# Supplementary Material for Network Reconstruction From High Dimensional Ordinary Differential Equations

Shizhe Chen, Ali Shojaie, and Daniela M. Witten

Department of Biostatistics
University of Washington
Box 357232
Seattle, WA 98195-7232

# A. PROOFS

## A.1 Outline

In this section, we prove Theorems 1 and 2 from Section 4 in the main paper. The remaining subsections are organized as follows. In Section A.2, we list the additional assumptions for Theorem 1 in the main paper and give the proof of Theorem 1 in the main paper. In Section A.3, we prove a theorem on variable selection consistency for group lasso regression with errors in variables, which itself is of independent interest. In Section A.4, we introduce Assumption S4 on the bases $\psi(\cdot)$, and several technical lemmas that are useful in proving Theorem 2 in the main paper. In Section A.5, we finish the proof of Theorem 2 in the main paper. And in Section A.6, we prove Proposition 1 in the main paper. The proofs of the technical lemmas presented in Section A.4 are provided in Section B.

## A.2 Proof of Theorem 1

In this section, we follow closely the notation in Section 1.6 of Tsybakov (2009). We first present some necessary notation and assumptions. Denote the local polynomial estimator of degree $\ell$ as

$$\hat{X}(t;h) = \sum_{i=1}^{n} Y_i W_{ni}(t;h), \tag{S1}$$

where

$$W_{ni}(t;h) = \frac{1}{nh} U^{\mathrm{T}}(0) B_{nt}^{-1} U\left(\frac{t_i - t}{h}\right) K\left(\frac{t_i - t}{h}\right), \tag{S2}$$

$$B_{nt} = \frac{1}{nh} \sum_{i=1}^{n} U\left(\frac{t_i - t}{h}\right) U^{\mathrm{T}}\left(\frac{t_i - t}{h}\right) K\left(\frac{t_i - t}{h}\right),$$

$$U(u) = \left(1, u, u^2/2!, \ldots, u^\ell/\ell!\right)^{\mathrm{T}},$$

and $K(\cdot)$ is a kernel function. In (S2), $W_{ni}(t;h)$ is the weight for observation $Y_i$ in (S1), which satisfies

$$\sum_{i=1}^{n} W_{ni}(t;h) = 1. \tag{S3}$$

See e.g., Proposition 1.12 in Tsybakov (2009), for a rigorous proof of (S3). We introduce the following assumptions on the kernel function $K(\cdot)$ and the time points $t_1, \ldots, t_n$. These assumptions are common in the study of local polynomial estimators (see e.g. Tsybakov, 2009).

**Assumption S1.** There exists a real number $\lambda_0 > 0$ and a positive integer $n_0$ such that the smallest eigenvalue $\Lambda_{\min}(B_{nt})$ of $B_{nt}$ satisfies

$$\Lambda_{\min}(B_{nt}) \geq \lambda_0$$

for all $n \geq n_0$ and any $t \in [0,1]$.

**Assumption S2.** The time points $t_1, \ldots, t_n$ are evenly-spaced on the interval $[0,1]$.

**Assumption S3.** The kernel $K$ has compact support belonging to $[-1,1]$, and there exists a number $K_{\max} < \infty$ such that $|K(u)| \leq K_{\max}$, $\forall u \in \mathbb{R}$.

These assumptions lead to the following lemma (Lemma 1.3 in Tsybakov, 2009).

**Lemma S1.** *Under Assumptions S1–S3, for all $n \geq n_0$, $h \geq 1/(2n)$, and $t \in [0,1]$, the weights $W_{ni}$ in* (S2) *satisfy:*

*i.* $\sup_{i,t} |W_{ni}(t;h)| \le C_3/nh$;

*ii.* $\sum_{i=1}^n |W_{ni}(t;h)| \le C_3$,

*where the constant $C_3$ depends only on $\lambda_0$ and $K_{\max}$.*

Recall that we also assume the unknown true solutions $X_j^*$, $j = 1, \ldots, p$, belong to a Hölder class in Assumption 2 in the main paper. We state the definition here for completeness.

**Definition S1.** Let $T$ be an interval in $\mathbb{R}$ and let $\beta_1$ and $L_1$ be two positive numbers. The Hölder class $\Sigma(\beta_1, L_1)$ on $T$ is defined as the set of $\ell = \lfloor \beta_1 \rfloor$ times differentiable functions $f : T \to \mathbb{R}$ whose $\ell$th order derivative $f^{(\ell)}(\cdot)$ satisfies

$$|f^{(\ell)}(x) - f^{(\ell)}(x')| \le L_1 |x - x'|^{\beta_1 - \ell}, \quad \forall x, x' \in T.$$

We are now ready to prove Theorem 1 of the main paper.

*Proof.*

$$\begin{aligned}\left|\left|\left|\hat{X}_j - X_j^*\right|\right|\right|^2 &= \int_0^1 \{\hat{X}_j(u;h) - X_j^*(u)\}^2 \, du = \int_0^1 \left\{\sum_{i=1}^n Y_{ij} W_{ni}(u;h) - X_j^*(u)\right\}^2 du \\ &= \int_0^1 \left[\sum_{i=1}^n \{X_j^*(t_i) + \epsilon_{ji}\} W_{ni}(u;h) - X_j^*(u)\right]^2 du.\end{aligned}$$

Using the property (S3) of the weights $W_{ni}$ and the fact that $(a+b)^2 \le 2a^2 + 2b^2$,

$$\begin{aligned} \left\| \hat{X}_j - X_j^* \right\|^2 \le\, & 2\int_0^1 \left[ \sum_{i=1}^n \{X_j^*(t_i) - X_j^*(u)\} W_{ni}(u;h) \right]^2 du \\ & + 2\int_0^1 \left\{ \sum_{i=1}^n \epsilon_{ji} W_{ni}(u;h) \right\}^2 du \\ \equiv\, & 2\int_0^1 \mathrm{bias}^2(u)\, du + 2\int_0^1 g^2(\epsilon_j/\sigma, u, h)\, du, \end{aligned} \tag{S4}$$

where

$$\mathrm{bias}(u) = \sum_{i=1}^n \{X_j^*(t_i) - X_j^*(t)\} W_{ni}(u;h), \tag{S5}$$

$$g(a,u,h) = \sigma \sum_{i=1}^n a_i W_{ni}(u;h), \quad \epsilon_j = (\epsilon_{1j}, \ldots, \epsilon_{nj})^{\mathrm{T}}, \tag{S6}$$

and where $\sigma$ is defined in Assumption 1 in the main paper.

In what follows, for convenience, we denote the $\ell$th derivative of $X_j^*(t)$ as $X_j^{(\ell)}$. Under Assumption 2 in the main paper and Assumptions S1–S3, it follows from Proposition 1.13 in Tsybakov (2009) that $|\mathrm{bias}(u)| \le q_1 h^{\beta_1}$, where $q_1 = C_3 L_1/\ell!$. Therefore,

$$\int_0^1 \mathrm{bias}^2(u)\, du \le q_1^2 h^{2\beta_1}. \tag{S7}$$

Next, we bound $g(\epsilon_j/\sigma, t, h)$ in (S6) using Theorem 5.6 in Boucheron et al. (2013). The theorem states that if $Z = (Z_1, \ldots, Z_n)$ is a vector of $n$ independent standard normal random variables and $f$ is an $L$-Lipschitz function, then for all $v > 0$,

$$\Pr\{f(Z) - \mathbb{E}f(Z) \ge v\} \le \exp\{-v^2/(2L^2)\}.$$

Applying the theorem to $f(z)$ and $-f(z)$, we get

$$\Pr\{|f(Z) - \mathbb{E}f(Z)| \geq v\} \leq 2\exp\{-v^2/(2L^2)\}.$$

We now show that $g(x, t, h)$ is an $L_3$-Lipschitz function with $L_3 = \sigma C_3(nh)^{-0.5}$:

$$\begin{aligned}
|g(a, u, h) - g(b, u, h)| =& \sigma \left| \sum_{i=1}^{n} (a_i - b_i) W_{ni}(u; h) \right| \\
\leq& \sigma \left\{ \sum_{i=1}^{n} W_{ni}^2(u; h) \right\}^{\frac{1}{2}} \|a - b\|_2 \\
\leq& \sigma \left\{ \sup_{i,u} |W_{ni}(u; h)| \sum_{i=1}^{n} |W_{ni}(u, h)| \right\}^{\frac{1}{2}} \|a - b\|_2 \\
\leq& \sigma C_3 \sqrt{\frac{1}{nh}} \|a - b\|_2,
\end{aligned}$$

where the last inequality follows from Lemma S1. Hence, from Theorem 5.6 in Boucheron et al. (2013), we have

$$\Pr\{|g(\epsilon_j/\sigma, u, h) - \mathbb{E}g(\epsilon_j/\sigma, u, h)| \geq v\} \leq 2\exp\{-v^2/(2L_3^2)\}.$$

Letting $v = n^{\alpha/2-0.5} h^{-0.5}$ and noting that $\mathbb{E}[g(\epsilon_j/\sigma, u, h)] = 0$, we have

$$\Pr\{|g(\epsilon_j/\sigma, u, h)| \geq n^{\alpha/2-0.5} h^{-0.5}\} \leq 2\exp\{-n^\alpha/(2\sigma^2 C_3^2)\}. \tag{S8}$$

Combining (S4), (S7), and (S8), we have

$$\left|\left|\left|\hat{X}_j - X_j^*\right|\right|\right|^2 \le 2\int_0^1 \mathrm{bias}^2(u)\,du + 2\int_0^1 g^2(\epsilon_j/\sigma, u, h)\,du \\ \le 2q_1^2 h^{2\beta_1} + 2n^{\alpha-1}h^{-1}, \tag{S9}$$

with probability at least $1 - 2\exp\{-n^\alpha/(2\sigma^2 C_3^2)\}$.

Minimizing the right-hand side of (S9) with respect to $h$, we find that the minimizer $h_n$ satisfies

$$2\beta_1 q_1^2 h_n^{2\beta_1+1} = n^{\alpha-1}.$$

Thus, for $h_n \propto n^{(\alpha-1)/(2\beta_1+1)}$, the error bound is

$$\left|\left|\left|\hat{X}_j - X_j^*\right|\right|\right|^2 \le C_2 n^{\frac{2\beta_1}{2\beta_1+1}(\alpha-1)},$$

for some global constant $C_2$. $\square$

## A.3 Variable selection consistency of group lasso in error-in-variable models

We first review some notation that is heavily used in this section. In (17c) of the main paper, we made use of the notation

$$\hat{\Psi}_0(t) = t;\ \hat{\Psi}_k(t) = \int_0^t \psi(\hat{X}_k(u;h))\,du,\ k = 1, \ldots, p.$$

Therefore, $\hat{\Psi}_k(t)$ is an $M$-vector for $k = 1, \ldots, p$ and a scalar for $k = 0$. We sometimes use sets, e.g. $S_j$ and $S_j^0$, as the subscripts. In this case, $\hat{\Psi}_{S_j}(t)$ is an $Ms_j$-vector, which is composed of $\hat{\Psi}_k$

for $k \in S_j$. Furthermore, $\hat{\Psi}_{S_j^0} = (\hat{\Psi}_0(t), \hat{\Psi}_{S_j}^{\mathrm{T}}(t))^{\mathrm{T}}$ is an $(Ms_j + 1)$-vector. Without subscripts, $\hat{\Psi}(t) \equiv (\hat{\Psi}_0(t), \hat{\Psi}_1^{\mathrm{T}}(t), \ldots, \hat{\Psi}_p^{\mathrm{T}}(t))^{\mathrm{T}}$ is of dimension $Mp + 1$. We will also apply subscripts to the quantities $\theta_j^*$, $\hat{\theta}_j$, $\hat{g}$, and $R$. For instance, $\hat{\theta}_{jk} = (\theta_{jk1}, \ldots, \theta_{jkM})^{\mathrm{T}}$ for $k = 1, \ldots, p$, and $\hat{\theta}_j = (\hat{\theta}_{j0}, \hat{\theta}_{j1}^{\mathrm{T}}, \ldots, \hat{\theta}_{jp}^{\mathrm{T}})^{\mathrm{T}}$. The products of these vectors are defined as usual, e.g., $\hat{\theta}_{jS_j^0}^{\mathrm{T}} \hat{\Psi}_{S_j^0}(t)$ is a scalar, and $\hat{\Psi}_{S_j^0}(t)\hat{\Psi}_{S_j^0}^{\mathrm{T}}(t)$ is an $(Ms_j + 1) \times (Ms_j + 1)$ matrix.

The optimization problem (17a) in the main paper is a standardized group lasso problem (Simon and Tibshirani, 2012). Because the regressors $\hat{\Psi}_1, \ldots, \hat{\Psi}_p$ are estimated, establishing variable selection consistency requires extra attention. For ease of discussion, we re-state the optimization problem (17a),

$$\hat{\theta}_j = \underset{C_0 \in \mathbb{R}, \theta_{j0} \in \mathbb{R},\ \theta_{jk} \in \mathbb{R}^M}{\arg\min} \frac{1}{2n} \sum_{i=1}^{n} \left\{ Y_{ij} - C_0 - \theta_{j0}\hat{\Psi}_0(t_i) - \sum_{k=1}^{p} \theta_{jk}^{\mathrm{T}} \hat{\Psi}_k(t_i) \right\}^2 + \lambda_{n,j} \sum_{k=1}^{p} \left[ \frac{1}{n} \sum_{i=1}^{n} \{\theta_{jk}^{\mathrm{T}} \hat{\Psi}_k(t_i)\}^2 \right]^{1/2},$$

where

$$\hat{X}(\cdot; h) = \underset{Z(\cdot) \in \mathcal{X}(h)}{\arg\min} \sum_{i=1}^{n} \|Y_i - Z(t_i)\|_2^2,$$

$$\hat{\Psi}_0(t) = t;\ \hat{\Psi}_k(t) = \int_0^t \psi(\hat{X}_k(u; h))\, du,\ k = 1, \ldots, p.$$

In what follows, for simplicity we assume that $X_j^*(0) = 0$, and that $\lambda_{n,1} = \cdots = \lambda_{n,p} \equiv \lambda_n$. For any $1 \le j, k \le p$, let $\theta_{jk}^* \in \mathbb{R}^M$ be the coefficients of the true functions $f_{jk}^*$ on the bases $\psi(\cdot)$, i.e.,

$$f_{jk}^*(a) = \psi(a)^{\mathrm{T}} \theta_{jk}^* + \delta_{jk}(a), \tag{S10}$$

where $f_{jk}^*$ is introduced in Assumption 3 in the main paper. Here we establish variable selection consistency for group lasso regression with errors in variables. We extend the recent work of Loh and Wainwright (2012) for lasso regression; related results can be found in Ma and Li (2010) and Rosenbaum and Tsybakov (2010). In order for variable selection consistency to hold, we need four conditions. In Section A.5, we will show that these conditions hold with high probability given Assumptions 1–6 in the main paper and Assumptions S1–S4.

**Condition S1.** Suppose that

$$
\begin{aligned}
&0 < \frac{1}{2} C_{\min} \leq \Lambda_{\min} \left( \frac{1}{n} \sum_{i=1}^{n} \hat{\Psi}_{S_j^0}(t_i) \hat{\Psi}_{S_j^0}^{\mathrm{T}}(t_i) \right), \\
&\Lambda_{\max} \left( \frac{1}{n} \sum_{i=1}^{n} \hat{\Psi}_{S_j^0}(t_i) \hat{\Psi}_{S_j^0}^{\mathrm{T}}(t_i) \right) \leq 2C_{\max}, \\
&0 < \frac{1}{2} C_{\min} \leq \Lambda_{\min} \left( \frac{1}{n} \sum_{i=1}^{n} \hat{\Psi}_{k}(t_i) \hat{\Psi}_{k}^{\mathrm{T}}(t_i) \right), \quad k \notin S_j^0,
\end{aligned}
$$

where $C_{\min}$ and $C_{\max}$ are introduced in Assumption 4 in the main paper.

**Condition S2.** Assume that

$$
\max_{k \notin S_j^0} \left\| \left( \frac{1}{n} \sum_{i=1}^{n} \hat{\Psi}_{k}(t_i) \hat{\Psi}_{S_j^0}^{\mathrm{T}}(t_i) \right) \left( \frac{1}{n} \sum_{i=1}^{n} \hat{\Psi}_{S_j^0}(t_i) \hat{\Psi}_{S_j^0}^{\mathrm{T}}(t_i) \right)^{-1} \right\|_2 \leq 2\xi,
$$

where $\xi$ is introduced in Assumption 5.

The next condition was first proposed in Loh and Wainwright (2012) as the deviation condition. Specifically, (S11) is a special case of Equation 3.1 in Loh and Wainwright (2012). Recall that the true parameters $\theta_{j0}^*$ and $\theta_{jk}^*$ are introduced in Assumption 3 of the main paper and (S10), respectively.

**Condition S3.** For $j = 1, \ldots, p$, let $\Delta \equiv \max_{j=1,\ldots,p} \left|\left|\left| \hat{X}_j - X_j^* \right|\right|\right|$. Assume that

$$\left\| \frac{1}{n} \sum_{i=1}^{n} \hat{\Psi}_k(t_i) Y_{ij} - \frac{1}{n} \sum_{i=1}^{n} \hat{\Psi}_k(t_i) \hat{\Psi}_{S_j^0}^{\mathrm{T}}(t_i) \theta_{jS_j^0}^* \right\|_2 \leq \eta, \ k = 0, \ldots, p \tag{S11}$$

where $\eta = M^{1/2} \left\{ sM^{-\beta_2} Q^{1/2} B + BD \|\theta_S^*\|_1 \Delta + n^{\alpha/2-1/2} \right\}$.

Note that the global constant $Q$ in Condition S3 also appears in Assumption S4 in Section A.4.

Condition S4 places constraints on the quantities involved in the proof of Theorem S1. In the proof of Theorem 2 in the main paper, we will show that Condition S4 holds with high probability.

**Condition S4.** The following inequalities hold:

$$\frac{2\sqrt{s+1}}{C_{\min}} \eta + \lambda_n \frac{\sqrt{8sC_{\max}}}{C_{\min}} \leq \frac{2}{3} \theta_{\min},$$

$$\frac{2\xi\sqrt{s+1}+1}{\lambda_n} \eta + 2\xi\sqrt{s}\sqrt{2C_{\max}} < \sqrt{C_{\min}/2},$$

where $\theta_{\min} \equiv \min_{k \in S_j^0} \|\theta_{jk}^*\|_2$, and $\xi, \eta, C_{\min}$, and $C_{\max}$ are introduced in Assumptions 4–6 of the main paper.

We arrive at the following theorem.

**Theorem S1.** *Suppose that Conditions S1–S4 are met. Then the estimator $\hat{\theta}_j$ from* (17a) *has the correct support, i.e. $\hat{S}_j = S_j$ for all $j = 1, \ldots, p$.*

*Proof.* We establish variable selection consistency using the primal-dual witness method (Wainwright, 2009). For simplicity, we drop the subscript $j$ in what follows: for instance, we drop the subscript $j$ in $Y_{ij}$ and $\hat{\theta}_j$ in (17a), and in the estimated neighbourhood $\hat{S}_j$.

A vector $\hat{\theta}$ solves the optimization problem (17a) in the main paper if it satisfies the Karush-Kuhn-Tucker (KKT) condition, which is

$$\frac{1}{n}\sum_{i=1}^{n}\hat{\Psi}_k(t_i)\left\{\hat{\Psi}^{\mathrm{T}}(t_i)\hat{\theta}-Y_i\right\}+\lambda_n\hat{g}_k=0, \quad k=1,\ldots,p, \tag{S12}$$

with

$$\begin{aligned}\hat{g}_k &= \frac{\sum_{i=1}^{n}\hat{\Psi}_k(t_i)\hat{\Psi}_k^{\mathrm{T}}(t_i)\hat{\theta}_k/n}{\sqrt{\hat{\theta}_k^{\mathrm{T}}\sum_{i=1}^{n}\hat{\Psi}_k(t_i)\hat{\Psi}_k^{\mathrm{T}}(t_i)\hat{\theta}_k/n}} \quad \text{if } \hat{\theta}_k\neq 0,\\ \hat{g}_k^{\mathrm{T}}&\left(\frac{1}{n}\sum_{i=1}^{n}\hat{\Psi}_k(t_i)\hat{\Psi}_k^{\mathrm{T}}(t_i)\right)^{-1}\hat{g}_k<1 \quad \text{if } \hat{\theta}_k=0.\end{aligned} \tag{S13}$$

The KKT condition for $\hat{\theta}_0$ is

$$\frac{1}{n}\sum_{i=1}^{n}\hat{\Psi}_0(t_i)\left\{\hat{\Psi}^{\mathrm{T}}(t_i)\hat{\theta}-Y_i\right\}=0. \tag{S14}$$

Note that, in the previous equations, we drop the parameter $C_0$ that appears in (17a) of the main paper to avoid cumbersome bookkeeping.

We will construct an oracle estimator $\hat{\theta}$ and will verify that it satisfies the KKT conditions (S12), (S13), and (S14), which means that it solves the optimization problem (17a) in the main paper.

We construct an oracle primal-dual pair $(\hat{\theta},\hat{g})$ as follows:

1. Set $\hat{\theta}_k=0$ for $k\notin S^0$.

2. Let

$$\hat{\theta}_{S^0}=\underset{\theta_{S^0}\in\mathbb{R}^{sM+1}}{\arg\min}\frac{1}{2n}\sum_{i=1}^{n}\left\{Y_i-\theta_{S^0}^{\mathrm{T}}\hat{\Psi}_{S^0}(t_i)\right\}^2+\lambda_n\sum_{k\in S}\left[\frac{1}{n}\sum_{i=1}^{n}\{\theta_{jk}^{\mathrm{T}}\hat{\Psi}_k(t_i)\}^2\right]^{1/2}. \tag{S15}$$

3. Define $\hat{g}_{S^0} = (0, \hat{g}_S^{\mathrm{T}})^{\mathrm{T}}$ as in (S13).

4. Solve $\hat{g}_k$ from the sub-gradient condition (S12) for $k \notin S^0$.

We will verify the support recovery consistency

$$\max_{k \in S} \|\hat{\theta}_k - \theta_k^*\|_2 \leq \frac{2}{3}\theta_{\min} \tag{S16}$$

and strict dual feasibility

$$\max_{k \notin S^0} \hat{g}_k^{\mathrm{T}} \left( \frac{1}{n} \sum_{i=1}^n \hat{\Psi}_k(t_i) \hat{\Psi}_k^{\mathrm{T}}(t_i) \right)^{-1} \hat{g}_k < 1. \tag{S17}$$

(S16) implies that the oracle estimator $\hat{\theta}$ recovers the support of $\theta^*$ exactly, and (S17) implies that $\hat{\theta}$ solves (17a).

Further, if the optimal solution to (17a) is unique, then the oracle estimator is the unique estimator. If the optimal solution is not unique, then from Theorem 2 in Roth and Fischer (2008), the null set of any optimal solution should contain $S^c$, and thus any optimal solution satisfies the construction of the oracle estimator. Therefore, the statement of Theorem S1 holds for any optimal solution for (17a).

We now establish (S16). The subgradient condition for the constrained problem (S15) is

$$\frac{1}{n} \sum_{i=1}^n \hat{\Psi}_{S^0}(t_i) \left\{ \hat{\Psi}_{S^0}^{\mathrm{T}}(t_i) \hat{\theta}_{S^0} - Y_i \right\} + \lambda_n \hat{g}_{S^0} = 0. \tag{S18}$$

Adding and subtracting $\frac{1}{n}\sum_{i=1}^{n}\hat{\Psi}_{S^0}(t_i)\hat{\Psi}_{S^0}^{\mathrm{T}}(t_i)\theta_{S^0}^*$, we get

$$\frac{1}{n}\sum_{i=1}^{n}\left\{\hat{\Psi}_{S^0}(t_i)\hat{\Psi}_{S^0}^{\mathrm{T}}(t_i)\hat{\theta}_{S^0}-\hat{\Psi}_{S^0}(t_i)\hat{\Psi}_{S^0}^{\mathrm{T}}(t_i)\theta_{S^0}^*\right\}+$$
$$\frac{1}{n}\sum_{i=1}^{n}\left\{\hat{\Psi}_{S^0}(t_i)\hat{\Psi}_{S^0}^{\mathrm{T}}(t_i)\theta_{S^0}^*-\hat{\Psi}_{S^0}(t_i)Y_i\right\}+\lambda_n\hat{g}_{S^0}=0.$$

Rearranging the terms and letting

$$R_{S^0}\equiv\frac{1}{n}\sum_{i=1}^{n}\hat{\Psi}_{S^0}(t_i)\hat{\Psi}_{S^0}^{\mathrm{T}}(t_i)\theta_{S^0}^*-\frac{1}{n}\sum_{i=1}^{n}\hat{\Psi}_{S^0}(t_i)Y_i, \tag{S19}$$

we get

$$\hat{\theta}_{S^0}-\theta_{S^0}^*=-\left(\frac{1}{n}\sum_{i=1}^{n}\hat{\Psi}_{S^0}(t_i)\hat{\Psi}_{S^0}^{\mathrm{T}}(t_i)\right)^{-1}\left(R_{S^0}+\lambda_n\hat{g}_{S^0}\right). \tag{S20}$$

By the definition of $R_{S^0}$ in (S19), for each $k\in S$, we have that

$$R_k=\frac{1}{n}\sum_{i=1}^{n}\hat{\Psi}_k(t_i)\hat{\Psi}_{S^0}^{\mathrm{T}}(t_i)\theta_{S^0}^*-\frac{1}{n}\sum_{i=1}^{n}\hat{\Psi}_k(t_i)Y_i, \tag{S21}$$

and $R_0=\frac{1}{n}\sum_{i=1}^{n}t_i\{\hat{\Psi}_{S^0}^{\mathrm{T}}(t_i)\theta_{S^0}^*-Y_i\}$. By Condition S3, we know that $\|R_k\|_2\le\eta$ for $k\in S^0$. Hence,

$$\|R_{S^0}\|_2\le\eta\sqrt{s+1}. \tag{S22}$$

By Condition S1, we have that

$$\Lambda_{\max}\left\{\left(\frac{1}{n}\sum_{i=1}^{n}\hat{\Psi}_{S^0}(t_i)\hat{\Psi}_{S^0}^{\mathrm{T}}(t_i)\right)^{-1}\right\}\le\frac{2}{C_{\min}}. \tag{S23}$$

From (S13) and the fact that the largest eigenvalue of a submatrix is no greater than the largest eigenvalue of the matrix,

$$\frac{1}{2C_{\max}}\|\hat{g}_k\|_2^2 \leq \hat{g}_k^{\mathrm{T}}\left(\frac{1}{n}\sum_{i=1}^{n}\hat{\Psi}_k(t_i)\hat{\Psi}_k^{\mathrm{T}}(t_i)\right)^{-1}\hat{g}_k = 1, \quad k \in S.$$

Furthermore, $\hat{g}_0 = 0$ by construction. Hence,

$$\|\hat{g}_{S^0}\|_2 = \left\{\|\hat{g}_0\|_2^2 + \|\hat{g}_S\|_2^2\right\}^{1/2} \leq \sqrt{2sC_{\max}}. \tag{S24}$$

Therefore, combining (S20), (S22), (S23), and (S24), it follows that

$$\max_{k\in S}\|\hat{\theta}_k - \theta_k^*\|_2 \leq \|\hat{\theta}_{S^0} - \theta_{S^0}^*\|_2 \leq \frac{2\eta\sqrt{s+1}}{C_{\min}} + \lambda_n\frac{\sqrt{8sC_{\max}}}{C_{\min}} \leq \frac{2}{3}\theta_{\min},$$

where the last inequality follows from Condition S4.

Next, we verify strict feasibility (S17). For $k \notin S^0$, from (S12),

$$\frac{1}{n}\sum_{i=1}^{n}\hat{\Psi}_k(t_i)\big(\hat{\Psi}_{S^0}^{\mathrm{T}}(t_i)\hat{\theta}_{S^0} - Y_i\big) + \lambda_n\hat{g}_k = 0.$$

Adding and subtracting $\frac{1}{n}\sum_{i=1}^{n}\hat{\Psi}_k(t_i)\hat{\Psi}_{S^0}^{\mathrm{T}}(t_i)\theta_{S^0}^*$ yields

$$\begin{aligned}&\frac{1}{n}\sum_{i=1}^{n}\left\{\hat{\Psi}_k(t_i)\hat{\Psi}_{S^0}^{\mathrm{T}}(t_i)\hat{\theta}_{S^0} - \hat{\Psi}_k(t_i)\hat{\Psi}_{S^0}^{\mathrm{T}}(t_i)\theta_{S^0}^*\right\} + \\ &\frac{1}{n}\sum_{i=1}^{n}\left\{\hat{\Psi}_k(t_i)\hat{\Psi}_{S^0}^{\mathrm{T}}(t_i)\theta_{S^0}^* - \hat{\Psi}_k(t_i)Y_i\right\} + \lambda_n\hat{g}_k = 0.\end{aligned}$$

Rearranging the terms and plugging in (S20) and (S21), we get

$$\lambda_n \hat{g}_k = \frac{1}{n}\sum_{i=1}^{n} \hat{\Psi}_k(t_i)\hat{\Psi}_{S^0}^{\mathrm{T}}(t_i)\left(\frac{1}{n}\sum_{i=1}^{n} \hat{\Psi}_{S^0}(t_i)\hat{\Psi}_{S^0}^{\mathrm{T}}(t_i)\right)^{-1}(R_{S^0} + \lambda_n \hat{g}_{S^0}) - R_k.$$

By Condition S2, we know that

$$\max_{k \notin S^0}\left\|\left(\frac{1}{n}\sum_{i=1}^{n} \hat{\Psi}_k(t_i)\hat{\Psi}_{S^0}^{\mathrm{T}}(t_i)\right)\left(\frac{1}{n}\sum_{i=1}^{n} \hat{\Psi}_{S^0}(t_i)\hat{\Psi}_{S^0}^{\mathrm{T}}(t_i)\right)^{-1}\right\|_2 \leq 2\xi.$$

Recall from Condition S3 that $\|R_k\|_2 \leq \eta$ for $1 \leq k \leq p$. Using (S22) and (S24), we have that

$$\|\hat{g}_k\|_2 \leq \frac{2\xi\sqrt{s+1}+1}{\lambda_n}\eta + 2\xi\sqrt{s}\sqrt{2C_{\max}}, \quad k \notin S^0.$$

By Condition S4, $\|\hat{g}_k\|_2 < \sqrt{C_{\min}/2}$, and thus, applying Condition S1,

$$\hat{g}_k^{\mathrm{T}}\left(\frac{1}{n}\sum_{i=1}^{n} \hat{\Psi}_k(t_i)\hat{\Psi}_k^{\mathrm{T}}(t_i)\right)^{-1}\hat{g}_k \leq \frac{2\|\hat{g}_k\|_2^2}{C_{\min}} < 1, \quad k \notin S^0.$$

Therefore, we have established (S17). □

## A.4 Assumption S4 and technical lemmas

Theorem S1 characterizes the samples on which the GRADE estimator is able to reconstruct the true network. We must now establish that with high probability, the observations satisfy Conditions S1–S4. In Section A.5, Lemmas S3–S5, stated below, will be used to show that Conditions S1–S4, needed for Theorem S1, hold with high probability. Lemma S2 is used to prove Lemmas S3–S5. Lemmas S2 – S5 are proven in Appendix B.

First, we state the regularity condition on the bases $\psi$ mentioned in Section 4 in the main paper.

**Assumption S4.** The basis functions are orthonormal, i.e., $\int_0^1 \psi_{jk}\big(X_k^*(u)\big)\psi_{jk}^{\mathrm{T}}\big(X_k^*(u)\big)du = I_M$, where $I_M$ is an $M \times M$ identity matrix. The basis functions are bounded and have bounded first order derivative, i.e. $|\psi_m(x)| \leq B,\ |\psi_m'(x)| \leq D, m = 1, \ldots, M$. Further, under Assumption 3 in the main paper, for any $j, k$,

$$\int_0^1 \delta_{jk}^2(u)du = \int_0^1 \left\{f_{jk}^*(X_k^*(u)) - \psi^{\mathrm{T}}(X_k^*(u))\theta_{jk}^*\right\}^2 du \leq Q(M+1)^{-2\beta_2}, \tag{S25}$$

where $\theta_{jk}^*$ is defined in (S10) and $Q$ is a global constant.

*Remark* 1. Assumption S4 holds, for instance, when $\psi(\cdot)$ is the set of trigonometric basis functions (see, e.g., Section 1.7.3 in Tsybakov (2009)).

We next state the technical lemmas used in the proof of Theorem 2 in the main paper.

**Lemma S2.** *Suppose that Assumption 3 in the main paper and Assumption S4 hold, and $\psi(t) = (\psi_0(t), \psi_1(t), \ldots, \psi_M(t))^{\mathrm{T}}$ is of degree $M$. Then,*

$$\left| \|\theta_{jk}^*\|_2 - \left\{\int_0^1 \left[f_{jk}^*(X_k^*(u))\right]^2 du\right\}^{1/2} \right| \leq \sqrt{Q}M^{-\beta_2}. \tag{S26}$$

$$\left\|\left| X_j^* - \Psi_{S^0}^{\mathrm{T}}\theta_{S^0}^* \right|\right\| \leq s\sqrt{QM^{-2\beta_2}}, \tag{S27}$$

*and*

$$\frac{1}{n}\sum_{i=1}^n \{X_j^*(t_i) - \Psi_{S^0}^{\mathrm{T}}(t_i)\theta_{S^0}^*\}^2 \leq s^2 QM^{-2\beta_2} + o\left(n^{-2}\right), \tag{S28}$$

*where $\theta_{jk}^*$ is defined in* (S10) *and $Q$ is a constant in Assumption S4.*

**Lemma S3.** *Suppose that Assumptions 3 and 4 in the main paper and Assumption S4 hold. Let* $\Delta \equiv \max_{j=1,\ldots,p} \left\|\left\|\hat{X}_j - X_j^*\right\|\right\|$*. The following bounds on the eigenvalues of* $\sum_{i=1}^n \hat{\Psi}_{S^0}\hat{\Psi}_{S^0}^{\mathrm{T}}/n$ *hold:*

$$\begin{aligned} \Lambda_{\min}\left(\frac{1}{n}\sum_{i=1}^n \hat{\Psi}_{S^0}(t_i)\hat{\Psi}_{S^0}^{\mathrm{T}}(t_i)\right) &\geq C_{\min} - \left(2BD\Delta + \frac{BD+B^2}{6n^2}\right)(Ms+1), \\ \Lambda_{\max}\left(\frac{1}{n}\sum_{i=1}^n \hat{\Psi}_{S^0}(t_i)\hat{\Psi}_{S^0}^{\mathrm{T}}(t_i)\right) &\leq C_{\max} + \left(2BD\Delta + \frac{BD+B^2}{6n^2}\right)(Ms+1), \\ \text{and} \quad \Lambda_{\min}\left(\frac{1}{n}\sum_{i=1}^n \hat{\Psi}_{k}(t_i)\hat{\Psi}_{k}^{\mathrm{T}}(t_i)\right) &\geq C_{\min} - \left(2BD\Delta + \frac{BD+B^2}{6n^2}\right)M, \quad k \notin S_j^0. \end{aligned} \tag{S29}$$

**Lemma S4.** *Suppose that Assumptions 3 and 5 in the main paper and Assumption S4 hold. Let* $\Delta \equiv \max_{j=1,\ldots,p} \left\|\left\|\hat{X}_j - X_j^*\right\|\right\|$*. Then,*

$$\begin{aligned} &\left\|\left(\frac{1}{n}\sum_{i=1}^n \hat{\Psi}_{k}(t_i)\hat{\Psi}_{S^0}^{\mathrm{T}}(t_i)\right)\left(\frac{1}{n}\sum_{i=1}^n \hat{\Psi}_{S^0}(t_i)\hat{\Psi}_{S^0}^{\mathrm{T}}(t_i)\right)^{-1}\right\|_2 \leq \\ &\xi + \left\{c_1\hat{C}_{\min}^{-2}M(Ms+1)^3\Delta^2\right\}^{1/2} + \left\{c_2M(Ms+1)\Delta^2\right\}^{1/2} + \left\{c_3M(Ms+1)^3/6n^2\right\}^{1/2}, \end{aligned} \tag{S30}$$

*where* $\hat{C}_{\min} \equiv C_{\min} - \left(2BD\Delta + \frac{BD+B^2}{6n^2}\right)(Ms+1)$*, and* $c_1, c_2, c_3$ *are constants.*

**Lemma S5.** *Suppose Assumptions 1, 2, and 3 in the main paper and Assumption S4 hold. Let* $\Delta \equiv \max_{j=1,\ldots,p} \left\|\left\|\hat{X}_j - X_j^*\right\|\right\|$*. For each* $k = 0, \ldots, p$*,*

$$\left\|\frac{1}{n}\sum_{i=1}^n \hat{\Psi}_{k}(t_i)Y_{ij} - \frac{1}{n}\sum_{i=1}^n \hat{\Psi}_{k}(t_i)\hat{\Psi}_{S^0}^{\mathrm{T}}(t_i)\theta_{S^0}^*\right\|_2 \leq \eta, \tag{S31}$$

*where*

$$\eta \equiv M^{1/2}\left\{sM^{-\beta_2}Q^{1/2}B + BD\|\theta_S^*\|_1\Delta + n^{\alpha/2-1/2}\right\}$$

*with probability at least* $1 - 2M\exp\{-n^\alpha/(2B^2\sigma^2)\}$.

## A.5 Proof of Theorem 2

*Proof.* Notice that Theorem S1 offers the desired result of Theorem 2 in the main paper. We now verify that Conditions S1–S4 hold with high probability given the assumptions for Theorem 2 of the main paper. This completes the proof of Theorem 2 of the main paper.

First of all, Lemma S5 tells us that Condition S3 holds with probability at least $1-2pM\exp^{-n^{\alpha}/(2B^2\sigma^2)}$. This probability converges to unity as $p$ and $n$ grow, because $M \propto n^{\frac{2}{2\beta_2+1}\frac{\beta_1}{2\beta_1+1}(1-\alpha)} = o(n)$ and $pn\exp(-C_4 n^{\alpha}/\sigma^2) = o(1)$ as required in Theorem 2 of the main paper, where $C_4 \equiv \min\{1/(2B^2), 1/(2C_3^2)\}$. Thus, Condition S3 holds with high probability.

Next, we verify that Condition S4 holds with high probability. Given Assumptions 1–2 and S1–S3, we know from Theorem 1 in the main paper that

$$\max_j \left\| \hat{X}_j - X_j^* \right\| \equiv \Delta = O\left(n^{\frac{\beta_1}{2\beta_1+1}(\alpha-1)}\right), \tag{S32}$$

with probability at least $1-2p\exp\{-n^{\alpha}/(2C_3\sigma^2)\}$. Recall that in Theorem 2 of the main paper we require that $s = O(n^{\gamma})$ and $M \propto n^{\frac{2}{2\beta_2+1}\frac{\beta_1}{2\beta_1+1}(1-\alpha)}$. Furthermore, $\|\theta_k^*\|_1 < \sqrt{M}\|\theta_k^*\|_2$, and $\|\theta_k^*\|_2$ is bounded by a constant due to the fact that $f_{jk}^*$ is bounded and (S26). Combining these with (S32), we know that the three terms of $\eta$ in Condition S3 satisfy

$$sM^{-\beta_2+1/2}Q^{1/2}B = O\left(n^{-\frac{2\beta_2-1}{2\beta_2+1}\frac{\beta_1}{2\beta_1+1}(1-\alpha)+\gamma}\right),$$

$$M^{1/2}BD\|\theta_S^*\|_1\Delta = O\left(n^{-\frac{2\beta_2-1}{2\beta_2+1}\frac{\beta_1}{2\beta_1+1}(1-\alpha)+\gamma}\right),$$

and

$$M^{1/2} n^{\alpha/2-1/2} = O\left(n^{(\frac{1}{2\beta_2+1}\frac{\beta_1}{2\beta_1+1}-\frac{1}{2})(1-\alpha)}\right).$$

These lead to

$$\left\| \frac{1}{n}\sum_{i=1}^{n} \hat{\Psi}_k(t_i) Y_{ij} - \frac{1}{n}\sum_{i=1}^{n} \hat{\Psi}_k(t_i) \hat{\Psi}_{S^0}^{\mathrm{T}}(t_i) \theta_{S^0}^{*} \right\|_2 \leq \eta = O\left(n^{-\frac{2\beta_2-1}{2\beta_2+1}\frac{\beta_1}{2\beta_1+1}(1-\alpha)+\gamma}\right) \tag{S33}$$

with probability at least $1 - 2pM\exp\{-n^{\alpha}/(2B^2\sigma^2)\}$ for all $k = 0, \ldots, p$, from Lemma S5.

In Theorem 2 of the main paper, we require that $\lambda_n \propto n^{-\frac{\beta_1}{2\beta_1+1}\frac{2\beta_2-1}{2\beta_2+1}(1-\alpha)+2\gamma}$. Given (S33) and $s = O(n^{\gamma})$, we know that $\sqrt{s}\eta = o(\lambda_n)$. Furthermore, define

$$H_1(\beta_1, \beta_2, \alpha) \equiv \min\left\{ \frac{\beta_1}{2\beta_1+1}\frac{2\beta_2-1}{4\beta_2+2}(1-\alpha), \frac{2}{3}\frac{\beta_1}{2\beta_1+1}\frac{2\beta_2-3}{2\beta_2+1}(1-\alpha) \right\}. \tag{S34}$$

Then,

$$-\frac{\beta_1}{2\beta_1+1}\frac{2\beta_2-1}{2\beta_2+1}(1-\alpha) + 2\gamma \leq -2H_1(\beta_1, \beta_2, \alpha) + 2\gamma.$$

Thus, $\lambda_n = o(1)$ for $\gamma < H_1(\beta_1, \beta_2, \alpha)$. Further notice that $M^{-\beta_2} \propto n^{-\frac{2\beta_2}{2\beta_2+1}\frac{\beta_1}{2\beta_1+1}(1-\alpha)} = o(1)$, which implies that $\theta_{\min} \geq 3f_{\min}/4$ for sufficiently large $n$ from (S26) in Lemma S2. As a result, the two inequalities in Condition S4 become

$$o(\lambda_n) + \lambda_n \frac{\sqrt{8sC_{\max}}}{C_{\min}} \leq \frac{f_{\min}}{2},$$

$$o(1) + 2\xi\sqrt{s}\sqrt{2C_{\max}} < \sqrt{C_{\min}/2},$$

which hold for sufficiently large $n$ under Assumption 6 of the main paper.

Note that the probability that (S32) and (S33) both hold is at least $1 - 2pM\exp\{-n^{\alpha}/(2B^2\sigma^2)\} - 2p\exp\{-n^{\alpha}/(2C_3^2\sigma^2)\}$. Letting $C_4 = \min\{1/(2B^2), 1/(2C_3^2)\}$, we know from Theorem 2 that $pn\exp(-C_4 n^{\alpha}/\sigma^2) = o(1)$. Combining this with $M \propto n^{\frac{2}{2\beta_2+1}\frac{\beta_1}{2\beta_1+1}(1-\alpha)} = o(n)$, we know that

$1-2pM\exp\{-n^{\alpha}/(2B^2\sigma^2)\}-2p\exp\{-n^{\alpha}/(2C_3^2\sigma^2)\}$ converges to 1 as $p$, $s$, and $n$ grow. Therefore, Condition S4 holds with high probability.

Finally, we establish that Conditions S1 and S2 hold with high probability. Note that the dominant terms not involving $C_{\min}, C_{\max}$ or $\xi$ in the bounds in (S29) in Lemma S3 and (S30) in Lemma S4 involve $sM\Delta$ and $s^{3/2}M^2\Delta$, respectively. Given (S32), one can check that

$$sM\Delta \propto n^{\frac{\beta_1}{2\beta_1+1}\frac{2\beta_2-1}{2\beta_2+1}(1-\alpha)+\gamma} = o(1), \text{ and} \tag{S35}$$

$$s^{3/2}M^2\Delta \propto n^{\frac{\beta_1}{2\beta_1+1}\frac{2\beta_2-3}{2\beta_2+1}(1-\alpha)+\frac{3}{2}\gamma} = o(1), \tag{S36}$$

where we have used the fact that $\beta_2 \geq 3$ in Assumption 3 in the main paper as well as the fact that $\gamma < H_1(\beta_1, \beta_2, \alpha)$ from the statement of Theorem 2 in the main paper. Since (S32) and (S33) hold with high probability, combining the inequalities in Lemmas S3 and S4 with (S35) and (S36), we see that Conditions S1 and S2 hold with high probability given Assumptions 3, 4 and 5 in the main paper.

In summary, we have shown that Conditions S1–S4 hold with high probability. Applying Theorem S1 establishes that the GRADE estimator $\hat{S}_j$ in (17) in the main paper recovers the true support $S_j^*$. □

## A.6 Proof of Proposition 1

In Proposition 1, the choice of bandwidth $h_n$ is different from that in Theorems 1 and 2 of the main paper. In order to prove Proposition 1 of the main paper, we establish the following concentration inequality for $\left|\left|\left|\hat{X}_j - X_j^*\right|\right|\right|$, where the bandwidth is chosen as specified in Proposition 1 of the main paper.

**Proposition S1.** *Suppose that Assumptions 1–2 in the main paper and S1–S3 hold. Let $\hat{X}_j$ be the local polynomial regression estimator of order $\ell = \lfloor \beta_1 \rfloor$ with bandwidth*

$$h_n \propto n^{-1/(2\beta_1+1)}.$$

*There exists a constant $C_2 < \infty$ such that for each $j = 1, \ldots, p$,*

$$\left|\left|\left| \hat{X}_j - X_j^* \right|\right|\right|^2 \leq C_2 n^{\alpha - \frac{2\beta_1}{2\beta_1+1}}$$

*holds with probability at least $1 - 2\exp\{-n^\alpha/(2\sigma^2 C_3^2)\}$.*

The proof of Proposition S1 is similar to that for Theorem 1 in the main paper by plugging in $h_n \propto n^{-1/(2\beta_1+1)}$ in (S9).

Given Proposition S1, the proof of Proposition 1 in the main paper follows from a similar argument as in the proof of Theorem 2 in the main paper, and is thus omitted here. The constant $H_2(\beta_1, \beta_2, \alpha)$ is defined as

$$H_2(\beta_1, \beta_2, \alpha) \equiv \min\left\{ \frac{\beta_1}{2\beta_1+1}\frac{2\beta_2-1}{2\beta_2+1} - \alpha, \frac{1}{3}\frac{\beta_1}{2\beta_1+1}\frac{2\beta_2-3}{2\beta_2+1} - \alpha \right\}.$$

# B. PROOFS OF TECHNICAL LEMMAS

## B.1 Proof of Lemma S2

In this section, in the interest of clarity, we bring back the subscript $j$ in $\theta_j^*, \theta_{jk}^*, \theta_{jS^0}^*$ and $f_{jk}^*$.

*Proof.* Recall that in Assumption S4, (S25) says that

$$\int_0^1 \delta_{jk}^2(u)du = \int_0^1 \left\{f_{jk}^*(X_k^*(u)) - \psi^{\mathrm{T}}(X_k^*(u))\theta_{jk}^*\right\}^2 du \le Q(M+1)^{-2\beta_2}.$$

It follows from the triangle inequality that

$$\left|\left\{\int_0^1 \left[\psi^{\mathrm{T}}(X_k^*(u))\theta_{jk}^*\right]^2 du\right\}^{1/2} - \left\{\int_0^1 \left[f_{jk}^*(X_k^*(u))\right]^2 du\right\}^{1/2}\right| \le \sqrt{Q}M^{-\beta_2}.$$

The orthogonality of $\psi$ in Assumption S4 then leads to (S26), i.e.,

$$\left|\|\theta_{jk}^*\|_2 - \left\{\int_0^1 \left[f_{jk}^*(X_k^*(u))\right]^2 du\right\}^{1/2}\right| \le \sqrt{Q}M^{-\beta_2}.$$

From (S25), we can also see that

$$\begin{aligned}\left|\int_0^t \delta_{jk}(u)\,du\right| \le \left\{\int_0^t \delta_{jk}^2(u)\,du\right\}^{1/2}\left\{\int_0^t 1^2\,du\right\}^{1/2} \le \left\{\int_0^1 \delta_{jk}^2(u)\,du\right\}^{1/2} \\ \le \sqrt{Q(M+1)^{-2\beta_2}} \le \sqrt{QM^{-2\beta_2}},\end{aligned}$$

where we use the fact that $t \in [0,1]$.

Recall from (15) in the main paper and (S10) that

$$X_j^*(t) = \theta_{j0}^* t + \sum_{k=1}^p \Psi_k^{\mathrm{T}}(t)\theta_{jk}^* + \sum_{k=1}^p \int_0^t \delta_{jk}(u)\,du,$$

where we let $X_j^*(0) = 0$ for ease of discussion. We know that both $\theta_{jk}^*$ and $\delta_{jk}$ are zero for $k \notin S$.

Thus, the errors that result from the use of truncated bases are bounded by

$$\left|\left|\left|X_j^* - \Psi_{S_j^0}^{\mathrm{T}}\theta_{jS_j^0}^*\right|\right|\right| = \left|\left|\left|X_j^* - \theta_{j0}^* t - \sum_{k\in S_j}\Psi_k^{\mathrm{T}}\theta_{jk}^*\right|\right|\right| = \left[\int_0^1 \left\{\sum_{k\in S_j}\int_0^t \delta_{jk}(u)du\right\}^2 dt\right]^{\frac{1}{2}}$$
$$\leq \left[\int_0^1 \left\{s\sqrt{QM^{-2\beta_2}}\right\}^2 dt\right]^{\frac{1}{2}} \leq s\sqrt{QM^{-2\beta_2}}.$$

The error bound in (S27) is on the whole trajectories, whereas we only observe discrete measurements of the trajectories in reality. The bound in (S28) addresses this case and is proved below.

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^n\{X_j^*(t_i) - \Psi_{S_j^0}^{\mathrm{T}}(t_i)\theta_{jS_j^0}^*\}^2 &= \frac{1}{n}\sum_{i=1}^n\left\{\sum_{k\in S_j}\int_0^{t_i}\delta_{jk}(u)du\right\}^2\\ &\leq \int_0^1\left\{\sum_{k\in S_j}\int_0^t\delta_{jk}(u)du\right\}^2 dt + o\left(\frac{1}{n^2}\right)\\ &\leq s^2QM^{-2\beta_2} + o\left(n^{-2}\right),\end{aligned}$$

where the last inequality follows from (S27) and the second to last inequality follows from the trapezoidal rule on a uniform grid.

□

## B.2 Proof of Lemma S3

We first review some known results on matrix norms and eigenvalues. For an $m \times n$ matrix $A$,

$$\|A\|_2 = \sup_{x\in\mathbb{R}^n}\frac{\|Ax\|_2}{\|x\|_2} = \sup_{\|x\|_2=1}\left\{\sum_{i=1}^m\left(\sum_{j=1}^n a_{ij}x_j\right)^2\right\}^{\frac{1}{2}} \leq \left(\sum_{i=1}^m\sum_{j=1}^n a_{ij}^2\right)^{\frac{1}{2}} \equiv \|A\|_F, \tag{S37}$$

where $\|\cdot\|_F$ is the Frobenius norm. We remind the reader that for a symmetric matrix $A$ that is not positive semi-definite, $\Lambda_{\max}(A) \leq \|A\|_2$. The following two inequalities are useful in the proofs. Let $A$ and $\hat{A}$ be two $n \times n$ symmetric matrices.

1. Weyl's inequality (Weyl, 1912) states that

$$\Lambda_{\min}(A) - \Lambda_{\max}(\hat{A} - A) \leq \Lambda_{\min}(\hat{A}), \text{ and } \Lambda_{\max}(\hat{A}) \leq \Lambda_{\max}(A) + \Lambda_{\max}(\hat{A} - A),$$

   which leads to

$$\Lambda_{\min}(A) - \|\hat{A} - A\|_2 \leq \Lambda_{\min}(\hat{A}), \text{ and } \Lambda_{\max}(\hat{A}) \leq \Lambda_{\max}(A) + \|\hat{A} - A\|_2. \tag{S38}$$

2. The Gershgorin circle theorem (Gershgorin, 1931) states that

$$\|\hat{A} - A\|_2 \leq \max_i \sum_{j=1}^{n} |(\hat{A} - A)_{ij}| \leq n\|\hat{A} - A\|_\infty, \tag{S39}$$

   where the norm $\|\cdot\|_\infty$ is defined as $\|A\|_\infty = \max_{i,j} |A_{ij}|$.

We are now ready to prove Lemma S3.

*Proof.* Let $A \equiv \int_0^1 \Psi_{S^0}(t)\Psi_{S^0}^T(t)\,dt$, $A_n \equiv \frac{1}{n}\sum_{i=1}^n \Psi_{S^0}(t_i)\Psi_{S^0}^T(t_i)$, $\hat{A}_n \equiv \frac{1}{n}\sum_{i=1}^n \hat{\Psi}_{S^0}(t_i)\hat{\Psi}_{S^0}^T(t_i)$, which are $(Ms+1) \times (Ms+1)$ matrices. Then,

$$\begin{aligned} \Lambda_{\min}(\hat{A}_n) \geq & \Lambda_{\min}(A) - \|\hat{A}_n - A\|_2 \\ \geq & \Lambda_{\min}(A) - \|A_n - A\|_2 - \|\hat{A}_n - A_n\|_2, \end{aligned} \tag{S40}$$

where the first inequality follows from (S38) and the second follows from the triangle inequality.

Furthermore,

$$
\begin{aligned}
\|\hat{A}_n - A_n\|_2 \leq & (Ms+1)\|\hat{A}_n - A_n\|_\infty \\
\leq & (Ms+1)\left\|\frac{1}{n}\sum_{i=1}^{n}\left\{\hat{\Psi}_{S^0}(t_i)\hat{\Psi}_{S^0}^{\mathrm{T}}(t_i) - \Psi_{S^0}(t_i)\Psi_{S^0}^{\mathrm{T}}(t_i)\right\}\right\|_\infty \\
\leq & \frac{Ms+1}{n}\left\|\sum_{i=1}^{n}\hat{\Psi}_{S^0}(t_i)\left\{\hat{\Psi}_{S^0}^{\mathrm{T}}(t_i) - \Psi_{S^0}^{\mathrm{T}}(t_i)\right\}\right\|_\infty + \\
& \frac{Ms+1}{n}\left\|\sum_{i=1}^{n}\Psi_{S^0}(t_i)\left\{\hat{\Psi}_{S^0}^{\mathrm{T}}(t_i) - \Psi_{S^0}^{\mathrm{T}}(t_i)\right\}\right\|_\infty \\
\leq & \frac{Ms+1}{n}\left\|\sum_{i=1}^{n}\hat{\Psi}_{S^0}(t_i)D\Delta\right\|_\infty + \frac{Ms+1}{n}\left\|\sum_{i=1}^{n}\Psi_{S^0}(t_i)D\Delta\right\|_\infty \\
\leq & \frac{2Ms+2}{n}\|nBD\Delta\|_\infty = 2(Ms+1)BD\Delta,
\end{aligned}
\tag{S41}
$$

where the first inequality follows from (S39), the last inequality follows from the bounds in Assumption S4, and the second to last inequality follows from the following inequality: for $k \in S^0$ and $m = 1, \ldots, M$,

$$
\begin{aligned}
|\hat{\Psi}_{km}(t_i) - \Psi_{km}(t_i)| &= \left|\int_0^{t_i}\psi_m(\hat{X}_k(u))\,du - \int_0^{t_i}\psi_m(X_k^*(u))\,du\right| \\
&= \left|\int_0^{t_i}\{\psi_m(\hat{X}_k(u)) - \psi_m(X_k^*(u))\}\,du\right| \\
&\leq \left|\int_0^{t_i}|D\{\hat{X}_k(u) - X_k^*(u)\}|\,du\right| \\
&\leq \left\{\int_0^{t_i}D^2\,du\right\}^{1/2}\left\{\int_0^{t_i}(\hat{X}_k(u) - X_k^*(u))^2\,du\right\}^{1/2} \\
&\leq D\left|\!\left|\!\left|\hat{X}_k - X_k^*\right|\!\right|\!\right| \leq D\Delta.
\end{aligned}
\tag{S42}
$$

Here the first inequality follows from the mean-value theorem and the bounds in Assumption S4.

Now, from (S39),

$$
\|A_n - A\|_2 \leq (Ms+1)\|A_n - A\|_\infty \leq (Ms+1)\frac{BD+B^2}{6n^2}, \tag{S43}
$$

where for each element of the matrix $A_n - A = \frac{1}{n}\sum_{i=1}^n \Psi_{S^0}(t_i)\Psi_{S^0}^{\mathrm{T}}(t_i) - \int_0^1 \Psi_{S^0}(t)\Psi_{S^0}^{\mathrm{T}}(t)\,dt$,

$$\begin{aligned}
&\left|\frac{1}{n}\sum_{i=1}^{n}\Psi_{km_1}(t_i)\Psi_{lm_2}(t_i) - \int_0^1 \Psi_{km_1}(t)\Psi_{lm_2}(t)\,dt\right| \\
\leq & \frac{\left|\{\Psi_{km_1}(u)\Psi_{lm_2}(u)\}''\right|}{12n^2} \leq \frac{\left|2\Psi'_{km_1}(u)\Psi'_{lm_2}(u) + \Psi''_{km_1}(u)\Psi_{lm_2}(u) + \Psi'_{km_1}(u)\Psi''_{lm_2}(u)\right|}{12n^2} \\
\leq & \frac{2B^2 + BD + BD}{12n^2} = \frac{BD + B^2}{6n^2},
\end{aligned}$$

where derivatives are taken with respect to $t$. By the trapezoid rule on a uniform grid, the first inequality holds for some $u \in [0,1]$. The second inequality makes use of the bounds in Assumption S4, which imply that

$$|\Psi'_{km}(t)| = \left|\left(\int_0^t \psi_{km}(s)\,ds\right)'\right| = |\psi_{km}(t)| \leq B$$

and

$$|\Psi''_{km}(t)| = \left|\left(\int_0^t \psi_{km}(s)\,ds\right)''\right| = |\psi'_{km}(t)| \leq D.$$

In summary, combining (S40), (S41), and (S43),

$$\begin{aligned}
\Lambda_{\min}(\hat{A}_n) \geq & \Lambda_{\min}(A) - \left(2BD\Delta + \frac{BD + B^2}{6n^2}\right)(Ms+1) \\
\geq & C_{\min} - \left(2BD\Delta + \frac{BD + B^2}{6n^2}\right)(Ms+1).
\end{aligned}$$

The upper bound for $\Lambda_{\max}(\hat{A}_n)$ and the lower bound for $\Lambda_{\min}\left(\frac{1}{n}\sum_{i=1}^n \hat{\Psi}_k(t_i)\hat{\Psi}_k^{\mathrm{T}}(t_i)\right)$ can be established in a similar manner. $\square$

### B.3 Proof of Lemma S4

*Proof.* Define $A$, $A_n$, and $\hat{A}_n$ as in the proof for Lemma S3. We let $F = \int_0^1 \Psi_k \Psi_{S^0}^{\mathrm{T}}\, dt$, $F_n = \sum_{i=1}^n \Psi_k(t_i)\Psi_{S^0}^{\mathrm{T}}(t_i)/n$, and $\hat{F}_n = \sum_{i=1}^n \hat{\Psi}_k(t_i)\hat{\Psi}_{S^0}^{\mathrm{T}}(t_i)/n$. $F, F_n$, and $\hat{F}_n$ are $M \times (Ms+1)$ matrices. We let $\hat{C}_{\min}$ denote the lower bound of $\Lambda_{\min}(\hat{A}_n)$ established in Lemma S3, i.e.,

$$\hat{C}_{\min} \equiv C_{\min} - \left(2BD\Delta + \frac{BD+B^2}{6n^2}\right)(Ms+1).$$

To prove the result, we need to bound $\|\hat{F}_n \hat{A}_n^{-1}\|_2$. Note that

$$\begin{aligned}\|\hat{F}_n \hat{A}_n^{-1}\|_2 &\leq \|\hat{F}_n \hat{A}_n^{-1} - \hat{F}_n A_n^{-1} + \hat{F}_n A_n^{-1} - F_n A_n^{-1} + F_n A_n^{-1}\|_2 \\ &\leq \|\hat{F}_n(\hat{A}_n^{-1} - A_n^{-1})\|_2 + \|(\hat{F}_n - F_n)A_n^{-1}\|_2 + \|F_n A_n^{-1}\|_2 \\ &\equiv \|\mathrm{I}\|_2 + \|\mathrm{II}\|_2 + \|\mathrm{III}\|_2.\end{aligned}$$

Using sub-multiplicity of the $\ell_2$-norm of matrices,

$$\|\mathrm{I}\|_2^2 \leq \|\hat{F}_n\|_2^2 \|\hat{A}_n^{-1} - A_n^{-1}\|_2^2.$$

Applying (S37) to $\hat{F}_n$, we get

$$\|\mathrm{I}\|_2^2 \leq M(Ms+1)\left(\max_{i,j} \hat{F}_{n,ij}^2\right)\|\hat{A}_n^{-1} - A_n^{-1}\|_2^2.$$

Recalling that $\hat{F}_n = \sum_{i=1}^n \hat{\Psi}_k(t_i)\hat{\Psi}_{S^0}^{\mathrm{T}}(t_i)/n$ and that $|\hat{\Psi}_{km}(t_i)| \leq B$,

$$\|\mathbf{I}\|_2^2 \leq M(Ms+1)(\sum_{i=1}^n B^2/n)^2\|\hat{A}_n^{-1} - A_n^{-1}\|_2^2.$$

Note that $\hat{A}_n^{-1} - A_n^{-1} = \hat{A}_n^{-1}(A_n - \hat{A}_n)A_n^{-1}$. Thus,

$$\begin{aligned}
\|\mathbf{I}\|_2^2 \leq& M(Ms+1)B^4\|\hat{A}_n^{-1}\|_2^2\|\hat{A}_n - A_n\|_2^2\|A_n^{-1}\|_2^2 \\
\leq& M(Ms+1)B^4\hat{C}_{\min}^{-2}\|\hat{A}_n - A_n\|_2^2 C_{\min}^{-2} \\
\leq& M(Ms+1)B^4\{2(Ms+1)DB\Delta\}^2\hat{C}_{\min}^{-2}C_{\min}^{-2}, \\
\equiv& c_1\hat{C}_{\min}^{-2}M(Ms+1)^3\Delta^2,
\end{aligned}$$

where the last two inequalities follow from the proof of Lemma S3.

Next, note that

$$\begin{aligned}
\|\mathbf{II}\|_2^2 =& \|(\hat{F}_n - F_n)A_n^{-1}\|_2^2 \\
\leq& C_{\min}^{-2}\left\|\frac{1}{n}\sum_{i=1}^n \hat{\Psi}_k(t_i)\hat{\Psi}_{S^0}^{\mathrm{T}}(t_i) - \frac{1}{n}\sum_{i=1}^n \Psi_k(t_i)\Psi_{S^0}^{\mathrm{T}}(t_i)\right\|_2^2 \\
\leq& C_{\min}^{-2}\left\|\frac{1}{n}\sum_{i=1}^n \hat{\Psi}_k(t_i)\left\{\hat{\Psi}_{S^0}^{\mathrm{T}}(t_i) - \Psi_{S^0}^{\mathrm{T}}(t_i)\right\}\right\|_2^2 + \\
& C_{\min}^{-2}\left\|\frac{1}{n}\sum_{i=1}^n \left\{\hat{\Psi}_k(t_i) - \Psi_k(t_i)\right\}\Psi_{S^0}^{\mathrm{T}}(t_i)\right\|_2^2 \\
\leq& 2C_{\min}^{-2}B^2D^2\Delta^2M(Ms+1) \\
\equiv& c_2M(Ms+1)\Delta^2,
\end{aligned}$$

where the first inequality follows from sub-multiplicity of norms of matrices, and the last from (S37), (S42), and the bounds in Assumption S4.

Finally,

$$
\begin{aligned}
\|\mathrm{III}\|_2 =& \|F_n A_n^{-1}\|_2 = \|F_n(A_n^{-1} - A^{-1}) + (F_n - F)A^{-1} + FA^{-1}\|_2 \\
\leq& \xi + \|F_n\|_2 \|A_n^{-1} - A^{-1}\|_2 + \|(F_n - F)A^{-1}\|_2 \\
\leq& \xi + \left\{M(Ms+1)B^4 \|A_n^{-1} - A^{-1}\|_2^2\right\}^{1/2} + \{\|F_n - F\|_2^2 C_{\min}^{-2}\}^{1/2} \\
\leq& \xi + \left\{M(Ms+1)B^4 \|A_n^{-1}\|_2^2 \|A_n - A\|_2^2 \|A^{-1}\|_2^2\right\}^{1/2} + \{\|F_n - F\|_2^2 C_{\min}^{-2}\}^{1/2} \\
\leq& \xi + \left\{M(Ms+1)B^4 \hat{C}_{\min}^{-2} C_{\min}^{-2} \|A_n - A\|_2^2\right\}^{1/2} + \{\|F_n - F\|_2^2 C_{\min}^{-2}\}^{1/2} \\
\leq& \xi + \left\{M(Ms+1)B^4 \hat{C}_{\min}^{-2} C_{\min}^{-2} (Ms+1)^2 \frac{BD + B^2}{6n^2}\right\}^{1/2} + \\
& \left\{M(Ms+1)C_{\min}^{-2} \frac{BD + B^2}{6n^2}\right\}^{1/2} \\
\leq& \xi + \left\{c_3 M(Ms+1)^3/6n^2\right\}^{1/2},
\end{aligned}
$$

where the first inequality follows from Assumption 5 in the main paper and the second to last inequality follows from (S43).

In summary,

$$
\left\| \frac{1}{n} \sum_{i=1}^{n} \hat{\Psi}_k(t_i) \hat{\Psi}_{S^0}^{\mathrm{T}}(t_i) \left( \frac{1}{n} \sum_{i=1}^{n} \hat{\Psi}_{S^0}(t_i) \hat{\Psi}_{S^0}^{\mathrm{T}}(t_i) \right)^{-1} \right\|_2 \leq
$$

$$
\xi + \left\{c_1 M(Ms+1)^3 \Delta^2\right\}^{1/2} + \left\{c_2 M(Ms+1)\Delta^2\right\}^{1/2} + \left\{c_3 M(Ms+1)^3/6n^2\right\}^{1/2}.
$$

where $c_1, c_2, c_3$ are constants. $\square$

### B.4 Proof of Lemma S5

*Proof.* For $k = 1, \ldots, p$,

$$\begin{aligned}
&\left\| \frac{1}{n} \sum_{i=1}^{n} \hat{\Psi}_k(t_i) Y_{ij} - \frac{1}{n} \sum_{i=1}^{n} \hat{\Psi}_k(t_i) \hat{\Psi}_{S^0}^{\mathrm{T}}(t_i) \theta_{S^0}^* \right\|_2 \\
=&\frac{1}{n} \left\| \sum_{i=1}^{n} \hat{\Psi}_k(t_i) X_j^*(t_i) + \sum_{i=1}^{n} \hat{\Psi}_k(t_i) \epsilon_{ji} - \sum_{i=1}^{n} \hat{\Psi}_k(t_i) \Psi_{S^0}^{\mathrm{T}}(t_i) \theta_{S^0}^* + \right. \\
&\left. \sum_{i=1}^{n} \hat{\Psi}_k(t_i) \Psi_{S^0}^{\mathrm{T}}(t_i) \theta_{S^0}^* - \sum_{i=1}^{n} \hat{\Psi}_k(t_i) \hat{\Psi}_{S^0}^{\mathrm{T}}(t_i) \theta_{S^0}^* \right\|_2 \\
\leq& \left\| \frac{1}{n} \sum_{i=1}^{n} \hat{\Psi}_k(t_i) \{ X_j^*(t_i) - \Psi_{S^0}^{\mathrm{T}}(t_i) \theta_{S^0}^* \} \right\|_2 + \left\| \frac{1}{n} \sum_{i=1}^{n} \hat{\Psi}_k(t_i) \{ \Psi_{S^0}^{\mathrm{T}}(t_i) - \hat{\Psi}_{S^0}^{\mathrm{T}}(t_i) \} \theta_{S^0}^* \right\|_2 + \\
&\left\| \frac{1}{n} \sum_{i=1}^{n} \hat{\Psi}_k(t_i) \epsilon_{ji} \right\|_2 \\
\equiv& \|\mathrm{I}\|_2 + \|\mathrm{II}\|_2 + \|\mathrm{III}\|_2.
\end{aligned}$$

First, applying the Cauchy-Schwarz inequality to $\|\mathrm{I}\|_2^2$,

$$\|\mathrm{I}\|_2^2 \leq \sum_{m=1}^{M} \left[ \frac{1}{n^2} \sum_{i=1}^{n} \hat{\Psi}_{km}^2(t_i) \sum_{i=1}^{n} \left\{ X_j^*(t_i) - \Psi_{S^0}^{\mathrm{T}}(t_i) \theta_{S^0}^* \right\}^2 \right].$$

From the bounds in Assumption S4 and (S28) ,

$$\|\mathrm{I}\|_2^2 \leq M \left\{ \frac{1}{n^2} \left( nB^2 \right) \left( s^2 n Q M^{-2\beta_2} \right) \right\} = s^2 M^{-2\beta_2+1} Q B^2.$$

Next, note that $\hat{\Psi}_0(t_i) - \Psi_0(t_i) = t_i - t_i = 0$, we have $\left\{ \Psi_{S^0}^{\mathrm{T}}(t_i) - \hat{\Psi}_{S^0}^{\mathrm{T}}(t_i) \right\} \theta_{S^0}^* = \left\{ \Psi_S(t_i) - \hat{\Psi}_S(t_i) \right\}^{\mathrm{T}} \theta_S^*$.

Thus applying the Cauchy-Schwarz inequality to $\|\mathrm{II}\|_2^2$,

$$\|\mathrm{II}\|_2^2 \le \sum_{m=1}^{M} \left( \frac{1}{n^2} \sum_{i=1}^{n} \hat{\Psi}_{km}^2(t_i) \sum_{i=1}^{n} \left[ \left\{ \Psi_S(t_i) - \hat{\Psi}_S(t_i) \right\}^{\mathrm{T}} \theta_S^* \right]^2 \right).$$

Applying the norm inequality $a^{\mathrm{T}} b \le \|a\|_\infty \|b\|_1$ to $\left\{ \Psi_S(t_i) - \hat{\Psi}_S(t_i) \right\}^{\mathrm{T}} \theta_S^*$ and using the inequality (S42) as well as the bounds in Assumption S4, we get

$$\|\mathrm{II}\|_2^2 \le M \left\{ \frac{1}{n^2} n B^2 \sum_{i=1}^{n} \|\theta_S^*\|_1^2 D^2 \Delta^2 \right\} \le M B^2 D^2 \|\theta_S^*\|_1^2 \Delta^2.$$

Finally, $\mathrm{III} = \frac{1}{n} \sum_{i=1}^{n} \hat{\Psi}_k(t_i) \epsilon_{ji}$ is an $M$-vector. For each $m = 1, \ldots, M$, we let $g(\epsilon_j/\sigma) = \sum_{i=1}^{n} \hat{\Psi}_{km}(t_i) \epsilon_{ji}/n$. Then, for $a, b \in \mathbb{R}^p$,

$$\begin{aligned} |g(a) - g(b)| &= \left| \sigma \sum_{i=1}^{n} \hat{\Psi}_{km}(t_i)(a_i - b_i)/n \right| \\ &\le \frac{\sigma}{n} \left\{ \sum_{i=1}^{n} \hat{\Psi}_{km}^2(t_i) \right\}^{0.5} \|a - b\|_2 \le \frac{\sigma}{n} \sqrt{nB^2} \|a - b\|_2. \end{aligned}$$

This shows that $g(\cdot)$ is an $L_3$-Lipshitz function with $L_3 = \sigma B/\sqrt{n}$. Note that $\mathbb{E} g(\epsilon_j/\sigma) = 0$. Thus, by Theorem 5.6 in Boucheron et al. (2013) presented in Section A.2, we have

$$\Pr(|g(\epsilon_j/\sigma)| \ge v) \le 2 \exp\{-v^2 n/(2B^2\sigma^2)\}.$$

Letting $v = n^{\alpha/2 - 0.5}$, $\|\mathrm{III}\|_2^2 \le n^{\alpha - 1} M$ holds with probability at least $1 - 2M \exp\{-n^\alpha/(2B^2\sigma^2)\}$.

Combining all of the pieces, we find that

$$\|\mathrm{I}\|_2 + \|\mathrm{II}\|_2 + \|\mathrm{III}\|_3 \le \eta \equiv M^{1/2} \left\{ s M^{-\beta_1} Q^{1/2} B + B D \|\theta_S^*\|_1 \Delta + n^{\frac{\alpha}{2} - \frac{1}{2}} \right\}$$

with probability at least $1 - 2M\exp\{-n^\alpha/(2B^2\sigma^2)\}$.

For $k = 0$,

$$\left\|\frac{1}{n}\sum_{i=1}^{n}\hat{\Psi}_0(t_i)\left\{Y_{ij} - \hat{\Psi}_{S^0}^{\mathrm{T}}(t_i)\theta_{S^0}^*\right\}\right\|_2 = \left\|\frac{1}{n}\sum_{i=1}^{n}t_i\left\{Y_{ij} - \hat{\Psi}_{S^0}^{\mathrm{T}}(t_i)\theta_{S^0}^*\right\}\right\|_2$$
$$\leq \left\|\frac{1}{n}\sum_{i=1}^{n}t_i\{X_j^*(t_i) - \Psi_{S^0}^{\mathrm{T}}(t_i)\theta_{S^0}^*\}\right\|_2 + \left\|\frac{1}{n}\sum_{i=1}^{n}t_i\{\Psi_{S^0}^{\mathrm{T}}(t_i) - \hat{\Psi}_{S^0}^{\mathrm{T}}(t_i)\}\theta_{S^0}^*\right\|_2 +$$
$$\left\|\frac{1}{n}\sum_{i=1}^{n}t_i\epsilon_{ji}\right\|_2.$$

Recall that $t \in [0,1]$ and, without loss of generality, let $B \geq 1$. Thus, we can see from the same argument that $\left\|\frac{1}{n}\sum_{i=1}^{n}t_i\left\{Y_{ij} - \hat{\Psi}_{S^0}^{\mathrm{T}}(t_i)\theta_{S^0}^*\right\}\right\|_2 \leq \eta$ holds with the same probability.

□

# C. DETAILS ABOUT DATA GENERATION

In this section, we provide details about the parameters used for generating data in Section 5.1 of the main paper (see Equation 26). Three pairs of variables, $(X_1, X_2), (X_3, X_4), (X_5, X_6)$, are solutions of (26) in the main paper with the following parameters and initial values:

1. $(X_1, X_2)$ are generated according to (26) from the main paper with $\theta_{1,0} = 0$, $\theta_{1,1} = (1.2, 0.3, -0.6)^{\mathrm{T}}$, $\theta_{1,2} = (0.1, 0.2, 0.2)^{\mathrm{T}}$, $\theta_{2,0} = 0.4$, $\theta_{2,1} = (-2, 0, 0.4)^{\mathrm{T}}$, $\theta_{2,2} = (0.5, 0.2, -0.3)^{\mathrm{T}}$, and initial values $X_1(0) = -2, X_2(0) = 2$.

2. $(X_3, X_4)$ are generated according to (26) from the main paper with $\theta_{3,0} = -0.2, \theta_{3,3} = (0,0,0)^{\mathrm{T}}, \theta_{3,4} = (-0.3, 0.4, 0.1)^{\mathrm{T}}, \theta_{4,0} = -0.2, \theta_{4,3} = (0.2, -0.1, -0.2)^{\mathrm{T}}, \theta_{4,4} = (0,0,0)^{\mathrm{T}}$, and initial values $X_3(0) = 2, X_4(0) = -2$.

3. $(X_5, X_6)$ are generated according to (26) from the main paper with $\theta_{5,0} = 0.05$, $\theta_{5,5} = (0,0,0)^{\mathrm{T}}$, $\theta_{5,6} = (0.1, 0, -0.8)^{\mathrm{T}}$, $\theta_{6,0} = -0.05$, $\theta_{6,5} = (0,0,0.5)^{\mathrm{T}}$, $\theta_{6,6} = (0,0,0)^{\mathrm{T}}$, and

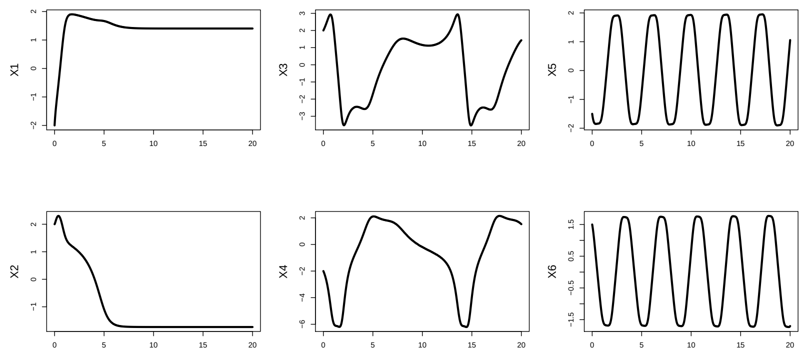


Figure S1: The curves $X_1, \ldots, X_6$ on $[0, 20]$ described in Section 5.1 of the main paper and Section C of the supplementary material.

initial values $X_5(0) = -1.5, X_6(0) = 1.5$.

Solution trajectories of $X_1, \ldots, X_6$ are shown in Figure S1. For $X_7, \ldots, X_{10}$, we drew the initial values $X_j(0), j = 7, \ldots, 10$, and the $\theta_{j,0}, j = 7, \ldots, 10$, from a normal distribution. All other parameters were set to zero, so that $X_7, \ldots, X_{10}$ represent "noise" variables. The directed graph of $X_1, \ldots, X_{10}$ is showing in Figure S2.

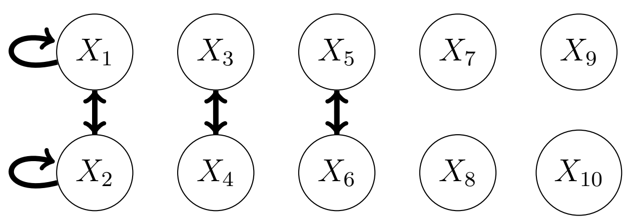


Figure S2: The network of $\{X_1, \ldots, X_{10}\}$. A directed edge $j \to k$ indicates that the $j$th node regulates the $k$th node.

## References

Boucheron, S., G. Lugosi, and P. Massart (2013). *Concentration inequalities*. Oxford University Press, Oxford. A nonasymptotic theory of independence, With a foreword by Michel Ledoux.

Gershgorin, S. (1931). über die abgrenzung der eigenwerte einer matrix. *Bulletin de l'Académie des Sciences de l'URSS. Classe des sciences mathématiques et na* (6), 749–754.

Loh, P.-L. and M. J. Wainwright (2012). High-dimensional regression with noisy and missing data: provable guarantees with nonconvexity. *Ann. Statist. 40*(3), 1637–1664.

Ma, Y. and R. Li (2010). Variable selection in measurement error models. *Bernoulli 16*(1), 274–300.

Rosenbaum, M. and A. B. Tsybakov (2010). Sparse recovery under matrix uncertainty. *Ann. Statist. 38*(5), 2620–2651.

Roth, V. and B. Fischer (2008). The group-lasso for generalized linear models: uniqueness of solutions and efficient algorithms. In *Proceedings of the 25th International Conference on Machine Learning*, ICML '08, New York, NY, USA, pp. 848–855. ACM.

Simon, N. and R. J. Tibshirani (2012). Standardization and the group lasso penalty. *Statist. Sinica 22*(3), 983–1001.

Tsybakov, A. B. (2009). *Introduction to nonparametric estimation*. Springer Series in Statistics. Springer, New York. Revised and extended from the 2004 French original, Translated by Vladimir Zaiats.

Wainwright, M. J. (2009). Sharp thresholds for high-dimensional and noisy sparsity recovery using $\ell_1$-constrained quadratic programming (Lasso). *IEEE Trans. Inform. Theory 55*(5), 2183–2202.

Weyl, H. (1912). Das asymptotische Verteilungsgesetz der Eigenwerte linearer partieller Differentialgleichungen (mit einer Anwendung auf die Theorie der Hohlraumstrahlung). *Math. Ann. 71*(4), 441–479.