}} > \max\left(\frac{4}{3}\sqrt{6}\cdot\sigma\tau^{1/2},\ 8D_{2,1}^*\right)\cdot\sqrt{\frac{\log q}{n}},$$

which is equivalent to

$$\frac{3n\lambda_2^2}{16\sigma^2\tau} - \log q > \frac{3n\lambda_2^2}{32\sigma^2\tau}$$

and

$$\frac{n\lambda_2^2}{32(D_{2,1}^*)^2} - \log q > \frac{n\lambda_2^2}{64(D_{2,1}^*)^2}.$$

We have (S.19) holds for some $D^{(3)} < D_2^*\left((\sigma+1)^2\tau\right)^{-1} < 64^{-1}\min\left\{6\sigma^{-2}\tau^{-1}, (D_{2,1}^*)^{-2}\right\}$, where D_1^* and D_2^* are universal constants. □

Proof of Lemma S.6

Proof. Recall $\left|\left|\left|\mathscr{T}^{(\mathscr{S},\mathscr{S})}\right|\right|\right|_{2,2} = \rho_1$ and define $t = \frac{u^2 n}{C^2\rho_1^2 q}$ for some constant $C > 0$, then by Corollary 2 in Koltchinskii and Lounici (2017),

$$\begin{aligned}
&\mathbb{P}\left(\sqrt{q}\left\|\left\|\boldsymbol{\mathscr{T}}^{(\mathscr{S},\mathscr{S})}-\boldsymbol{\mathscr{T}}_n^{(\mathscr{S},\mathscr{S})}\right\|\right\|_{2,2}\geq u\right)\\
=&\mathbb{P}\left(\left\|\left\|\boldsymbol{\mathscr{T}}^{(\mathscr{S},\mathscr{S})}-\boldsymbol{\mathscr{T}}_n^{(\mathscr{S},\mathscr{S})}\right\|\right\|_{2,2}\geq C\left\|\left\|\boldsymbol{\mathscr{T}}^{(\mathscr{S},\mathscr{S})}\right\|\right\|_{2,2}\sqrt{\frac{t}{n}}\right) \\
\leq& e^{-t}
\end{aligned} \tag{S.60}$$

as long as

$$\sqrt{\frac{t}{n}}=\max\left(\sqrt{\frac{r(\boldsymbol{\mathscr{T}}^{(\mathscr{S},\mathscr{S})})}{n}},\frac{r(\boldsymbol{\mathscr{T}}^{(\mathscr{S},\mathscr{S})})}{n},\sqrt{\frac{t}{n}},\frac{t}{n}\right), \tag{S.61}$$

where

$$r(\boldsymbol{\mathscr{T}}^{(\mathscr{S},\mathscr{S})})=\frac{(\mathbb{E}\|X_1\|_2)^2}{\|\|\boldsymbol{\mathscr{T}}^{(\mathscr{S},\mathscr{S})}\|\|_{2,2}}\leq\frac{\mathbb{E}\|X_1\|_2^2}{\|\|\boldsymbol{\mathscr{T}}^{(\mathscr{S},\mathscr{S})}\|\|_{2,2}}\leq\frac{q\tau}{\rho_1}$$

by Jensen's inequality and Condition C.2. Hence (S.61) holds when

$$\frac{q\tau}{\rho_1}<t<n,$$

which amounts to (S.20).

□

S.2.4 Additional technical lemmas

Lemma S.7. *For any* $0 < \nu < 1$,

$$\left\| \left(\mathscr{T}^{(\mathscr{S},\mathscr{S})}\right)^{\nu} \left(\widetilde{\boldsymbol{f}}_{\mathscr{S}} - \boldsymbol{f}_{0\mathscr{S}}\right) \right\|_2 \leq (1-\nu)^{1-\nu}\nu^{\nu}\lambda_3^{\nu} \left\|\boldsymbol{f}_{0\mathscr{S}}\right\|_2 .$$

Proof. Write $\mathscr{T}^{(\mathscr{S},\mathscr{S})} = \sum_{k\geq 1} \rho_k \boldsymbol{\phi}_k \otimes \boldsymbol{\phi}_k$ and $\boldsymbol{f}_{0\mathscr{S}} = \sum_{k\geq 1} f_k \boldsymbol{\phi}_k$. Then

$$\widetilde{\boldsymbol{f}}_{\mathscr{S}} = \sum_{k\geq 1} \frac{\rho_k f_k}{\lambda_3 + \rho_k} \boldsymbol{\phi}_k.$$

Therefore

$$\begin{aligned} \left\| \left(\mathscr{T}^{(\mathscr{S},\mathscr{S})}\right)^{\nu} \left(\widetilde{\boldsymbol{f}}_{\mathscr{S}} - \boldsymbol{f}_{0\mathscr{S}}\right) \right\|_2^2 &= \sum_{k\geq 1} \rho_k^{2\nu} \left(\frac{\lambda_3 f_k}{\lambda_3 + \rho_k}\right)^2 \leq \max_{k\geq 1} \frac{\lambda_3^2 \rho_k^{2\nu}}{(\lambda_3+\rho_k)^2} \sum_{k\geq 1} f_k^2 \\ &\leq (1-\nu)^{2(1-\nu)}\nu^{2\nu}\lambda_3^{2\nu} \left\|\boldsymbol{f}_{0\mathscr{S}}\right\|_2^2 . \end{aligned}$$

The last inequality follows from Young's inequality: $\lambda_3+\rho_k \geq (1-\nu)^{-(1-\nu)}\nu^{-\nu}\lambda_3^{1-\nu}\rho_k^{\nu}$.

□

Lemma S.8. *For* $0 < \nu < 1$,

$$\left\| \left\| \left(\mathscr{T}^{(\mathscr{S},\mathscr{S})}\right)^{\nu} \left(\mathscr{T}_{\lambda}^{(\mathscr{S},\mathscr{S})}\right)^{-1} \right\| \right\|_{2,2} \leq (1-\nu)^{1-\nu}\nu^{\nu}\lambda^{\nu-1}.$$

Proof. For any $\boldsymbol{f} \in \mathbb{L}_2^q$ such that $\|\boldsymbol{f}\|_2 \leq 1$, write $\boldsymbol{f} = \sum_{k\geq 1} f_k \boldsymbol{\phi}_k$, we have

$$\left\| \left(\mathscr{T}^{(\mathscr{S},\mathscr{S})}\right)^{\nu} \left(\mathscr{T}_{\lambda}^{(\mathscr{S},\mathscr{S})}\right)^{-1} \boldsymbol{f} \right\|_2 = \sqrt{\sum_{k\geq 1} \frac{\rho_k^{2\nu}}{(\rho_k+\lambda)^2} f_k^2} \leq \max_{k\geq 1} \left\{ \frac{\rho_k^{\nu}}{\rho_k+\lambda} \right\} \leq (1-\nu)^{1-\nu} \nu^{\nu} \lambda^{\nu-1}.$$

□

Lemma S.9. *Assume Condition C.5-C.6 hold. For $0 < \nu \leq 1/2$, $r > 1/2$*

$$\left\| \left(\mathscr{T}^{(\mathscr{S},\mathscr{S})}\right)^{\nu} \left(\mathscr{T}_{\lambda_3}^{(\mathscr{S},\mathscr{S})}\right)^{-1} \boldsymbol{g}_{\mathscr{S}} \right\|_2 = O_p\left(\left(\frac{n}{q} \cdot \lambda_3^{1-2\nu+\frac{1}{2r}} \right)^{-\frac{1}{2}} \right).$$

Proof. For $0 \leq \nu \leq 1/2$,

$$\begin{aligned}
\left\| \left(\mathscr{T}^{(\mathscr{S},\mathscr{S})}\right)^{\nu} \left(\mathscr{T}_{\lambda_3}^{(\mathscr{S},\mathscr{S})}\right)^{-1} \boldsymbol{g}_{\mathscr{S}} \right\|_2^2 &= \sum_{k\geq 1} \left\langle \left(\mathscr{T}^{(\mathscr{S},\mathscr{S})}\right)^{\nu} \left(\mathscr{T}_{\lambda_3}^{(\mathscr{S},\mathscr{S})}\right)^{-1} \boldsymbol{g}_{\mathscr{S}},\ \boldsymbol{\phi}_k \right\rangle_2^2 \\
&= \sum_{k\geq 1} \left\langle \left(\mathscr{T}^{(\mathscr{S},\mathscr{S})}\right)^{\nu} \left(\mathscr{T}_{\lambda_3}^{(\mathscr{S},\mathscr{S})}\right)^{-1} \boldsymbol{\phi}_k,\ \boldsymbol{g}_{\mathscr{S}}, \right\rangle_2^2 \\
&= \sum_{k\geq 1} \left\langle \frac{\rho_k^{\nu}}{\rho_k+\lambda_3} \boldsymbol{\phi}_k,\ \frac{1}{n} \sum_{i=1}^{n} \epsilon_i \widetilde{\boldsymbol{X}}_{i\mathscr{S}} \right\rangle_2^2 \\
&= \sum_{k\geq 1} \frac{\rho_k^{2\nu}}{(\rho_k+\lambda_3)^2} \left\{ \frac{1}{n} \sum_{i=1}^{n} \epsilon_i \left\langle \boldsymbol{\phi}_k,\ \widetilde{\boldsymbol{X}}_{i\mathscr{S}} \right\rangle_2 \right\}^2.
\end{aligned}$$

Therefore

$$\mathbb{E} \left\| \left(\mathscr{T}^{(\mathscr{S},\mathscr{S})}\right)^{\nu} \left(\mathscr{T}_{\lambda_3}^{(\mathscr{S},\mathscr{S})}\right)^{-1} \boldsymbol{g}_{\mathscr{S}} \right\|_2^2 = \frac{\sigma^2}{n} \sum_{k\geq 1} \frac{\rho_k^{2\nu}}{(\rho_k+\lambda_3)^2} \cdot \mathbb{E} \left\langle \boldsymbol{\phi}_k,\ \widetilde{\boldsymbol{X}}_{1\mathscr{S}} \right\rangle_2^2$$

$$
\begin{aligned}
&= \frac{\sigma^2}{n} \sum_{k\geq 1} \frac{\rho_k^{2\nu+1}}{(\rho_k+\lambda_3)^2} \\
&\leq \frac{\sigma^2}{n\lambda_3^{1-2\nu}} \sum_{k\geq 1} \frac{\rho_k^{2\nu+1}}{(\rho_k+\lambda_3)^{1+2\nu}} \\
&\leq C\sigma^2 \left((n/q)\cdot \lambda_3^{1-2\nu+\frac{1}{2r}} \right)^{-1}
\end{aligned}
$$

for some constant $C > 0$. The last inequality is obtained by Lemma S.11. The proof can be completed by Markov inequality. □

Lemma S.10. *Assume Condition C.5-C.6 hold. Then for any $r > 1/2$, $0 < \nu < 1/2 - 1/(4r)$,*

$$
(1).\quad \left\| \left(\mathscr{T}^{(\mathscr{S},\mathscr{S})} \right)^{\nu} \left(\mathscr{T}_{\lambda_3}^{(\mathscr{S},\mathscr{S})} \right)^{-1} \left(\mathscr{T}_n^{(\mathscr{S},\mathscr{S})} - \mathscr{T}^{(\mathscr{S},\mathscr{S})} \right) \left(\mathscr{T}^{(\mathscr{S},\mathscr{S})} \right)^{-\nu} \right\|_{2,2} = O_p\left(q^{\frac{1}{2}} \left(\frac{n}{q} \lambda_3^{1-2\nu+\frac{1}{2r}} \right)^{-\frac{1}{2}} \right).
$$

$$
(2).\quad \left\| \left(\mathscr{T}^{(\mathscr{S},\mathscr{S})} \right)^{1/2} \left(\mathscr{T}_{\lambda_3}^{(\mathscr{S},\mathscr{S})} \right)^{-1} \left(\mathscr{T}_n^{(\mathscr{S},\mathscr{S})} - \mathscr{T}^{(\mathscr{S},\mathscr{S})} \right) \left(\mathscr{T}^{(\mathscr{S},\mathscr{S})} \right)^{-\nu} \right\|_{2,2} = O_p\left(q^{\frac{1}{2}} \left(\frac{n}{q} \lambda_3^{\frac{1}{2r}} \right)^{-\frac{1}{2}} \right).
$$

Proof. **(1).** Write $\boldsymbol{g} = \sum_{k\geq 1} g_k \boldsymbol{\phi}_k$, $\boldsymbol{h} = \sum_{k\geq 1} h_k \boldsymbol{\phi}_k$. We have

$$
\begin{aligned}
&\left\| \left(\mathscr{T}^{(\mathscr{S},\mathscr{S})} \right)^{\nu} \left(\mathscr{T}_{\lambda_3}^{(\mathscr{S},\mathscr{S})} \right)^{-1} \left(\mathscr{T}_n^{(\mathscr{S},\mathscr{S})} - \mathscr{T}^{(\mathscr{S},\mathscr{S})} \right) \left(\mathscr{T}^{(\mathscr{S},\mathscr{S})} \right)^{-\nu} \right\|_{2,2} \\
&= \sup_{\|g\|\leq 1, \|h\|\leq 1} \left| \left\langle \boldsymbol{g},\ \left(\mathscr{T}^{(\mathscr{S},\mathscr{S})} \right)^{\nu} \left(\mathscr{T}_{\lambda_3}^{(\mathscr{S},\mathscr{S})} \right)^{-1} \left(\mathscr{T}_n^{(\mathscr{S},\mathscr{S})} - \mathscr{T}^{(\mathscr{S},\mathscr{S})} \right) \left(\mathscr{T}^{(\mathscr{S},\mathscr{S})} \right)^{-\nu} \boldsymbol{h} \right\rangle_2 \right| \\
&= \sup_{\|g\|\leq 1, \|h\|\leq 1} \left| \left\langle \left(\mathscr{T}_{\lambda_3}^{(\mathscr{S},\mathscr{S})} \right)^{-1} \left(\mathscr{T}^{(\mathscr{S},\mathscr{S})} \right)^{\nu} \boldsymbol{g},\ \left(\mathscr{T}_n^{(\mathscr{S},\mathscr{S})} - \mathscr{T}^{(\mathscr{S},\mathscr{S})} \right) \left(\mathscr{T}^{(\mathscr{S},\mathscr{S})} \right)^{-\nu} \boldsymbol{h} \right\rangle_2 \right| \\
&= \sup_{\|g\|\leq 1, \|h\|\leq 1} \left| \left\langle \sum_{k\geq 1} \frac{\rho_k^{\nu} g_k}{\rho_k+\lambda_3} \boldsymbol{\phi}_k,\ \sum_{l\geq 1} \rho_l^{-\nu} h_k \left(\mathscr{T}_n^{(\mathscr{S},\mathscr{S})} - \mathscr{T}^{(\mathscr{S},\mathscr{S})} \right) \boldsymbol{\phi}_l \right\rangle_2 \right|
\end{aligned}
$$

$$
\begin{aligned}
&= \sup_{\|g\|\leq 1, \|h\|\leq 1} \left| \sum_{k,l\geq 1} \frac{\rho_k^{\nu}\rho_l^{-\nu} g_k h_l}{\rho_k+\lambda_3} \left\langle \boldsymbol{\phi}_k,\ \left(\mathscr{T}_n^{(\mathscr{S},\mathscr{S})} - \mathscr{T}^{(\mathscr{S},\mathscr{S})}\right)\boldsymbol{\phi}_l \right\rangle_2 \right| \\
&\leq \sup_{\|g\|\leq 1, \|h\|\leq 1} \left| \sum_{k,l\geq 1} g_k^2 h_l^2 \right|^{1/2} \left| \sum_{k,l\geq 1} \frac{\rho_k^{2\nu}\rho_l^{-2\nu}}{(\rho_k+\lambda_3)^2} \left\langle \boldsymbol{\phi}_k,\ \left(\mathscr{T}_n^{(\mathscr{S},\mathscr{S})} - \mathscr{T}^{(\mathscr{S},\mathscr{S})}\right)\boldsymbol{\phi}_l \right\rangle_2^2 \right|^{1/2} \\
&\leq \left| \sum_{k,l\geq 1} \frac{\rho_k^{2\nu}\rho_l^{-2\nu}}{(\rho_k+\lambda_3)^2} \left\langle \boldsymbol{\phi}_k,\ \left(\mathscr{T}_n^{(\mathscr{S},\mathscr{S})} - \mathscr{T}^{(\mathscr{S},\mathscr{S})}\right)\boldsymbol{\phi}_l \right\rangle_2^2 \right|^{1/2}.
\end{aligned}
$$

The second inequality from the bottom follows from the Cauchy-Schwarz inequality. By Jensen's inequality

$$
\begin{aligned}
&\mathbb{E}\left| \sum_{k,l\geq 1} \frac{\rho_k^{2\nu}\rho_l^{-2\nu}}{(\rho_k+\lambda_3)^2} \left\langle \boldsymbol{\phi}_k,\ \left(\mathscr{T}_n^{(\mathscr{S},\mathscr{S})} - \mathscr{T}^{(\mathscr{S},\mathscr{S})}\right)\boldsymbol{\phi}_l \right\rangle_2^2 \right|^{1/2} \\
&\leq \left| \sum_{k,l\geq 1} \frac{\rho_k^{2\nu}\rho_l^{-2\nu}}{(\rho_k+\lambda_3)^2} \mathbb{E}\left\langle \boldsymbol{\phi}_k,\ \left(\mathscr{T}_n^{(\mathscr{S},\mathscr{S})} - \mathscr{T}^{(\mathscr{S},\mathscr{S})}\right)\boldsymbol{\phi}_l \right\rangle_2^2 \right|^{1/2}.
\end{aligned}
$$

Note that

$$
\begin{aligned}
&\mathbb{E}\left\langle \boldsymbol{\phi}_k,\ \left(\mathscr{T}_n^{(\mathscr{S},\mathscr{S})} - \mathscr{T}^{(\mathscr{S},\mathscr{S})}\right)\boldsymbol{\phi}_l \right\rangle_2^2 \\
&= \mathbb{E}\left\langle \boldsymbol{\phi}_k,\ \left(\frac{1}{n}\sum_{i=1}^n \widetilde{\boldsymbol{X}}_{i\mathscr{S}} \otimes \widetilde{\boldsymbol{X}}_{i\mathscr{S}}^{\top} - \mathbb{E}\widetilde{\boldsymbol{X}}_{1\mathscr{S}} \otimes \widetilde{\boldsymbol{X}}_{1\mathscr{S}}^{\top}\right)\boldsymbol{\phi}_l \right\rangle_2^2 \\
&= \frac{1}{n}\mathbb{E}\left\langle \boldsymbol{\phi}_k,\ \left(\widetilde{\boldsymbol{X}}_{1\mathscr{S}} \otimes \widetilde{\boldsymbol{X}}_{1\mathscr{S}}^{\top} - \mathbb{E}\widetilde{\boldsymbol{X}}_{1\mathscr{S}} \otimes \widetilde{\boldsymbol{X}}_{1\mathscr{S}}^{\top}\right)\boldsymbol{\phi}_l \right\rangle_2^2 \\
&\leq \frac{1}{n}\mathbb{E}\left\langle \boldsymbol{\phi}_k,\ \left(\widetilde{\boldsymbol{X}}_{1\mathscr{S}} \otimes \widetilde{\boldsymbol{X}}_{1\mathscr{S}}^{\top}\right)\boldsymbol{\phi}_l \right\rangle_2^2 \\
&= \frac{1}{n}\mathbb{E}\left\langle \boldsymbol{\phi}_k,\ \widetilde{\boldsymbol{X}}_{1\mathscr{S}} \right\rangle_2^2 \left\langle \boldsymbol{\phi}_l,\ \widetilde{\boldsymbol{X}}_{1\mathscr{S}} \right\rangle_2^2 \\
&\leq \frac{1}{n}\mathbb{E}^{1/2}\left\langle \boldsymbol{\phi}_k,\ \widetilde{\boldsymbol{X}}_{1\mathscr{S}} \right\rangle_2^4 \mathbb{E}^{1/2}\left\langle \boldsymbol{\phi}_l,\ \widetilde{\boldsymbol{X}}_{1\mathscr{S}} \right\rangle_2^4
\end{aligned}
$$

$$
\begin{aligned}
&= \frac{3}{n}\mathbb{E}\left\langle \boldsymbol{\phi}_k,\ \widetilde{\boldsymbol{X}}_{1\mathscr{S}} \right\rangle_2^2 \mathbb{E}\left\langle \boldsymbol{\phi}_l,\ \widetilde{\boldsymbol{X}}_{1\mathscr{S}} \right\rangle_2^2 \\
&= \frac{3}{n}\rho_k\rho_l.
\end{aligned}
$$

The last inequality follows from the Cauchy-Schwarz inequality. The second-to-last equality from the bottom is derived from the property of Gaussian kurtosis, $\mathbb{E}\langle \boldsymbol{\phi}_k, \widetilde{\boldsymbol{X}}_{1\mathscr{S}}\rangle_2^4 = 3\left(\mathbb{E}\langle \boldsymbol{\phi}_k, \widetilde{\boldsymbol{X}}_{1\mathscr{S}}\rangle_2^2\right)^2$, where $\langle \boldsymbol{\phi}_k, \widetilde{\boldsymbol{X}}_{1\mathscr{S}}\rangle_2$ follows a Gaussian distribution with mean 0 and variance smaller than ρ_1. Therefore

$$
\begin{aligned}
&\mathbb{E}\left|\left|\left| \left(\mathscr{T}^{(\mathscr{S},\mathscr{S})}\right)^{\nu} \left(\mathscr{T}_{\lambda_3}^{(\mathscr{S},\mathscr{S})}\right)^{-1} \left(\mathscr{T}_n^{(\mathscr{S},\mathscr{S})} - \mathscr{T}^{(\mathscr{S},\mathscr{S})}\right) \left(\mathscr{T}^{(\mathscr{S},\mathscr{S})}\right)^{-\nu} \right|\right|\right|_{2,2} \\
&\leq \left(\frac{3}{n}\sum_{k\geq 1}\frac{\rho_k^{1+2\nu}}{(\rho_k+\lambda_3)^2}\sum_{l\geq 1}\rho_l^{1-2\nu}\right)^{1/2} \\
&\leq \left(3\sum_{l\geq 1}\rho_l^{1-2\nu}\right)^{1/2}\left(\frac{1}{n\lambda_3^{1-2\nu}}\sum_{k\geq 1}\frac{\rho_k^{1+2\nu}}{(\rho_k+\lambda_3)^{1+2\nu}}\right)^{1/2}. \qquad \text{(S.62)}
\end{aligned}
$$

By Corollary 2, we have

$$
\sum_{l\geq 1}\rho_l^{1-2\nu} = \sum_{j=1}^{q}\sum_{k\geq 1}\left(\rho_{q(k-1)+j}\right)^{1-2\nu} \leq (bc)^{1-2\nu} q \sum_{k\geq 1} k^{-2r(1-2\nu)} = O(q).
$$

The last equation holds because $1-2\nu > 1/(2r)$. By Lemma S.11, the expression (S.62) can be bounded by $Cq^{1/2}\left((n/q)\cdot\lambda_3^{1-2\nu+\frac{1}{2r}}\right)^{-1/2}$ for some $C > 0$. The proof is completed by applying the Markov inequality.

(2). Similarly, we can show that

$$
\begin{aligned}
&\mathbb{E}\left\|\left|\left(\mathscr{T}^{(\mathscr{S},\mathscr{S})}\right)^{1/2}\left(\mathscr{T}_{\lambda_3}^{(\mathscr{S},\mathscr{S})}\right)^{-1}\left(\mathscr{T}_n^{(\mathscr{S},\mathscr{S})}-\mathscr{T}^{(\mathscr{S},\mathscr{S})}\right)\left(\mathscr{T}^{(\mathscr{S},\mathscr{S})}\right)^{-\nu}\right|\right\|_{2,2} \\
&\leq\left(3\sum_{l\geq 1}\rho_l^{1-2\nu}\right)^{1/2}\left(\frac{1}{n}\sum_{k\geq 1}\frac{\rho_k^2}{(\rho_k+\lambda_3)^2}\right)^{1/2} \\
&\leq C' q^{1/2}\left((n/q)\cdot\lambda_3^{\frac{1}{2r}}\right)^{-1/2}
\end{aligned}
$$

for some $C'>0$. The proof is completed by applying the Markov inequality.

□

Lemma S.11. *For $\lambda<1$, suppose $\mathscr{T}^{(\mathscr{S},\mathscr{S})}$ satisfies Condition C.5, $\{\rho_l\}_{l\geq 1}$ are the eigenvalues of $\mathscr{T}^{(\mathscr{S},\mathscr{S})}$. Then there exist constants $c'>0$ depending only on b,c,r,ν such that*

$$
\sum_{l\geq 1}\frac{\rho_l^{1+2\nu}}{(\lambda+\rho_l)^{1+2\nu}}\leq c' q\left(1+\lambda^{-1/(2r)}\right),
$$

where b,c are defined in Condition C.5.

Proof. Let $C=bc$, according to Corollary 2, it is straightforward that

$$
\begin{aligned}
\sum_{l\geq 1}\frac{\rho_l^{1+2\nu}}{(\lambda+\rho_l)^{1+2\nu}} &= \sum_{j=1}^{q}\sum_{k\geq 1}\left(\frac{\rho_{q(k-1)+j}}{\lambda+\rho_{q(k-1)+j}}\right)^{1+2\nu} \\
&\leq q\sum_{k\geq 1}\left(\frac{Ck^{-2r}}{\lambda+Ck^{-2r}}\right)^{1+2\nu}
\end{aligned}
$$

$$
\begin{aligned}
&= qC^{1+2\nu} \sum_{k\geq 1} \frac{1}{(\lambda k^{2r}+C)^{1+2\nu}} \\
&\leq qC^{1+2\nu} \left(C^{-(1+2\nu)} + \int_1^\infty \frac{dx}{(\lambda x^{2r}+C)^{1+2\nu}} \right) \\
&\leq qC^{1+2\nu} \left(C^{-(1+2\nu)} + \lambda^{-\frac{1}{2r}} \int_0^\infty \frac{dy}{(y^{2r}+C)^{1+2\nu}} \right) \\
&< qc' \left(1 + \lambda^{-\frac{1}{2r}} \right).
\end{aligned}
$$

The last inequality holds because for $r > 1/2$,

$$
\int_0^\infty \frac{dy}{(y^{2r}+C)^{1+2\nu}} < \sum_{k=0}^\infty \frac{1}{(k^{2r}+C)^{1+2\nu}} < C^{-(1+2\nu)} + \sum_{k=1}^\infty k^{-2r(1+2\nu)} < \infty.
$$

□

Lemma S.12. *Suppose that $\boldsymbol{U}_1, \boldsymbol{U}_2$ are jointly Gaussian processes with means $\boldsymbol{\mu}_1, \boldsymbol{\mu}_2$, (auto) covariance operators $\mathscr{G}_{11}, \mathscr{G}_{22}$ and cross covariance operator $\mathscr{G}_{12} = \mathscr{G}_{21}^*$. Then, conditional on $\boldsymbol{U}_1$, $\boldsymbol{U}_2$ is a Gaussian process with mean $\boldsymbol{\mu}_2 + \mathscr{G}_{21}\mathscr{G}_{11}^-(\boldsymbol{U}_1 - \boldsymbol{\mu}_1)$ and covariance operator $\mathscr{G}_{22} - \mathscr{G}_{21}\mathscr{G}_{11}^-\mathscr{G}_{12}$, where $\mathscr{G}_{11}^-$ is the Moore-Penrose generalized inverse of $\mathscr{G}_{11}$, and therefore*

$$
\boldsymbol{U}_2 \overset{d}{=} \boldsymbol{\mu}_2 + \mathscr{G}_{21}\mathscr{G}_{11}^-(\boldsymbol{U}_1 - \boldsymbol{\mu}_1) + \boldsymbol{Z}
$$

where $\boldsymbol{Z}$ is a zero-mean process independent of U_1 and has covariance operator $\mathscr{G}_{22} - \mathscr{G}_{21}\mathscr{G}_{11}^-\mathscr{G}_{12}$.

Lemma S.13. *Suppose* $U_l \overset{iid}{\sim} \mathscr{GP}(0, \mathscr{G})$, $l = 1, \ldots, L$, *with* $\mathrm{tr}(\mathscr{G}) < \infty$, *then for any* $s > 1$,

(i)

$$\mathbb{P}\left(\sum_{l=1}^{L} \|U_l\|_2^2 > x\right) \leq \left(\frac{s}{s-1}\right)^{L/2} \exp\left(-\frac{x}{2s \cdot \mathrm{tr}(\mathscr{G})}\right);$$

(ii) if we further have $x > (1 + s/2)L \cdot \mathrm{tr}(\mathscr{G})$, *then*

$$\mathbb{P}\left(\sum_{l=1}^{L} \|U_l\|_2^2 > x\right) \leq \exp\left(-\frac{(1 - s^{-1/2})^2}{2\|\mathscr{G}\|_2}(x - L \cdot \mathrm{tr}(\mathscr{G}))\right).$$

The proof of this result is a straightforward application of the following Lemma S.14.

Lemma S.14. *Suppose that* $\xi_{lk}, 1 \leq m \leq L, 1 \leq k \leq K$, *are independent random variables where* $L < \infty, K \leq \infty$, $\xi_{lk} \sim N(0, \theta_k)$ *for all* l, k *with* $\|\theta\|_1 < \infty$, *where* $\|\theta\|_1 = \sum_{k=1}^{K} \theta_k$, *further define* $\|\theta\|_\infty = \max_{\{k=1,\ldots,K\}} \theta_k$, *then for any* $s > 1$,

(i)

$$\mathbb{P}\left(\sum_{l=1}^{L} \sum_{k=1}^{K} \xi_{lk}^2 > x\right) \leq \left(\frac{s}{s-1}\right)^{L/2} \exp\left(-\frac{x}{2s\|\theta\|_1}\right); \tag{S.63}$$

(ii) if we further have $x > (1+s/2)L\|\theta\|_1$, then

$$\mathbb{P}\left(\sum_{l=1}^{L}\sum_{k=1}^{K}\xi_{lk}^2 > x\right) \le \exp\left(-\frac{(1-s^{-1/2})^2}{2\|\theta\|_\infty}(x - L\|\theta\|_1)\right). \qquad \text{(S.64)}$$

Proof. For (i), by Markov's inequality,

$$\mathbb{P}\left(\sum_{l=1}^{L}\sum_{k=1}^{K}\xi_{lk}^2 > x\right) \le e^{-tx}\left\{\prod_{k=1}^{K}\mathbb{E}\left(e^{t\xi_{1k}^2}\right)\right\}^L = e^{-tx}\prod_{k=1}^{K}(1-2t\theta_k)^{-L/2}.$$

Letting $t = (2s\sum_{k=1}^{\infty}\theta_k)^{-1}$, $s > 1$, we obtain

$$\prod_{k=1}^{K}(1-2t\theta_k)^{-L/2} = \prod_{k=1}^{K}\left(1 - \frac{\theta_k}{s\sum_{k=1}^{K}\theta_k}\right)^{-L/2} \le \left(\frac{s}{s-1}\right)^{L/2},$$

where the maximum is attained when $\theta_1 \neq 0, \theta_2 = \theta_3 = \cdots = 0$. To see why the above statement is true, define $r_k = \theta_k(\sum_{k=1}^{K}\theta_k)^{-1}$, then we have $0 \le r_k \le 1$, $\sum_{k=1}^{K} r_k = 1$, denote $\boldsymbol{r}_K = (r_1, \ldots, r_K)^\top$, define

$$g_K(\boldsymbol{r}_K) = -\frac{L}{2}\sum_{k=1}^{K}\log\left(1 - \frac{r_k}{s}\right).$$

It is straightforward to determine that the function g_K has a compact support and is differentiable. By setting the gradient of g_K with respect to $\boldsymbol{r}_K$ equal to zero, we obtain $r_k \equiv 1/K$, $k = 1, \ldots, K$, and this leads to the attainment of the function's minimum value. Note that function g_K only have one critical point, as a result, the maximum value must be attained at the boundary of the support of $\boldsymbol{r}_K$. Without loss of generality, we have $r_K = 0$, then the minimum value of g_{K-1} is attained at $r_k \equiv 1/(K-1)$, $k = 1, \ldots, K-1$, the maximum value of g_{K-1} must be attained at the boundary of $\boldsymbol{r}_{K-1}$. Recursively using this fact, we have $r_1 = 1$, $r_2 = \cdots = r_K = 0$.

For (ii), the proof utilizes a modified version of the Laurent-Massart inequality (Laurent and Massart, 2000), as follows. Suppose $Z_j \overset{i.i.d.}{\sim} N(0,1)$, $a_j \geq 0$ $(j = 1, \ldots, n)$, define $c = 2\|a\|_\infty$ and $v^2 = 2\|a\|_2^2$. Then, for any $y > 0$,

$$\mathbb{P}\left(\sum_{j=1}^{n} a_j(Z_j^2 - 1) > y\right) \leq \exp\left\{-\frac{v^2}{2c^2}\left((1 + 2v^{-2}cy)^{1/2} - 1\right)^2\right\}.$$

Back to our setting, letting $\xi_{lk} = \theta_k^{1/2} Z_{lk}$, $v^2 = 2L\|\theta\|_2^2$, $c = 2\|\theta\|_\infty$, and assuming $y > 2^{-1}sL\|\theta\|_1$ $(s > 1)$, we have $2cy/s > 2L\|\theta\|_1\|\theta\|_\infty \geq 2L\|\theta\|_2^2 = v^2$. Then, $2v^{-2}cy > s > 1$, and in this case

$$\frac{v^2}{2c^2}\left((1 + 2v^{-2}cy)^{1/2} - 1\right)^2 > \frac{v^2}{2c^2}\left((2v^{-2}cy)^{1/2} - 1\right)^2 > \frac{\left(1 - s^{-1/2}\right)^{-2}}{c} y.$$

Subsequently,

$$\mathbb{P}\left(\sum_{l=1}^{L}\sum_{k=1}^{K}(\xi_{lk}^2 - \theta_k) > y\right) \leq \exp\left(-\frac{(1 - s^{-1/2})^2}{2\|\theta\|_\infty} y\right).$$

Let $x = y + L\|\theta\|_1$. Then, for $x > (1 + s/2)L\|\theta\|_1$, (S.64) holds.

$\square$

The proofs of the following lemmas are straightforward and are omitted.

Lemma S.15. *For operator-valued matrices* $\boldsymbol{A}$ *and* $\boldsymbol{B}$*,*

(i) $|||\boldsymbol{A}\boldsymbol{B}|||_{\alpha,\beta} \leq |||\boldsymbol{A}|||_{\eta,\beta}|||\boldsymbol{B}|||_{\alpha,\eta}$ *for* $\alpha, \beta, \eta \in \{2, \infty\}$*;*

(ii) if $\boldsymbol{A}$ *has dimension* $q \times q$*, then* $\frac{1}{\sqrt{q}}|||\boldsymbol{A}|||_{2,2} \leq |||\boldsymbol{A}|||_{\infty,\infty} \leq \sqrt{q}|||\boldsymbol{A}|||_{2,2}$.

Lemma S.16. *For a* $q \times q$ *operator-valued covariance matrix* $\boldsymbol{R}$*, suppose* ρ_1 *is the largest eigenvalue of* $\boldsymbol{R}$*, then for any* $\lambda > 0$

$$\left|\left|\left|\boldsymbol{R}(\boldsymbol{R} + \lambda\boldsymbol{\mathscr{I}})^{-1}\right|\right|\right|_{\infty,\infty} \geq \frac{\rho_1}{\rho_1 + \lambda}.$$

S.3. Substantiating examples for the technical conditions

We now provide examples of functional predictors that satisfy technical conditions such as C.3 and C.4. As described in Remark 2, we consider functional predictors with partially separable covariance structure (Zapata et al., 2021) such that

$$\boldsymbol{\mathscr{T}}^{(\mathscr{S},\mathscr{S})} = \sum_{k=1}^{\infty} \boldsymbol{A}_k \psi_k \otimes \psi_k, \tag{S.65}$$

where $\{\psi_k, k \geq 1\}$ are orthonormal functions in $\mathbb{L}_2[0,1]$ and $\{\boldsymbol{A}_k, k \geq 1\}$ are a sequence of $q \times q$ covariance matrices. Further, consider $\boldsymbol{A}_k = \nu_k \boldsymbol{R}$, where $\nu_1 \geq \nu_2 \geq \cdots > 0$ are a sequence of eigenvalues and $\boldsymbol{R}$ is a $q \times q$ correlation matrix, e.g. a $MA(1)$ correlation matrix. In this setting, $\{X_j, j \in \mathscr{S}\}$ share the same eigenvalues and eigenfunctions, and their principal component scores have the same correlation structure across different order k. To satisfy Condition C.2, $\nu_1 = 1$ and $\{\nu_k\}$ decay to 0 fast enough such that $\sum_{k \geq 1} \nu_k < \infty$. To verify C.3,

$$\boldsymbol{\mathscr{T}}^{(\mathscr{S},\mathscr{S})}(\boldsymbol{\mathscr{T}}_\lambda^{(\mathscr{S},\mathscr{S})})^{-1} = \sum_{k=1}^{\infty} \boldsymbol{A}_k(\boldsymbol{A}_k + \lambda \boldsymbol{I})^{-1} \psi_k \otimes \psi_k \equiv \sum_{k=1}^{\infty} \boldsymbol{B}_k \psi_k \otimes \psi_k.$$

Under the setting considered, $\boldsymbol{B}_k = \boldsymbol{R}(\boldsymbol{R} + \vartheta_k \boldsymbol{I})^{-1}$, where $\vartheta_k = \lambda/\nu_k \to \infty$ as $k \to \infty$.

S.3.1 MA(1) correlation

We first focus on MA(1) correlation

$$\boldsymbol{R} = \begin{pmatrix} 1 & \rho & 0 & 0 & \cdots & 0 \\ \rho & 1 & \rho & 0 & \cdots & 0 \\ 0 & \rho & 1 & \rho & \cdots & 0 \\ \vdots & 0 & \ddots & \ddots & \ddots & \vdots \\ 0 & \vdots & & \rho & 1 & \rho \\ 0 & 0 & \cdots & 0 & \rho & 1 \end{pmatrix}.$$

In order for $\boldsymbol{R}$ to be a legitimate correlation matrix, we need $|\rho| < 1/2$. We will focus on the case $0 \leq \rho < 1/2$; the same conclusion can be reached for $\rho \in (-1/2, 0)$ using similar arguments. We have

$$\boldsymbol{B}_k = \boldsymbol{I} - \vartheta_k(\boldsymbol{R} + \vartheta_k \boldsymbol{I})^{-1} = \boldsymbol{I} - \frac{\vartheta_k}{1+\vartheta_k}\widetilde{\boldsymbol{R}}_k^{-1}$$

where

$$\widetilde{\boldsymbol{R}}_k = \begin{pmatrix} 1 & \widetilde{\rho}_k & 0 & 0 & \cdots & 0 \\ \widetilde{\rho}_k & 1 & \widetilde{\rho}_k & 0 & \cdots & 0 \\ 0 & \widetilde{\rho}_k & 1 & \widetilde{\rho}_k & \cdots & 0 \\ \vdots & 0 & \ddots & \ddots & \ddots & \vdots \\ 0 & \vdots & & \widetilde{\rho}_k & 1 & \widetilde{\rho}_k \\ 0 & 0 & \cdots & 0 & \widetilde{\rho}_k & 1 \end{pmatrix},$$

with $\widetilde{\rho}_k = \rho/(1+\vartheta_k)$. Note that both $\boldsymbol{B}_k$ and $\widetilde{\boldsymbol{R}}_k^{-1}$ are positive definite, all diagonal values for both matrices should be greater than 0, hence $|B_{k,jj}| < 1$ for all k, j. Let $\widetilde{R}^{jj'}$ be the (j, j')th element of $\widetilde{\boldsymbol{R}}^{-1}$ and denote

$$\theta_k = \frac{1-\sqrt{1-4\widetilde{\rho}_k^2}}{2\widetilde{\rho}_k} = \frac{2\widetilde{\rho}_k}{1+\sqrt{1-4\widetilde{\rho}_k^2}}.$$

One can easily verify that θ_k is an increasing function of $\widetilde{\rho}_k$ and $|\theta_k| < 1$. Hence, θ_k decreases to 0 as $\vartheta_k \to \infty$ with k.

By Shaman (1969),

$$\begin{aligned} |\widetilde{R}_k^{jj'}| &\leq \frac{1}{\sqrt{1-4\widetilde{\rho}_k^2}}\theta_k^{|j-j'|} \\ &\leq \frac{1}{\sqrt{1-4\widetilde{\rho}_1^2}}\theta_1^{|j-j'|} \end{aligned}$$

$$\leq \frac{1}{\sqrt{1-4\rho^2}}\theta^{|j-j'|}, \tag{S.66}$$

where $\theta = \frac{1-\sqrt{1-4\rho^2}}{2\rho} \in [0,1)$. Hence, for $j \neq j'$, $|B_{k,jj'}| \leq |\widetilde{R}_k^{jj'}| \leq \frac{1}{\sqrt{1-4\rho^2}}\theta^{|j-j'|}$ uniformly for all k. By (S.3)

$$\varkappa = \left|\!\left|\!\left| \mathscr{T}^{(\mathscr{S},\mathscr{S})}(\mathscr{T}_\lambda^{(\mathscr{S},\mathscr{S})})^{-1} \right|\!\right|\!\right|_{\infty,\infty} \leq 1 + \frac{1}{\sqrt{1-4\rho^2}}\frac{2\theta}{1-\theta}, \tag{S.67}$$

which is a constant not depending on λ or q. We continue to verify C.4 in this example:

$$(\mathscr{T}^{(\mathscr{S},\mathscr{S})} - \mathscr{Q}^{(\mathscr{S},\mathscr{S})})(\mathscr{Q}_\lambda^{(\mathscr{S},\mathscr{S})})^{-1} = \sum_{k=1}^{\infty} \frac{\nu_k}{\nu_k+\lambda}(\boldsymbol{R}-\boldsymbol{I})\psi_k \otimes \psi_k,$$

where $\boldsymbol{R}$ is the MA(1) correlation matrix above. Using the same argument as for (S.3),

$$\left|\!\left|\!\left| (\mathscr{T}^{(\mathscr{S},\mathscr{S})} - \mathscr{Q}^{(\mathscr{S},\mathscr{S})})(\mathscr{Q}_\lambda^{(\mathscr{S},\mathscr{S})})^{-1} \right|\!\right|\!\right|_{\infty,\infty} = 2\rho \max_k \frac{\nu_k}{\nu_k+\lambda} \leq 2\rho < 1,$$

which satisfies Condition C.4.

S.3.2 AR(1) correlation

We shift our focus towards AR(1) correlation

$$\boldsymbol{R} = \begin{pmatrix} 1 & \rho & \rho^2 & \rho^3 & \cdots & \rho^{q-1} \\ \rho & 1 & \rho & \rho^2 & \cdots & \rho^{q-2} \\ \rho^2 & \rho & 1 & \rho & \cdots & \rho^{q-3} \\ \vdots & \rho^2 & \ddots & \ddots & \ddots & \vdots \\ \rho^{q-2} & \vdots & \ddots & \rho & 1 & \rho \\ \rho^{q-1} & \rho^{q-2} & \cdots & \rho^2 & \rho & 1 \end{pmatrix},$$

and we will focus on the case $0 \leq \rho < 1$. Similarly, because $\boldsymbol{B}_k = \boldsymbol{I} - \vartheta_k(\boldsymbol{R} + \vartheta_k \boldsymbol{I})^{-1}$, we have $|B_{k,jj}| < 1$ for all k, j. Define $\widetilde{\boldsymbol{R}}_k = \boldsymbol{R} + \vartheta_k \boldsymbol{I}$, let $\widetilde{R}_k^{jj'}$ be the (j, j')th element of $\widetilde{\boldsymbol{R}}_k^{-1}$, we have $|B_{k,jj'}| \leq \vartheta_k |\widetilde{R}_k^{jj'}|$ for all $j' \neq j$.

Consider stochastic process Y_t with AR(1) mean and Gaussian white noise, i.e.

$$\begin{cases} Y_t = \mu_t + W_t, & W_t \overset{i.i.d}{\sim} \mathcal{N}(0, \vartheta) \\ \mu_t = \rho\mu_{t-1} + V_t, & V_t \overset{i.i.d}{\sim} \mathcal{N}(0, 1 - \rho^2) \end{cases}$$

then $\boldsymbol{Y}_{[1:q]} \sim \mathcal{N}(\boldsymbol{0}, \widetilde{\boldsymbol{R}})$, where $\widetilde{\boldsymbol{R}} = \boldsymbol{R} + \vartheta \boldsymbol{I}$. It can be shown that Y_t is an

ARMA(1,1) process

$$Y_t = \rho Y_{t-1} + U_t - \theta U_{t-1}, \quad U_t \overset{i.i.d}{\sim} \mathscr{N}(0, \kappa),$$

where $0 \le \theta < 1$, and (θ, κ) satisfies

$$\begin{cases} \mathrm{Var}(Y_t) = 1 + \vartheta = \dfrac{1 - 2\rho\theta + \theta^2}{1 - \rho^2}\kappa \\ \mathrm{Cov}(Y_t, Y_{t-h}) = \rho^{|h|} = \dfrac{(\rho - \theta)(1 - \rho\theta)}{1 - \rho^2}\rho^{|h|-1}\kappa, \end{cases}$$

then

$$\theta \le \rho, \quad \frac{\rho}{1 + \vartheta} = \frac{(\rho - \theta)(1 - \rho\theta)}{1 - 2\rho\theta + \theta^2}, \quad \frac{\vartheta}{\kappa} = \frac{\theta}{\rho}.$$

According to Tiao and Ali (1971), for $j' \ne j$, we have

$$\kappa|\widetilde{R}^{jj'}| \le C\left\{(1 - \rho\theta)^2\theta^{|j-j'|-1} + (\rho - \theta)^2\theta^{2q-|j-j'|-1} + (1 - \rho\theta)(\rho - \theta)\left(\theta^{j+j'-2} + \theta^{2q-j-j'}\right)\right\},$$

where

$$\begin{aligned} C &= \left\{1 + \frac{(\rho - \theta)^2(1 - \theta^{2q})}{(1 - \rho^2)(1 - \theta^2)}\right\}^{-1} \frac{(\rho - \theta)(1 - \rho\theta)}{(1 - \rho^2)(1 - \theta^2)^2} \\ &\le \frac{1}{(1 - \theta^2)^2}\frac{(\rho - \theta)(1 - \rho\theta)}{1 - 2\rho\theta + \theta^2} \end{aligned}$$

$$
\begin{aligned}
&= \frac{\rho}{(1+\vartheta)(1-\theta^2)^2} \\
&\le \frac{\rho}{(1-\theta^2)^2}.
\end{aligned}
$$

Also, note that

$$
\frac{\vartheta}{\kappa}C \le \frac{\theta}{(1-\theta^2)^2}; \quad 1-\rho\theta \le 1-\theta^2; \quad \rho-\theta \le 1-\theta \le 1-\theta^2,
$$

we have

$$
\begin{aligned}
|B_{k,jj'}| &\le \theta_k^{|j-j'|} + \theta_k^{2q-|j-j'|} + \theta_k^{j+j'-1} + \theta_k^{2q-j-j'+1} \\
&\le \rho^{|j-j'|} + \rho^{2q-|j-j'|} + \rho^{j+j'-1} + \rho^{2q-j-j'+1}.
\end{aligned}
$$

Applying some algebra, we have

$$
\begin{aligned}
&\max_j \sum_{j \ne j'} \rho^{|j-j'|} \le \frac{2\rho}{1-\rho}(1-\rho^{q-1}), \quad \max_j \sum_{j \ne j'} \rho^{2q-|j-j'|} = \sum_{k=q+1}^{2q-1} \rho^k \le \frac{\rho^{q+1}}{1-\rho}, \\
&\max_j \sum_{j' \ne j} \rho^{j+j'-1} + \rho^{2q-j-j'+1} \le \max_j \left(\rho^{j-1} + \rho^{q-j}\right) \sum_{k=1}^{q} \rho^k \le \frac{\rho}{1-\rho}(1+\rho^{q-1}).
\end{aligned}
$$

By (S.3) and the above derivation,

$$
\varkappa = \left|\!\left|\!\left| \mathscr{T}^{(\mathscr{S},\mathscr{S})} (\mathscr{T}_\lambda^{(\mathscr{S},\mathscr{S})})^{-1} \right|\!\right|\!\right|_{\infty,\infty} \le 1 + \frac{3\rho}{1-\rho} \tag{S.68}
$$

which is a constant not depending on λ or q. We continue to verify C. 4. Using the same argument as for (S.3),

$$
\begin{aligned}
\left\|\left\|\sum_{k=1}^{\infty}\frac{\nu_k}{\nu_k+\lambda}(\boldsymbol{R}-\boldsymbol{I})\psi_k\otimes\psi_k\right\|\right\|_{\infty,\infty} &\le \max_{1\le j\le q}\sum_{j'\neq j}\max_k\frac{\nu_k}{\nu_k+\lambda}\rho^{|j-j'|}\\
&\le \max_{1\le j\le q}\sum_{j'\neq j}\rho^{|j-j'|}\\
&= \frac{\rho}{1-\rho}\left(2-\rho^{\lceil (q-1)/2\rceil}-\rho^{\lfloor (q-1)/2\rfloor}\right)\\
&\le \frac{2\rho}{1-\rho}.
\end{aligned}
$$

Hence, for large q, we need $\rho \le 1/3$ in order that C. 4 holds.

S.4. Additional Simulation Results

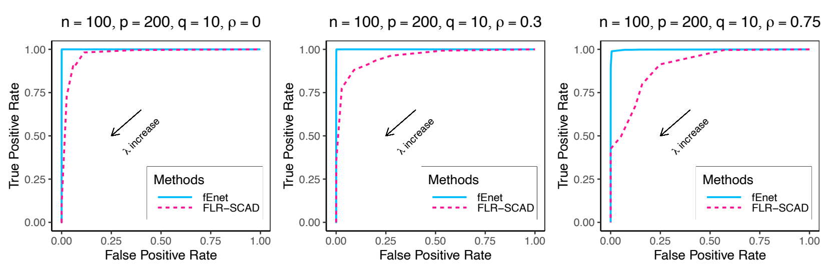

Figure S.1: Simulation Scenario Ⅱ: the ROC curves of fEnet and FLR-SCAD under the ultra high-dimensional case. The ROC curves are obtained by changing the value of λ and holding other hyperparameters as optimal.

Table S.1: Simulation Scenario II: summary of estimation, prediction, and variable selection performance of the proposed fEnet versus FLR-SCAD under different problem sizes.

n	p	q	Method	FPR (%)	FNR (%)	MND	RER
					$\rho = 0$		
500	50	5	fEnet	0 (0, 0)	0 (0, 0)	1.11 (0.61, 1.82)	0.0009 (0.0005, 0.0015)
			FLR-SCAD	0 (0, 0)	0 (0, 0)	1.80 (0.90, 3.59)	0.0014 (0.0008, 0.0028)
200	100	5	fEnet	0 (0, 0)	0 (0, 0)	1.57 (0.81, 2.37)	0.0025 (0.0015, 0.0040)
			FLR-SCAD	0 (0, 0)	0 (0, 0)	2.16 (1.18, 3.71)	0.0048 (0.0025, 0.0111)
100	200	10	fEnet	0 (0, 0.5)	0 (0, 0)	3.23 (2.01, 5.05)	0.0252 (0.0124, 0.0611)
			FLR-SCAD	5.8 (1.1, 13.2)	10 (0, 30)	7.49 (4.90, 15.18)	0.4896 (0.2332, 0.8809)
					$\rho = 0.3$		
500	50	5	fEnet	0 (0, 0)	0 (0, 0)	1.11 (0.68, 2.05)	0.0011 (0.0007, 0.0017)
			FLR-SCAD	0 (0, 0)	0 (0, 0)	1.96 (0.93, 4.11)	0.0016 (0.0009, 0.0033)
200	100	5	fEnet	0 (0, 0)	0 (0, 0)	1.66 (0.90, 2.52)	0.0028 (0.0016, 0.0049)
			FLR-SCAD	0 (0, 0)	0 (0, 0)	2.18 (1.03, 3.60)	0.0054 (0.0025, 0.0132)
100	200	10	fEnet	0 (0, 1.1)	0 (0, 0)	3.15 (1.95, 4.97)	0.0230 (0.0110, 0.0735)
			FLR-SCAD	8.4 (4.2, 14.2)	10 (0, 30)	7.60 (4.95, 12.37)	0.4162 (0.2522, 0.7676)
					$\rho = 0.75$		
500	50	5	fEnet	0 (0, 0)	0 (0, 0)	1.61 (0.82, 2.63)	0.0013 (0.0008, 0.0021)
			FLR-SCAD	0 (0, 0)	0 (0, 0)	3.08 (1.38, 6.41)	0.0018 (0.0010, 0.0040)
200	100	5	fEnet	0 (0, 0)	0 (0, 0)	1.95 (0.99, 3.25)	0.0032 (0.0018, 0.0055)
			FLR-SCAD	0 (0, 2.1)	0 (0, 0)	2.93 (1.41, 6.34)	0.0060 (0.0030, 0.0140)
100	200	10	fEnet	0 (0, 3.7)	0 (0, 10)	4.15 (2.73, 6.55)	0.0184 (0.0084, 0.0914)
			FLR-SCAD	4.7 (1.6, 10.6)	50 (30, 70)	8.16 (4.95, 16.04)	0.2345 (0.1581, 0.3791)

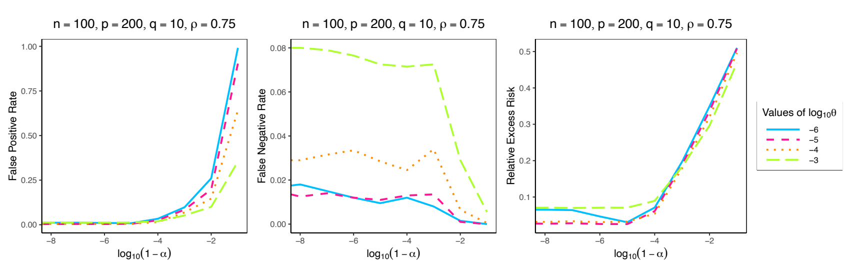

Figure S.2: Simulation Scenario II: the plots of FPR, FNR, and RER versus $\log_{10}(1-\alpha)$ for different values of θ under the ultra high-dimensional case.

Table S.2: Simulation Scenario III: summary of estimation, prediction, and variable selection performance of the proposed fEnet method versus FLR-SCAD under different problem sizes.

n	p	q	Method	FPR (%)	FNR (%)	MND	RER
					$\rho = 0$		
500	50	5	fEnet	0 (0, 0)	0 (0, 0)	0.55 (0.41, 0.82)	0.0213 (0.0113, 0.0340)
			FLR-SCAD	0 (0, 0)	0 (0, 0)	0.65 (0.46, 1.09)	0.0381 (0.0245, 0.0604)
200	100	5	fEnet	0 (0, 0)	0 (0, 0)	0.86 (0.57, 1.39)	0.0413 (0.0248, 0.0702)
			FLR-SCAD	9.5 (4.2, 17.9)	0 (0, 0)	0.92 (0.65, 1.37)	0.0612 (0.0405, 0.1034)
100	200	10	fEnet	0 (0, 0.5)	0 (0, 10)	1.49 (0.95, 4.18)	0.0784 (0.0429, 0.2346)
			FLR-SCAD	6.8 (2.6, 11.6)	0 (0, 30)	4.01 (2.86, 4.18)	0.4616 (0.2127, 0.7290)
					$\rho = 0.3$		
500	50	5	fEnet	0 (0, 0)	0 (0, 0)	0.58 (0.40, 0.89)	0.0274 (0.0172, 0.0491)
			FLR-SCAD	0 (0, 2.2)	0 (0, 0)	0.65 (0.48, 0.87)	0.0528 (0.0353, 0.0830)
200	100	5	fEnet	0 (0, 0)	0 (0, 0)	0.95 (0.61, 1.39)	0.0562 (0.0338, 0.1042)
			FLR-SCAD	9.5 (4.2, 15.8)	0 (0, 0)	0.96 (0.66, 1.41)	0.0797 (0.0503, 0.1410)
100	200	10	fEnet	0 (0, 1.1)	0 (0, 20)	1.84 (1.32, 4.18)	0.1048 (0.0618, 0.3288)
			FLR-SCAD	8.4 (3.7, 13.2)	20 (0, 50)	4.18 (3.88, 4.18)	0.5074 (0.3487, 0.7764)
					$\rho = 0.75$		
500	50	5	fEnet	2.2 (0, 6.7)	0 (0, 0)	0.86 (0.62, 1.42)	0.0504 (0.0276, 0.0926)
			FLR-SCAD	26.7 (13.3, 37.8)	0 (0, 0)	1.05 (0.73, 3.59)	0.0870 (0.0506, 0.1701)
200	100	5	fEnet	1.1 (0, 4.2)	0 (0, 20)	1.45 (0.90, 4.18)	0.1411 (0.0603, 0.3734)
			FLR-SCAD	9.5 (3.2, 16.8)	20 (0, 40)	4.18 (1.29, 4.18)	0.3056 (0.1227, 0.5523)
100	200	10	fEnet	0.5 (0, 1.6)	40 (20, 50)	4.18 (4.18, 4.18)	0.1518 (0.0878, 0.2769)
			FLR-SCAD	5.3 (2.1, 9.0)	60 (40, 70)	4.19 (4.18, 6.16)	0.2467 (0.1616, 0.3688)

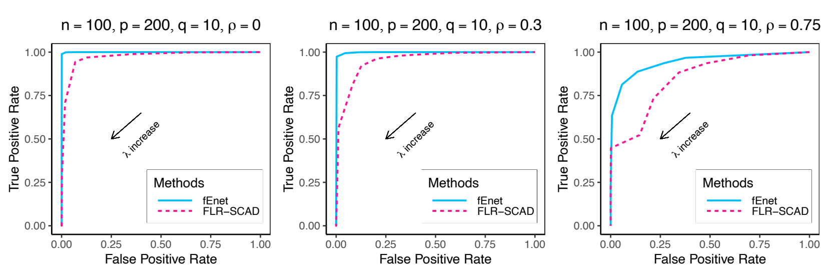

Figure S.3: Simulation Scenario III: the ROC curves of fEnet and FLR-SCAD under the ultra high-dimensional case. The ROC curves are obtained by changing the value of λ and holding other hyperparameters as optimal.

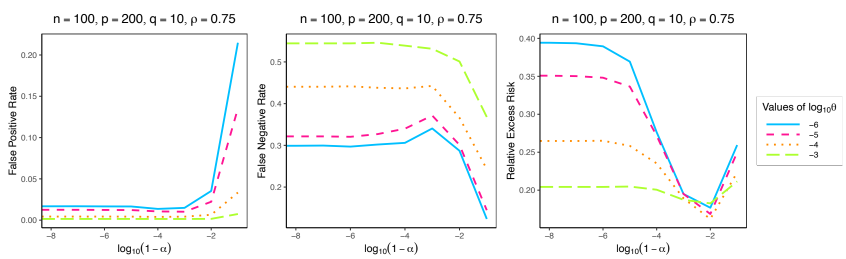

Figure S.4: Simulation Scenario III: the plots of FPR, FNR, and RER versus $\log_{10}(1-\alpha)$ for different values of θ under the ultra high-dimensional case.

Supplementary References

Hsing, T. and Eubank, R. (2015). *Theoretical foundations of functional data analysis, with an introduction to linear operators*, volume 997. John Wiley & Sons.

Koltchinskii, V. and Lounici, K. (2017). Concentration inequalities and moment bounds for sample covariance operators. *Bernoulli*, 23(1):110–133.

Laurent, B. and Massart, P. (2000). Adaptive estimation of a quadratic functional by model selection. *The Annals of Statistics*, 58(5):1302–1338.

Lu, L.-Z. and Pearce, C. E. (2000). Some new bounds for singular values and eigenvalues of matrix products. *Annals of Operations Research*, 98(1):141–148.

Shaman, P. (1969). On the inverse of the covariance matrix of a first order moving average. *Biometrika*, 56(3):595–600.

Tiao, G. and Ali, M. M. (1971). Analysis of correlated random effects: Linear model with two random components. *Biometrika*, 58(1):37–51.

Tsybakov, A. B. (2009). *Introduction to Nonparametric Estimation*. Springer, New York.

Zapata, J., Oh, S. Y., and Petersen, A. (2021). Partial separability and functional graphical models for multivariate Gaussian processes. *Biometrika*, 109(3):665–681.