Supplementary Material

A Proof of Section 3

A.1 Proof of Corollary 3.4

Proof. Firstly, we derive an error of $\widetilde{\mu}_{\boldsymbol{x}^*}$ when the observed data follow the regression model (5). Namely, we show the following equality with probability at least $1-\delta$,

$$\widetilde{\mu}_{\boldsymbol{x}^*} = \mu_{\boldsymbol{x}^*} + O\left(n^{-1/2}R\right) \pm O\left(\sqrt{\epsilon}L'^2 R\right),$$

with the model. This equality is an analogous of the inequality (15) without the assumption of the regression model (5).

We start with (14) and obtain

$$\widetilde{\mu}_{\boldsymbol{x}^*} = \frac{\langle \boldsymbol{v}^*, \boldsymbol{y}\rangle}{n} + \frac{\langle \boldsymbol{v} - \boldsymbol{v}^*, \boldsymbol{y}\rangle}{n} + \langle \mathfrak{v}, \pi(y_S) - y\rangle. \tag{8}$$

By the model (5), we have $\boldsymbol{y} = \boldsymbol{f} + \boldsymbol{\xi}$ where $\boldsymbol{f} := (f(\boldsymbol{x}_1), \dots, f(\boldsymbol{x}_n))^\top$ and $\boldsymbol{\xi} := (\xi_1, \dots, \xi_n)^\top$, then we obtain

$$\langle \boldsymbol{v} - \boldsymbol{v}^*, \boldsymbol{y}\rangle = \langle \boldsymbol{v} - \boldsymbol{v}^*, \boldsymbol{f}\rangle + \langle \boldsymbol{v} - \boldsymbol{v}^*, \boldsymbol{\xi}\rangle = \langle \Phi(\boldsymbol{v} - \boldsymbol{v}^*), \boldsymbol{\alpha}\rangle_{\mathcal{H}} + \sum_{i\in[n]} \xi_i (v_i - v_i^*).$$

About the second term $\sum_{i\in[n]} \xi_i(v_i - v_i^*)$, we define $\bar{v}_i := (v_i - v_i^*)$, then we have

$$\sum_{i\in[n]} \xi_i (v_i - v_i^*) \sim \mathcal{N}\left(0, \nu^2 \|\bar{\boldsymbol{v}}\|_2^2\right),$$

since $\xi_i \sim \mathcal{N}(0,\nu^2)$ independently and identically. Then, we apply the tail bound for Gaussian random variables and obtain

$$\left|\sum_{i\in[n]} \xi_i (v_i - v_i^*)\right| \le \sqrt{2}\nu \|\bar{\boldsymbol{v}}\|_2 \log^{1/2}(1/\delta),$$

with probability at least $1-\delta$ for any $\delta \in (0,1)$. By definition of $\boldsymbol{v}$, it has the same ℓ_∞ norm of $\boldsymbol{v}^*$, meaning $\|\bar{\boldsymbol{v}}\|_\infty \le 2R$. Since $\|\boldsymbol{u}\|_2 \le \sqrt{n}\|\boldsymbol{u}\|_\infty$ for any $\boldsymbol{u} \in \mathbb{R}^n$, we have $\|\bar{\boldsymbol{v}}\|_2 \le \sqrt{4n}R$, and

$$\left|\sum_{i\in[n]} \xi_i (v_i - v_i^*)\right| \le \sqrt{8n}\nu R \log^{1/2}(1/\delta).$$

Substituting the result into (8), and the Cauchy-Schwartz inequality with Lemma 2.3 as (15) yields

$$\begin{aligned}\widetilde{\mu}_{\boldsymbol{x}^*} &= \mu_{\boldsymbol{x}^*} \pm O\left(\frac{\sqrt{8}\nu R \log^{1/2}(1/\delta)}{\sqrt{n}}\right) \pm O\Big(\frac{\|\Phi(\boldsymbol{v}^* - \boldsymbol{v})\|_{\mathcal{H}} \|\boldsymbol{\alpha}\|_{\mathcal{H}}}{n} + \|\mathfrak{v}\|_\infty \|(\pi(y_S) - y)1^\top\|_\square \|1\|_\infty\Big) \\ &= \mu_{\boldsymbol{x}^*} \pm O\left(\frac{\sqrt{8}\nu R \log^{1/2}(1/\delta)}{\sqrt{n}}\right) \pm O\left(\sqrt{\epsilon}L'^2 R\right).\end{aligned}$$

Substituting $\delta = 0.01$, then we obtain

$$\widetilde{\mu}_{\boldsymbol{x}^*} = \mu_{\boldsymbol{x}^*} + O\left(n^{-1/2}R\right) \pm O\left(\sqrt{\epsilon}L'^2 R\right).$$

When we substitute $\epsilon = O(\log^{1/2} n)$, the second term $O\left(n^{-1/2}R\right)$ is negligible asymptotically in comparison with $O\left(\sqrt{\epsilon}L^2 R\right)$, hence we can ignore the second term as $n \to \infty$. $\square$

B Minimizing the Normalized Loss

In this section, we prove Theorem 3.1.

To show that $\min_{\boldsymbol{v}} \ell_{K,\boldsymbol{k},\lambda}(\boldsymbol{v})$ and $\min_{\widetilde{\boldsymbol{v}}} \ell_{K_{SS},\boldsymbol{k}_S,\lambda}(\widetilde{\boldsymbol{v}})$ are close, we want to say that K and K_{SS} are close in some sense. Here, we measure their distance by the cut norm of their corresponding graphons $\mathcal{K}$ and $\mathcal{K}_{SS}$ in order to exploit Lemma 2.2. In the case of $\boldsymbol{k}$ and $\boldsymbol{k}_S$, we measure their distance by the cut norm of the graphons $\mathscr{k}1^\top$ and $\mathscr{k}_S 1^\top$, where $1\colon [0,1] \to \mathbb{R}$ is a function with $1(x) = 1$.

As we work on graphons, it is useful to define an analog of (2) for graphons:

$$\ell_{\mathcal{K},\mathscr{k},\lambda}(f) = \|\mathcal{K}f - \mathscr{k}\|_2^2 + \lambda\langle f, \mathcal{K}f\rangle.$$

We show that the minima of $\ell_{K,\boldsymbol{k},\lambda}$ and $\ell_{K_{SS},\boldsymbol{k},\lambda}$ are close if $\mathcal{K}$ and $\mathcal{K}_{SS}$ are close in the cut norm up to a measure-preserving bijection and so do $\mathscr{k}1^\top$ and $\mathscr{k}_S 1^\top$.

Lemma B.1. *If a set $S \subseteq [n]$ satisfies*

$$\|\mathcal{K} - \pi(\mathcal{K}_{SS})\|_\square \leq \epsilon L \text{ and } \|\mathscr{k}1^\top - \pi(\mathscr{k}_S)1^\top\|_\square \leq \epsilon L$$

for some measure-preserving bijection $\pi\colon [0,1] \to [0,1]$, then we have

$$\min_{\widetilde{\boldsymbol{v}} \in \mathbb{R}^s} \ell_{K_{SS},\boldsymbol{k}_S,\lambda}(\widetilde{\boldsymbol{v}}) = \min_{\boldsymbol{v} \in \mathbb{R}^n} \ell_{K,\boldsymbol{k},\lambda}(\boldsymbol{v}) \pm O\Big(\epsilon L^2 R^2\Big).$$

Proof of Theorem 3.1. By applying Lemma 2.2 on K and $\boldsymbol{k}\boldsymbol{1}^\top$, we obtain

$$\|\mathcal{K} - \pi(\mathcal{K}_{SS})\|_\square \leq \epsilon L \text{ and } \|(\mathscr{k} - \pi(\mathscr{k}_S))1^\top\|_\square \leq \epsilon L$$

for a measure-preserving bijection $\pi\colon [0,1] \to [0,1]$ with probability at least 0.99. Then, the theorem follows by Lemma B.1. □

B.1 Proof of Lemma B.1

We say that a function $f\colon [0,1] \to \mathbb{R}$ is *n-block constant* if $f(x) = f(x')$ holds whenever $i_n(x) = i_n(x')$. For an n-block constant f, we can find $\boldsymbol{v} \in \mathbb{R}^n$ such that $\ell_{K,\boldsymbol{k},\lambda}(\boldsymbol{v}) = \ell_{\mathcal{K},\mathscr{k},\lambda}(f)$:

Lemma B.2. *Let $f\colon [0,1] \to \mathbb{R}$ be an n-block constant function and let $\boldsymbol{v} \in \mathbb{R}^n$ be a vector so that $v_j = f^*(x)$ for $x \in [0,1]$ with $i_n(x) = j$ (Note that $\boldsymbol{v}$ is uniquely determined). Then, we have*

$$\ell_{K,\boldsymbol{k},\lambda}(\boldsymbol{v}) = \ell_{\mathcal{K},\mathscr{k},\lambda}(f).$$

Proof. Note that we have

$$\ell_{K,\boldsymbol{k},\lambda}(\boldsymbol{v}) = \frac{1}{n^3}\|K\boldsymbol{v}\|_2^2 - \frac{2}{n^2}\langle \boldsymbol{k}, K\boldsymbol{v}\rangle + \frac{1}{n}\|\boldsymbol{k}\|_2^2 + \frac{\lambda}{n^2}\langle \boldsymbol{v}, K\boldsymbol{v}\rangle,$$
$$\ell_{\mathcal{K},\mathscr{k},\lambda}(\boldsymbol{v}) = \|\mathcal{K}f\|_2^2 - 2\langle \mathscr{k}, \mathcal{K}f\rangle + \|\mathscr{k}\|_2^2 + \lambda\langle f, \mathcal{K}f\rangle.$$

We show that each pair of corresponding terms are equal.

For the first pair of terms, we have

$$\|\mathcal{K}f\|_2^2 = \int_0^1 \Big(\int_0^1 \mathcal{K}(x,y)f(y)\mathrm{d}y\Big)^2 \mathrm{d}x = \sum_{i\in[n]} \int_{I_i^n} \Big(\sum_{j\in[n]} \int_{I_j^n} \mathcal{K}(x,y)f(y)\mathrm{d}y\Big)^2 \mathrm{d}x$$
$$= \sum_{i\in[n]} \int_{I_i^n} \Big(\sum_{j\in[n]} \int_{I_j^n} K_{ij}v_j \mathrm{d}y\Big)^2 \mathrm{d}x = \sum_{i\in[n]} \int_{I_i^n} \Big(\frac{1}{n}\sum_{j\in[n]} K_{ij}v_j\Big)^2 \mathrm{d}x$$
$$= \frac{1}{n^3}\sum_{i\in[n]} \Big(\sum_{j\in[n]} K_{ij}v_j\Big)^2 = \frac{1}{n^3}\|K\boldsymbol{v}\|_2^2.$$

For the second pair of terms, we have

$$\begin{aligned}\langle \mathscr{k}, \mathscr{K} f\rangle &= \int_0^1 \mathscr{k}(x)\Big(\int_0^1 \mathscr{K}(x,y)f(y)\mathrm{d}y\Big)\mathrm{d}x \\ &= \sum_{i\in[n]}\int_{I_i^n} \mathscr{k}(x)\Big(\sum_{j\in[n]}\int_{I_j^n}\mathscr{K}(x,y)f(y)\mathrm{d}y\Big)\mathrm{d}x = \sum_{i\in[n]}\int_{I_i^n} y_i\Big(\sum_{j\in[n]}\int_{I_j^n} K_{ij}v_j\mathrm{d}y\Big)\mathrm{d}x \\ &= \sum_{i\in[n]}\int_{I_i^n} y_i\Big(\frac{1}{n}\sum_{j\in[n]}K_{ij}v_j\Big)\mathrm{d}x = \frac{1}{n^2}\sum_{i\in[n]} y_i\Big(\sum_{j\in[n]}K_{ij}v_j\Big) = \frac{1}{n^2}\langle \boldsymbol{k}, K\boldsymbol{v}\rangle.\end{aligned}$$

For the third pair of terms, we have

$$\|\mathscr{k}\|_2^2 = \int_0^1 \mathscr{k}(x)^2\mathrm{d}x = \sum_{i\in[n]}\int_{I_i^n}\mathscr{k}(x)^2\mathrm{d}x = \sum_{i\in[n]}\int_{I_i^n} y_i^2\mathrm{d}x = \frac{1}{n}\sum_{i\in[n]} y_i^2 = \frac{1}{n}\|\boldsymbol{k}\|_2^2.$$

For the fourth pair of terms, we have

$$\begin{aligned}\langle f, \mathscr{K} f\rangle &= \int_0^1\int_0^1 \mathscr{K}(x,y)f(x)f(y)\mathrm{d}x\mathrm{d}y = \sum_{i\in[n]}\sum_{j\in[n]}\int_{I_i^n}\int_{I_j^m}\mathscr{K}(x,y)f(x)f(y)\mathrm{d}x\mathrm{d}y \\ &= \frac{1}{n^2}n\sum_{i\in[n]}\sum_{j\in[n]}K_{ij}v_iv_j = \frac{1}{n^2}\langle \boldsymbol{v}, K\boldsymbol{v}\rangle.\end{aligned}$$

Combining these equalities establishes the claim. ∎

The following lemma states that minimizing $\ell_{K,\boldsymbol{k},\lambda}$ and $\ell_{\mathscr{K},\mathscr{k},\lambda}$ are equivalent:

Lemma B.3. *For any $R\in\mathbb{R}_+$, we have*

$$\min_{\boldsymbol{v}\in\mathbb{R}^n:\|\boldsymbol{v}\|_\infty\le R}\ell_{K,\boldsymbol{k},\lambda}(\boldsymbol{v}) = \min_{f\colon[0,1]\to\mathbb{R}:\|f\|_\infty\le R}\ell_{\mathscr{K},\mathscr{k},\lambda}(f).$$

Proof. First, we show (RHS) $\le$ (LHS). Let $\boldsymbol{v}^*$ be a minimizer of the LHS and let $f\colon[0,1]\to\mathbb{R}$ with $f(x) = v^*_{i_n(x)}$. Note that $\|f\|_\infty = \|\boldsymbol{v}^*\|_\infty$. As f is n-block constant, by Lemma B.2, we have $\ell_{\mathscr{K},\mathscr{k},\lambda}(f) = \ell_{K,\boldsymbol{k},\lambda}(\boldsymbol{v}^*)$.

Next, we show (LHS) $\le$ (RHS). First, note that there exists a minimizer of the RHS because $\ell_{\mathscr{K},\mathscr{k},\lambda}$ is convex. Indeed, we can show that there is an n-block constant minimizer. To this end, let $f^*\colon[0,1]\to\mathbb{R}$ be an arbitrary minimizer of the RHS. For an integer $m\in\mathbb{N}$, let $f_m\colon[0,1]\to\mathbb{R}$ be the $(n\cdot m)$-block constant function, where the value of the i-th block is the average of the values of f^* in the corresponding interval. As the sequence $\{f_m\}_m$ (strongly) converges to f^* in the L^2 norm, $\ell_{\mathscr{K},\mathscr{k},\lambda}(f_m)$ converges to $\ell_{\mathscr{K},\mathscr{k},\lambda}(f^*)$.

Next for each $m\in\mathbb{N}$, we construct an n-block constant function $f'_m\colon[0,1]\to\mathbb{R}$ from f_m. As $\ell_{\mathscr{K},\mathscr{k},\lambda}$ is convex and is invariant under swapping $f_m(x)$ and $f_m(x')$ for any $x,x'\in[0,1]$ with $i_n(x)=i_n(x')$, we can replace $f_m(x)$ and $f_m(x')$ with their average without increasing the value of $\ell_{\mathscr{K},\mathscr{k},\lambda}$. By taking the limit of this process, we can construct an n-block constant function f'_m such that $\ell_{\mathscr{K},\mathscr{k},\lambda}(f'_m)\le\ell_{\mathscr{K},\mathscr{k},\lambda}(f_m)$.

Then, there is a subsequence of $\{f_m\}_m$ in which f'_m converges to an n-block constant function, f', and we replace $\{f_m\}_m$ with this subsequence. Then, we have $\ell_{\mathscr{K},\mathscr{k},\lambda}(f')\le\lim_{m\to\infty}\ell_{\mathscr{K},\mathscr{k},\lambda}(f'_m)\le\lim_{m\to\infty}\ell_{\mathscr{K},\mathscr{k},\lambda}(f_m) = \ell_{\mathscr{K},\mathscr{k},\lambda}(f^*)$. Hence, there is a n-block constant minimizer f'. Now as f' is n-block constant, Lemma B.2 gives a vector $\boldsymbol{v}\in\mathbb{R}^n$ such that $\ell_{K,\boldsymbol{k},\lambda}(\boldsymbol{v}) = \ell_{\mathscr{K},\mathscr{k},\lambda}(f')$. Also, $\|\boldsymbol{v}\|_\infty = \|f^*\|_\infty$. ∎

Proof of Lemma B.1. We have

$$\begin{aligned}\min_{\widetilde{\boldsymbol{v}}\in\mathbb{R}^s}\ell_{K_{SS},\boldsymbol{k}_S,\lambda}(\widetilde{\boldsymbol{v}}) &= \min_{\widetilde{\boldsymbol{v}}\in\mathbb{R}^s:\|\widetilde{\boldsymbol{v}}\|_\infty\le R}\ell_{K_{SS},\boldsymbol{k}_S,\lambda}(\widetilde{\boldsymbol{v}}) = \min_{\substack{f\colon[0,1]\to\mathbb{R}:\\ \|f\|_\infty\le R}}\ell_{\mathscr{K}_{SS},\mathscr{k}_S,\lambda}(f) \qquad \text{(By Lemma B.3)} \\ &= \min_{\substack{f\colon[0,1]\to\mathbb{R}:\\ \|f\|_\infty\le R}}\|\mathscr{K}_{SS}f\|_2^2 - 2\langle\mathscr{k}_S,\mathscr{K}_{SS}f\rangle + \|\mathscr{k}_S\|_2^2 + \lambda\langle f,\mathscr{K}_{SS}f\rangle\end{aligned}$$

$$
\begin{aligned}
&= \min_{\substack{f\colon [0,1]\to\mathbb{R}:\\ \|f\|_\infty \le R}} \|\pi(\mathcal{K}_{SS})f\|_2^2 - 2\langle \pi(\mathscr{k}_S), \pi(\mathcal{K}_{SS})f\rangle + \|\pi(\mathscr{k}_S)\|_2^2 + \lambda\langle f, \pi(\mathcal{K}_{SS})f\rangle \\
&= \min_{\substack{f\colon [0,1]\to\mathbb{R}:\\ \|f\|_\infty \le R}} \left\|\left(\pi(\mathcal{K}_{SS}) - \mathcal{K} + \mathcal{K}\right)f\right\|_2^2 - 2\left\langle \pi(\mathscr{k}_S) - \mathscr{k} + \mathscr{k}, \left(\pi(\mathcal{K}_{SS}) - \mathcal{K} + \mathcal{K}\right)f\right\rangle \\
&\qquad + \|\pi(\mathscr{k}_S) - \mathscr{k} + \mathscr{k}\|_2^2 + \lambda\left\langle f, \left(\pi(\mathcal{K}_{SS}) - \mathcal{K} + \mathcal{K}\right)f\right\rangle \\
&= \min_{\substack{f\colon [0,1]\to\mathbb{R}:\\ \|f\|_\infty \le R}} \|\mathcal{K}f\|_2^2 + 2\left\langle \left(\pi(\mathcal{K}_{SS}) - \mathcal{K}\right)f, \mathcal{K}f\right\rangle + \left\|\left(\pi(\mathcal{K}_{SS}) - \mathcal{K}\right)f\right\|_2^2 \\
&\qquad - 2\langle \mathscr{k}, \mathcal{K}f\rangle - 2\left\langle \mathscr{k}, \left(\pi(\mathcal{K}_{SS}) - \mathcal{K}\right)f\right\rangle - 2\langle \pi(\mathscr{k}_S) - \mathscr{k}, \mathcal{K}f\rangle \\
&\qquad - 2\left\langle \pi(\mathscr{k}_S) - \mathscr{k}, \left(\pi(\mathcal{K}_{SS}) - \mathcal{K}\right)f\right\rangle + \|\mathscr{k}\|_2^2 + 2\langle \pi(\mathscr{k}_S) - \mathscr{k}, \mathscr{k}\rangle + \left\|\pi(\mathscr{k}_S) - \mathscr{k}\right\|_2^2 \\
&\qquad + \lambda\langle f, \mathcal{K}f\rangle + \lambda\left\langle f, \left(\pi(\mathcal{K}_{SS}) - \mathcal{K}\right)f\right\rangle. \qquad (9)
\end{aligned}
$$

By Lemma 2.3 and using the fact that $\pi(\mathscr{k}_S) - \mathscr{k} = (\pi(\mathscr{k}_S) - \mathscr{k})1^\top 1$, we have

$$
\begin{aligned}
(9) &= \min_{\substack{f\colon [0,1]\to\mathbb{R}:\\ \|f\|_\infty \le R}} \|\mathcal{K}f\|_2^2 - 2\langle \mathscr{k}, \mathcal{K}f\rangle + \|\mathscr{k}\|_2^2 + \lambda\langle f, \mathcal{K}f\rangle \\
&\qquad \pm \Big(2\left\|\pi(\mathcal{K}_{SS}) - \mathcal{K}\right\|_\square \|\mathcal{K}\|_\square \|f\|_\infty^2 + \left\|\pi(\mathcal{K}_{SS}) - \mathcal{K}\right\|_\square^2 \|f\|_\infty^2 \\
&\qquad + 2\left\|\pi(\mathcal{K}_{SS}) - \mathcal{K}\right\|_\square \|\mathscr{k}\|_\infty \|f\|_\infty + 2\|\mathcal{K}\|_\square \left\|(\pi(\mathscr{k}_S) - \mathscr{k})1^\top\right\|_\square \|1\|_\infty \|f\|_\infty \\
&\qquad + 2\left\|\pi(\mathcal{K}_{SS}) - \mathcal{K}\right\|_\square \left\|(\pi(\mathscr{k}_S) - \mathscr{k})1^\top\right\|_\square \|1\|_\infty \|f\|_\infty \\
&\qquad + 2\left\|(\pi(\mathscr{k}_S) - \mathscr{k})1^\top\right\|_\square \|1\|_\infty \|\mathscr{k}\|_\infty + \left\|\pi(\mathscr{k}_S) - \mathscr{k}\right\|_2^2 + \lambda\left\|\pi(\mathcal{K}_{SS}) - \mathcal{K}\right\|_\square \|f\|_\infty^2\Big). \qquad (10)
\end{aligned}
$$

From the assumption, we have

$$
\begin{aligned}
&= \min_{f\colon [0,1]\to\mathbb{R}: \|f\|_\infty \le R} \|\mathcal{K}f\|_2^2 - 2\langle \mathscr{k}, \mathcal{K}f\rangle + \|\mathscr{k}\|_2^2 + \lambda\langle f, \mathcal{K}f\rangle \\
&\qquad \pm \left(2\epsilon L^2R^2 + \epsilon^2L^2R^2 + 2\epsilon L^2R + 2\epsilon L^2R + 2\epsilon^2L^2R + 2\epsilon L^2 + \epsilon^2L^2 + \lambda\epsilon LR^2\right) \\
&= \min_{\boldsymbol{v}\in\mathbb{R}^n: \|\boldsymbol{v}\|_\infty \le R} \frac{1}{n^3}\|K\boldsymbol{v}\|_2^2 - \frac{2}{n^2}\langle \boldsymbol{k}, K\boldsymbol{v}\rangle + \frac{1}{n}\|\boldsymbol{k}\|_2^2 + \frac{\lambda}{n^2}\langle \boldsymbol{v}, \mathcal{K}\boldsymbol{v}\rangle \pm O\left(\epsilon L^2R^2\right) \qquad \text{(By Lemma B.3)} \\
&= \min_{\boldsymbol{v}\in\mathbb{R}^n: \|\boldsymbol{v}\|_\infty \le R} \ell_{K,\boldsymbol{k},\lambda}(\boldsymbol{v}) \pm O\left(\epsilon L^2R^2\right) = \min_{\boldsymbol{v}\in\mathbb{R}^n} \ell_{K,\boldsymbol{k},\lambda}(\boldsymbol{v}) \pm O\left(\epsilon L^2R^2\right)
\end{aligned}
$$

as desired. □

C Proof of Theorem 3.2

The following lemma is a modification of Lemma B.1 for relating the solution of $\ell_{K_{SS},\boldsymbol{k}_S,\lambda}$ and that of $\ell_{K,\boldsymbol{k},\lambda}$ using a given measure-preserving bijection.

Lemma C.1. *If a set $S \subseteq [n]$ satisfies*

$$\|\mathcal{K} - \pi(\mathcal{K}_{SS})\|_\square \le \epsilon L \text{ and } \|\mathscr{k}1^\top - \pi(\mathscr{k}_S)1^\top\|_\square \le \epsilon L$$

for a measure-preserving bijection $\pi\colon [0,1] \to [0,1]$, then for any $\widetilde{\boldsymbol{v}} \in \mathbb{R}^s$ with $\|\widetilde{\boldsymbol{v}}\|_\infty \le R$, there exists $\boldsymbol{v} \in \mathbb{R}^n$ such that

$$\ell_{K_{SS},\boldsymbol{k}_S,\lambda}(\widetilde{\boldsymbol{v}}) = \ell_{K,\boldsymbol{k},\lambda}(\boldsymbol{v}) \pm O\left(\epsilon L^2R^2\right) \quad \text{and} \quad \pi(\widetilde{\mathscr{v}}) = \mathscr{v}.$$

Proof. Let $\widetilde{\mathscr{v}}\colon [0,1] \to \mathbb{R}$ be the function corresponding to $\widetilde{\boldsymbol{v}}$, that is, $\widetilde{\mathscr{v}}(x) = \widetilde{v}_{i_n(x)}$. Then, we have

$$\ell_{K_{SS},\boldsymbol{k}_S,\lambda}(\widetilde{\boldsymbol{v}}) = \frac{1}{s^3}\|K_{SS}\widetilde{\boldsymbol{v}}\|_2^2 - \frac{2}{s^2}\langle \boldsymbol{k}_S, K_{SS}\widetilde{\boldsymbol{v}}\rangle + \frac{1}{s}\|\boldsymbol{k}_S\|_2^2 + \frac{\lambda}{n^2}\langle \widetilde{\boldsymbol{v}}, K_{SS}\widetilde{\boldsymbol{v}}\rangle$$

$$
\begin{aligned}
&= \|\mathcal{K}_{SS}\widetilde{\mathfrak{v}}\|_2^2 - 2\langle \mathscr{k}_S, \mathcal{K}_{SS}\widetilde{\mathfrak{v}}\rangle + \|\mathscr{k}_S\|_2^2 + \lambda\langle \widetilde{\mathfrak{v}}, \mathcal{K}_{SS}\widetilde{\mathfrak{v}}\rangle && \text{(By Lemma B.2)} \\
&= \|\pi(\mathcal{K}_{SS})\pi(\widetilde{\mathfrak{v}})\|_2^2 - 2\langle \pi(\mathscr{k}_S), \pi(\mathcal{K}_{SS})\pi(\widetilde{\mathfrak{v}})\rangle + \|\pi(\mathscr{k}_S)\|_2^2 + \lambda\langle \pi(\widetilde{\mathfrak{v}}), \pi(\mathcal{K}_{SS})\pi(\widetilde{\mathfrak{v}})\rangle \\
&= (\pi(\mathcal{K}_{SS}) - \mathcal{K} + \mathcal{K})\pi(\widetilde{\mathfrak{v}})\|_2^2 - 2\langle \pi(\mathscr{k}_S) - \mathscr{k} + \mathscr{k}, (\pi(\mathcal{K}_{SS}) - \mathcal{K} + \mathcal{K})\pi(\widetilde{\mathfrak{v}})\rangle \\
&\quad + \|\pi(\mathscr{k}_S) - \mathscr{k} + \mathscr{k}\|_2^2 + \lambda\langle \pi(\widetilde{\mathfrak{v}}), (\pi(\mathcal{K}_{SS}) - \mathcal{K} + \mathcal{K})\pi(\widetilde{\mathfrak{v}})\rangle \\
&= \|\mathcal{K}\pi(\widetilde{\mathfrak{v}})\|_2^2 + 2\langle (\pi(\mathcal{K}_{SS}) - \mathcal{K})\pi(\widetilde{\mathfrak{v}}), \mathcal{K}\pi(\widetilde{\mathfrak{v}})\rangle + \|(\pi(\mathcal{K}_{SS}) - \mathcal{K})\pi(\widetilde{\mathfrak{v}})\|_2^2 \\
&\quad - 2\langle \mathscr{k}, \mathcal{K}\pi(\widetilde{\mathfrak{v}})\rangle - 2\langle \mathscr{k}, (\pi(\mathcal{K}_{SS}) - \mathcal{K})\pi(\widetilde{\mathfrak{v}})\rangle - 2\langle \pi(\mathscr{k}_S) - \mathscr{k}, \mathcal{K}\pi(\widetilde{\mathfrak{v}})\rangle - 2\langle \pi(\mathscr{k}_S) - \mathscr{k}, (\pi(\mathcal{K}_{SS}) - \mathcal{K})\pi(\widetilde{\mathfrak{v}})\rangle \\
&\quad + \|\mathscr{k}\|_2^2 + 2\langle \pi(\mathscr{k}_S) - \mathscr{k}, \mathscr{k}\rangle + \|\pi(\mathscr{k}_S) - \mathscr{k}\|_2^2 \\
&\quad + \lambda\langle \pi(\widetilde{\mathfrak{v}}), \mathcal{K}\pi(\widetilde{\mathfrak{v}})\rangle + \lambda\langle \pi(\widetilde{\mathfrak{v}}), (\pi(\mathcal{K}_{SS}) - \mathcal{K})\pi(\widetilde{\mathfrak{v}})\rangle. && (11)
\end{aligned}
$$

By Lemma 2.3 and using the assumption that $\pi(\mathscr{k}_S) - \mathscr{k} = (\pi(\mathscr{k}_S) - \mathscr{k})1^\top 1$, we have

$$
\begin{aligned}
(11) &= \|\mathcal{K}\pi(\widetilde{\mathfrak{v}})\|_2^2 - 2\langle \mathscr{k}, \mathcal{K}\pi(\widetilde{\mathfrak{v}})\rangle + \|\mathscr{k}\|_2^2 + \lambda\langle \pi(\widetilde{\mathfrak{v}}), \mathcal{K}\pi(\widetilde{\mathfrak{v}})\rangle \\
&\pm \Big(2\|\pi(\mathcal{K}_{SS}) - \mathcal{K}\|_\square \|\mathcal{K}\|_\square \|\pi(\widetilde{\mathfrak{v}})\|_\infty^2 + \|\pi(\mathcal{K}_{SS}) - \mathcal{K}\|_\square^2 \|\pi(\widetilde{\mathfrak{v}})\|_\infty^2 \\
&\qquad + 2\|\pi(\mathcal{K}_{SS}) - \mathcal{K}\|_\square \|\mathscr{k}\|_\infty \|\pi(\widetilde{\mathfrak{v}})\|_\infty + 2\|\mathcal{K}\|_\square \|(\pi(\mathscr{k}_S) - \mathscr{k})1^\top\|_\square \|1\|_\infty \|\pi(\widetilde{\mathfrak{v}})\|_\infty \\
&\qquad + 2\|\pi(\mathcal{K}_{SS}) - \mathcal{K}\|_\square \|(\pi(\mathscr{k}_S) - \mathscr{k})1^\top\|_\square \|1\|_\infty \|\pi(\widetilde{\mathfrak{v}})\|_\infty + 2\|(\pi(\mathscr{k}_S) - \mathscr{k})1^\top\|_\square \|1\|_\infty \|\mathscr{k}\|_\infty \\
&\qquad + \|\pi(\mathscr{k}) - \mathscr{k}\|_2^2 + \lambda\|\pi(\mathcal{K}_{SS}) - \mathcal{K}\|_2 \|\pi(\widetilde{\mathfrak{v}})\|_2^2\Big). && (12)
\end{aligned}
$$

Recall that π satisfies $i_n(\pi(x)) = i_n(\pi(y))$ whenever $i_n(x) = i_n(y)$. Then, $\pi(\widetilde{\mathfrak{v}})$ is n-block constant, and hence we can define a vector $\boldsymbol{v} \in \mathbb{R}^n$ corresponding to $\pi(\widetilde{\mathfrak{v}})$, that is, $v_i = \pi(\widetilde{\mathfrak{v}})(x)$ for any $x \in [0,1]$ with $i_n(x)$. Then, we have

$$
\begin{aligned}
(12) &= \|\mathcal{K}\pi(\widetilde{\mathfrak{v}})\|_2^2 - 2\langle \mathscr{k}, \mathcal{K}\pi(\widetilde{\mathfrak{v}})\rangle + \|\mathscr{k}\|_2^2 + \lambda\langle \pi(\widetilde{\mathfrak{v}}), \mathcal{K}\pi(\widetilde{\mathfrak{v}})\rangle \\
&\quad \pm \Big(2\epsilon L^2R^2 + \epsilon^2L^2R^2 + 2\epsilon L^2R + 2\epsilon L^2R + 2\epsilon^2L^2R + 2\epsilon L^2 + \epsilon^2L^2 + \lambda\epsilon LR^2\Big) \\
&= \frac{1}{n^3}\|K\boldsymbol{v}\|_2^2 - \frac{2}{n^2}\langle \boldsymbol{k}, K\boldsymbol{v}\rangle + \frac{1}{n}\|\boldsymbol{k}\|_2^2 + \frac{\lambda}{n^2}\langle \boldsymbol{v}, \mathcal{K}\boldsymbol{v}\rangle \pm O\Big(\epsilon L^2R^2\Big) && \text{(By Lemma B.3)} \\
&= \ell_{K,\boldsymbol{k},\lambda}(\boldsymbol{v}) \pm O\Big(\epsilon L^2R^2\Big)
\end{aligned}
$$

as desired. □

The following lemma states that, if $\boldsymbol{v} + \boldsymbol{\Delta}$ and $\boldsymbol{v}$ have similar normalized losses, then $\Phi(\boldsymbol{\Delta})$ must be small in $\mathcal{H}$-norm.

Lemma C.2. *For any vectors $\boldsymbol{v}$, and $\boldsymbol{\Delta} \in \mathbb{R}^n$, we have*

$$
\|\Phi(\boldsymbol{\Delta})\|_{\mathcal{H}} = O\left(n\sqrt{\frac{\ell_{K,\boldsymbol{k},\lambda}(\boldsymbol{v}+\boldsymbol{\Delta}) - \ell_{K,\boldsymbol{k},\lambda}(\boldsymbol{v})}{\lambda}}\right).
$$

Proof. Recall that

$$
\ell_{K,\boldsymbol{k},\lambda}(\boldsymbol{v}) = \frac{1}{n^3}\|K\boldsymbol{v}\|_2^2 - \frac{2}{n^2}\langle \boldsymbol{k}, K\boldsymbol{v}\rangle + \frac{1}{n}\|\boldsymbol{k}\|_2^2 + \frac{\lambda}{n^2}\langle \boldsymbol{v}, K\boldsymbol{v}\rangle.
$$

Then, we have

$$
\begin{aligned}
&\ell_{K,\boldsymbol{k},\lambda}(\boldsymbol{v}+\boldsymbol{\Delta}) - \ell_{K,\boldsymbol{k},\lambda}(\boldsymbol{v}) \\
&= \frac{1}{n^3}\|K(\boldsymbol{v}+\boldsymbol{\Delta})\|_2^2 - \frac{1}{n^3}\|K\boldsymbol{v}\|_2^2 - \frac{2}{n^2}\langle \boldsymbol{k}, K(\boldsymbol{v}+\boldsymbol{\Delta})\rangle + \frac{2}{n^2}\langle \boldsymbol{k}, K\boldsymbol{v}\rangle \\
&\quad + \frac{\lambda}{n^2}\langle (\boldsymbol{v}+\boldsymbol{\Delta}), K(\boldsymbol{v}+\boldsymbol{\Delta})\rangle - \frac{\lambda}{n^2}\langle \boldsymbol{v}, K\boldsymbol{v}\rangle \\
&= \frac{1}{n^3}\Big(2\langle K\boldsymbol{v}, K\boldsymbol{\Delta}\rangle + \|K\boldsymbol{\Delta}\|_2^2\Big) - \frac{2}{n^2}\langle \boldsymbol{k}, K\boldsymbol{\Delta}\rangle + \frac{\lambda}{n^2}\Big(2\langle \boldsymbol{\Delta}, K(\boldsymbol{v}+\boldsymbol{\Delta})\rangle + \langle \boldsymbol{\Delta}, K\boldsymbol{\Delta}\rangle\Big) \\
&= \frac{1}{n^3}\Big(2\langle K\boldsymbol{v}, K\boldsymbol{\Delta}\rangle + \|K\boldsymbol{\Delta}\|_2^2\Big) - \frac{2}{n^2}\langle \boldsymbol{k}, K\boldsymbol{\Delta}\rangle + \frac{\lambda}{n^2}\Big(2\langle \boldsymbol{v}, K\boldsymbol{\Delta}\rangle + 3\langle \boldsymbol{\Delta}, K\boldsymbol{\Delta}\rangle\Big). && (13)
\end{aligned}
$$

Let $\lambda_1 \leq Ln$ be the largest eigenvalue of K. Let $U\Sigma V^\top$ be the SVD of Φ, where $U \in \mathbb{R}^{p\times p}$, $\Sigma = \mathrm{diag}(\sigma_1, \ldots, \sigma_p)$ for $\sigma_1 \geq \cdots \geq \sigma_p$, and $V \colon \mathcal{H} \to \mathbb{R}^p$. As $K_{ij} = \langle \phi_{\boldsymbol{x}_i}, \phi_{\boldsymbol{x}_j} \rangle_{\mathcal{H}}$, we have $K = U\Sigma^2 U^\top$ and hence $\sigma_1 = \lambda_1^{1/2}$ By Cauchy-Schwarz, we have

$$\begin{aligned}(13) &\geq -\frac{2}{n^3}\|\Phi(K\boldsymbol{v})\|_{\mathcal{H}}\|\Phi(\boldsymbol{\Delta})\|_{\mathcal{H}} - \frac{2}{n^2}\|\Phi(\boldsymbol{k})\|_{\mathcal{H}}\|\Phi(\boldsymbol{\Delta})\|_{\mathcal{H}} + \frac{\lambda}{n^2}\Big(3\|\Phi(\boldsymbol{\Delta})\|_{\mathcal{H}}^2 - 2\|\Phi(\boldsymbol{v})\|_{\mathcal{H}}\|\Phi(\boldsymbol{\Delta})\|_{\mathcal{H}}\Big)\\
&\geq -\frac{2}{n^{5/2}}\lambda_{\max}^{3/2}R\|\Phi(\boldsymbol{\Delta})\|_{\mathcal{H}} - \frac{2\lambda_{\max}^{1/2}L}{n^{3/2}}\|\Phi(\boldsymbol{\Delta})\|_{\mathcal{H}} + \frac{\lambda}{n^2}\Big(3\|\Phi(\boldsymbol{\Delta})\|_{\mathcal{H}}^2 - 2\lambda_{\max}^{1/2}Rn^{1/2}\|\Phi(\boldsymbol{\Delta})\|_{\mathcal{H}}\Big)\\
&\geq \frac{3\lambda}{n^2}\|\Phi(\boldsymbol{\Delta})\|_{\mathcal{H}}^2 - \frac{2(L^{3/2}R + L^{3/2} + \lambda L^{1/2}R)}{n}\|\Phi(\boldsymbol{\Delta})\|_{\mathcal{H}}.\end{aligned}$$

Then for

$$a = \frac{3\lambda}{n^2} \quad \text{and} \quad b = \frac{2(L^{3/2}R + L^{3/2} + \lambda L^{1/2}R)}{n},$$

we have

$$\begin{aligned}\|\Phi(\boldsymbol{\Delta})\|_{\mathcal{H}} &\leq \frac{b - \sqrt{b^2 - 4a(\ell_{K,\boldsymbol{k},\lambda}(\boldsymbol{v}+\boldsymbol{\Delta}) - \ell_{K,\boldsymbol{k},\lambda}(\boldsymbol{v}))}}{2a} \leq \sqrt{\frac{\ell_{K,\boldsymbol{k},\lambda}(\boldsymbol{v}+\boldsymbol{\Delta}) - \ell_{K,\boldsymbol{k},\lambda}(\boldsymbol{v})}{a}}\\
&\leq n\sqrt{\frac{\ell_{K,\boldsymbol{k},\lambda}(\boldsymbol{v}+\boldsymbol{\Delta}) - \ell_{K,\boldsymbol{k},\lambda}(\boldsymbol{v})}{3\lambda}} = O\left(n\sqrt{\frac{\ell_{K,\boldsymbol{k},\lambda}(\boldsymbol{v}+\boldsymbol{\Delta}) - \ell_{K,\boldsymbol{k},\lambda}(\boldsymbol{v})}{\lambda}}\right)\end{aligned}$$

as desired. □

The lemma also holds for $\ell_{K_{SS},\boldsymbol{k}_S,\lambda}$ and $\ell_{\mathcal{K}_{SS},\mathscr{k}_S,\lambda}$.

Proof of Theorem 3.2. On applying Lemma 2.2 to K, $\boldsymbol{k}\boldsymbol{1}^\top$, and $\boldsymbol{y}\boldsymbol{1}^\top$, we have

$$\|\mathcal{K} - \pi(\mathcal{K}_{SS})\|_\square \leq \epsilon L, \|\mathscr{k} - \pi(\mathscr{k}_S)\|_\square \leq \epsilon L,$$

and

$$\|y - \pi(y_S)\|_\square \leq \epsilon L,$$

which holds for a some measure-preserving bijection $\pi\colon [0,1] \to [0,1]$ with a probability of at least 0.99. In what follows, we assume that this has happened.

Let $\widetilde{\boldsymbol{v}}^* \in \mathbb{R}^s$ be the minimizer of $\ell_{K_{SS},\boldsymbol{k}_S,\lambda}$ that is returned by Algorithm 1, and let $\boldsymbol{v} \in \mathbb{R}^n$ be the vector given by Lemma C.1 on $\widetilde{\boldsymbol{v}}^*$. Then, we have

$$\begin{aligned}\ell_{K,\boldsymbol{k},\lambda}(\boldsymbol{v}) &= \ell_{K_{SS},\boldsymbol{k}_S,\lambda}(\widetilde{\boldsymbol{v}}^*) + O\Big(\epsilon L^2R^2\Big)\\
&= \ell_{K,\boldsymbol{k},\lambda}(\boldsymbol{v}^*) + O\Big(\epsilon L^2R^2\Big).\end{aligned}$$

This means that $\|\Phi(\boldsymbol{v} - \boldsymbol{v}^*)\|_{\mathcal{H}} = O\big(\sqrt{\epsilon}LRn\big)$ by Lemma C.2. Let π be the measure-preserving bijection given by Lemma C.1. Then, we have

$$\begin{aligned}\widetilde{\mu}_{\boldsymbol{x}^*} &= \frac{\langle \widetilde{\boldsymbol{v}}, \boldsymbol{y}_S\rangle}{s} = \langle \widetilde{\mathscr{v}}, y_S\rangle = \langle \pi(\widetilde{\mathscr{v}}), \pi(y_S)\rangle = \langle \mathscr{v}, \pi(y_S)\rangle\\
&= \langle \mathscr{v}, y\rangle + \langle \mathscr{v}, \pi(y_S) - y\rangle = \frac{\langle \boldsymbol{v}, \boldsymbol{y}\rangle}{n} + \langle \mathscr{v}, \pi(y_S) - y\rangle\\
&= \frac{\langle \boldsymbol{v}^*, \boldsymbol{y}\rangle}{n} + \frac{\langle \boldsymbol{v} - \boldsymbol{v}^*, \boldsymbol{y}\rangle}{n} + \langle \mathscr{v}, \pi(y_S) - y\rangle\\
&= \mu_{\boldsymbol{x}^*} + \frac{\langle \Phi(\boldsymbol{v} - \boldsymbol{v}^*), \boldsymbol{\alpha}\rangle_{\mathcal{H}}}{n} + \langle \mathscr{v}, \pi(y_S) - y\rangle. \qquad (14)\end{aligned}$$

By Cauchy-Schwarz and Lemma 2.3, we have

$$(14) = \mu_{\boldsymbol{x}^*} \pm \Big(\frac{\|\Phi(\boldsymbol{v}^* - \boldsymbol{v})\|_{\mathcal{H}}\|\boldsymbol{\alpha}\|_{\mathcal{H}}}{n}$$

$$
\begin{aligned}
&\quad + \|\mathfrak{v}\|_\infty \|(\pi(y_S) - y)1^\top\|_\square \|1\|_\infty \Big) \\
&= \mu_{\boldsymbol{x}^*} \pm O\left(\sqrt{\epsilon} L^2 R\right). \qquad (15)
\end{aligned}
$$

Similarly, we have

$$
\begin{aligned}
\widetilde{\sigma}^2_{\boldsymbol{x}^*} &= k(\boldsymbol{x}^*, \boldsymbol{x}^*) - \frac{\langle \widetilde{\boldsymbol{v}}, \boldsymbol{k}_S \rangle}{s} = k(\boldsymbol{x}^*, \boldsymbol{x}^*) - \frac{\langle \widetilde{\mathfrak{v}}, \mathfrak{k}_S \rangle}{s} \\
&= k(\boldsymbol{x}^*, \boldsymbol{x}^*) - \langle \pi(\widetilde{\mathfrak{v}}), \pi(\mathfrak{k}_S) \rangle \\
&= k(\boldsymbol{x}^*, \boldsymbol{x}^*) - \langle \mathfrak{v}, \pi(\mathfrak{k}_S) \rangle \\
&= k(\boldsymbol{x}^*, \boldsymbol{x}^*) - \langle \mathfrak{v}, \mathfrak{k} \rangle - \langle \mathfrak{v}, \pi(\mathfrak{k}_S) - \mathfrak{k} \rangle \\
&= k(\boldsymbol{x}^*, \boldsymbol{x}^*) - \frac{\langle \boldsymbol{v}, \boldsymbol{k} \rangle}{n} - \langle \mathfrak{v}, \pi(\mathfrak{k}_S) - \mathfrak{k} \rangle \\
&= k(\boldsymbol{x}^*, \boldsymbol{x}^*) - \frac{\langle \boldsymbol{v}^*, \boldsymbol{k} \rangle}{n} - \frac{\langle \boldsymbol{v} - \boldsymbol{v}^*, \boldsymbol{k} \rangle}{n} - \langle \mathfrak{v}, \pi(\mathfrak{k}_S) - \mathfrak{k} \rangle \\
&= \sigma^2_{\boldsymbol{x}^*} - \frac{\langle \Phi(\boldsymbol{v} - \boldsymbol{v}^*), \phi_{\boldsymbol{x}^*} \rangle_{\mathcal{H}}}{n} - \langle \mathfrak{v}, \pi(\mathfrak{k}_S) - \mathfrak{k} \rangle. \qquad (16)
\end{aligned}
$$

By Cauchy-Schwarz and Lemma 2.3, we have

$$
\begin{aligned}
(16) &= \sigma^2_{\boldsymbol{x}^*} \pm \Bigg(\frac{\|\Phi(\boldsymbol{v} - \boldsymbol{v}^*)\|_{\mathcal{H}} \|\phi_{\boldsymbol{x}^*}\|_{\mathcal{H}}}{n} \\
&\quad + \|\mathfrak{v}\|_\infty \|(\pi(\mathfrak{k}_S) - \mathfrak{k})1^\top\|_\square \|1\|_\infty \Bigg) \\
&= \sigma^2_{\boldsymbol{x}^*} \pm O\left(\sqrt{\epsilon} L^2 R\right).
\end{aligned}
$$

□

D Proofs of Section 5

D.1 Proof of Theorem 5.1

Proof. We evaluate the difference between the cross-validated loss values as

$$
\begin{aligned}
&\mathrm{CV}(\theta_1) - \mathrm{CV}(\theta_2) \\
&= \frac{1}{q} \sum_{i \in Q} \left(y_i - \widehat{f}_{S,\theta_1}(\boldsymbol{x}_i)\right)^2 - \left(y_i - \widehat{f}_{S,\theta_2}(\boldsymbol{x}_i)\right)^2 \\
&= \frac{1}{q} \sum_{i \in Q} \left(f^*(\boldsymbol{x}_i) - \widehat{f}_{S,\theta_1}(\boldsymbol{x}_i)\right)^2 - \left(f^*(\boldsymbol{x}_i) - \widehat{f}_{S,\theta_2}(\boldsymbol{x}_i)\right)^2 \\
&\quad - 2\epsilon_i (f^*(\boldsymbol{x}_i) - \widehat{f}_{S,\theta_1}(\boldsymbol{x}_i)) + 2\epsilon_i (f^*(\boldsymbol{x}_i) - \widehat{f}_{S,\theta_2}(\boldsymbol{x}_i)) \\
&= \frac{1}{q} \sum_{i \in Q} \left(f^*(\boldsymbol{x}_i) - f^0_{S,\theta_1}(\boldsymbol{x}_i) - \omega_1(s)\right)^2 - \left(f^*(\boldsymbol{x}_i) - f^0_{S,\theta_2}(\boldsymbol{x}_i) - \omega_2(s)\right)^2 \\
&\quad - 2\epsilon_i (f^*(\boldsymbol{x}_i) - f^0_{S,\theta_1}(\boldsymbol{x}_i) + \omega_1(s) - f^*(\boldsymbol{x}_i) - f^0_{S,\theta_2}(\boldsymbol{x}_i) - \omega_2(s)) \\
&= \frac{1}{q} \sum_{i \in Q} \left(f^*(\boldsymbol{x}_i) - f^0_{S,\theta_1}(\boldsymbol{x}_i)\right)^2 - \omega_1(s)(f^*(\boldsymbol{x}_i) - f^0_{S,\theta_1}(\boldsymbol{x}_i)) + \omega_1(s)^2 \\
&\quad - \left(f^*(\boldsymbol{x}_i) - f^0_{S,\theta_2}(\boldsymbol{x}_i)\right)^2 + \omega_2(s)(f^*(\boldsymbol{x}_i) - f^0_{S,\theta_2}(\boldsymbol{x}_i)) - \omega_2(s)^2 \\
&\quad - 2\epsilon_i (f^*(\boldsymbol{x}_i) - f^0_{S,\theta_1}(\boldsymbol{x}_i) - f^*(\boldsymbol{x}_i) + f^0_{S,\theta_2}(\boldsymbol{x}_i)) - 2\epsilon_i (\omega_1(s) - \omega_2(s)).
\end{aligned}
$$

Here for $\ell = 1, 2$, by the Bernstein's inequality, we have

$$
\Pr\left(\left| \frac{1}{q} \sum_{i \in Q} \left(f^*(\boldsymbol{x}_i) - f^0_{S,\theta_\ell}(\boldsymbol{x}_i)\right)^2 - \mathrm{EL}(\theta_\ell) \right| \leq t_\ell \right)
$$

$$\geq 1 - 2\exp\left(-\frac{1}{2}\frac{t_\ell^2}{B_\sigma^2/q + 2B^2 t_\ell/(3q)}\right) =: 1 - p_\ell(t_\ell, q),$$

for any $t_\ell > 0$. Also, the Chebyshev's inequality yields

$$\Pr\left(\left|\nu^2 - \frac{1}{q}\sum_{i\in Q}\epsilon_i^2\right| \leq t\right) \geq 1 - \frac{\nu^2}{qt} =: 1 - p_\nu(t, q),$$

for all $t > 0$. Then, with probability $1 - p_1(t_1, q) - p_2(t_2, q) - p_\nu(t_3, q)$, we obtain

$$\begin{aligned}
&\mathrm{CV}(\theta_1) - \mathrm{CV}(\theta_2)\\
&\leq \mathrm{EL}(\theta_1) + t_1 - \mathrm{EL}(\theta_2) + t_2\\
&\quad + \frac{1}{q}\sum_{i\in Q} -\omega_1(s)(f^*(\boldsymbol{x}_i) - f^0_{S,\theta_1}(\boldsymbol{x}_i)) + \omega_1(s)^2 + \omega_2(s)(f^*(\boldsymbol{x}_i) - f^0_{S,\theta_2}(\boldsymbol{x}_i)) - \omega_2(s)^2\\
&\quad - 2\epsilon_i(f^*(\boldsymbol{x}_i) - f^0_{S,\theta_1}(\boldsymbol{x}_i) - f^*(\boldsymbol{x}_i) + f^0_{S,\theta_2}(\boldsymbol{x}_i)) - 2\epsilon_i(\omega_1(s) - \omega_2(s))\\
&\leq -\Xi + t_1 + t_2 + 3\omega(s)^2 + 2\omega(s)B + \nu^2(4B + \omega(s)) + t_3(4B + 2\omega(s)),
\end{aligned}$$

by applying the Cauchy-Schwarz inequality and $\omega(s) = \omega_1(s) \vee \omega_2(s)$.

Then, we can state that

$$\mathrm{CV}(\theta_1) \leq \mathrm{CV}(\theta_2),$$

when the following holds;

$$t_1 + t_2 + t_3(4B + 2\omega) \leq \Xi - 3\omega(s)^2 - 2\omega(s)B - \nu^2(4B + \omega) =: \widetilde{\Xi}(s).$$

We set $t_1 = t_2 = t_3(4B + 2\omega) = \widetilde{\Xi}/3$ and substitute them, then we have

$$\begin{aligned}
&1 - p_1(t_1, q) - p_2(t_2, q) - p_\nu(t_3, q)\\
&= 1 - 4\exp\left(-\frac{1}{2}\frac{t^2}{B_\sigma^2/q + B^2 t/(3q)}\right) - \frac{3\nu^2(4B + 2\omega(s))}{q\widetilde{\Xi}(s)}\\
&\geq 1 - 4\exp\left(-\frac{1}{2}\left(\frac{3qt}{B^2} - \frac{B_\sigma^2 9q}{B^4}\right)\right) - \frac{3\nu^2(4B + 2\omega(s))}{q\widetilde{\Xi}(s)}\\
&= 1 - 4\exp\left(-\frac{q}{2}\left(\frac{1}{B^2}\widetilde{\Xi}(s) - \frac{9B_\sigma^2}{B^4}\right)\right) - \frac{3\nu^2(4B + 2\omega(s))}{q\widetilde{\Xi}(s)}.
\end{aligned}$$

Then, we obtain the result. □

E Approximation Accuracy with Various Kernels with Other Data Sets

Figures 5–9 show the approximation errors with various kernel functions as shown in Section 6.1, with different datasets.

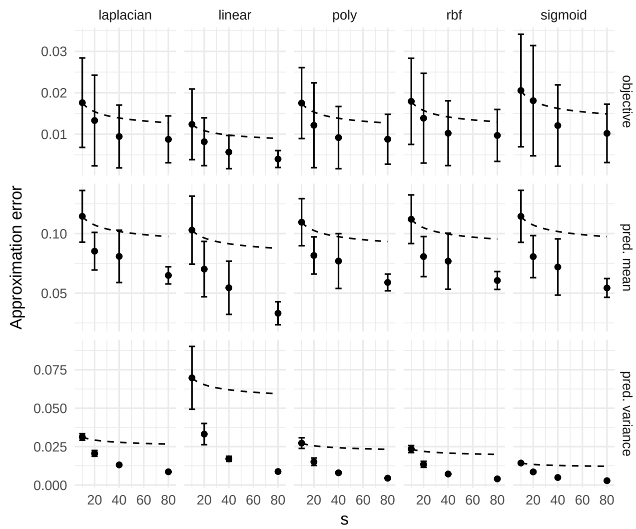

Figure 5: Approximation errors on `abalone` data set. The setting is the same as Section 6.1.

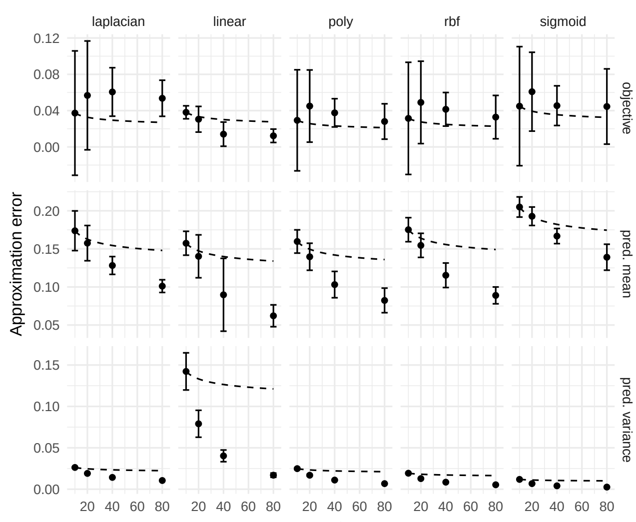

Figure 6: Approximation errors on `cpusmall` data set.

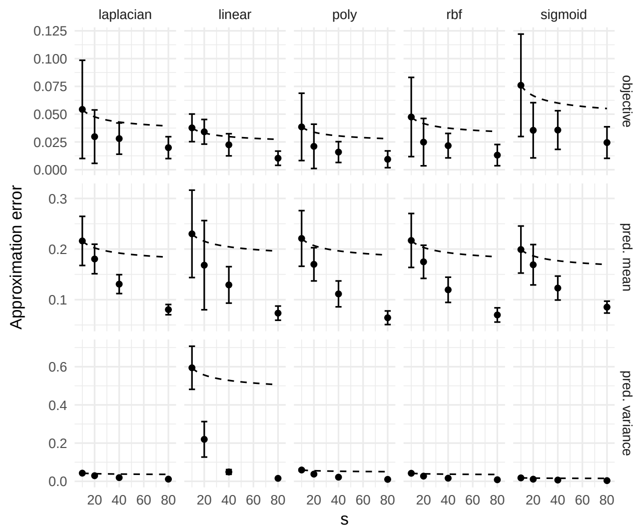

Figure 7: Approximation errors on `housing` data set.

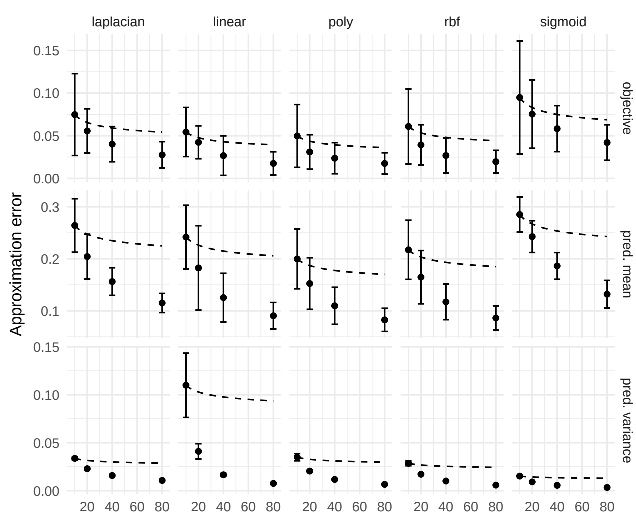

Figure 8: Approximation errors on `mg` data set.

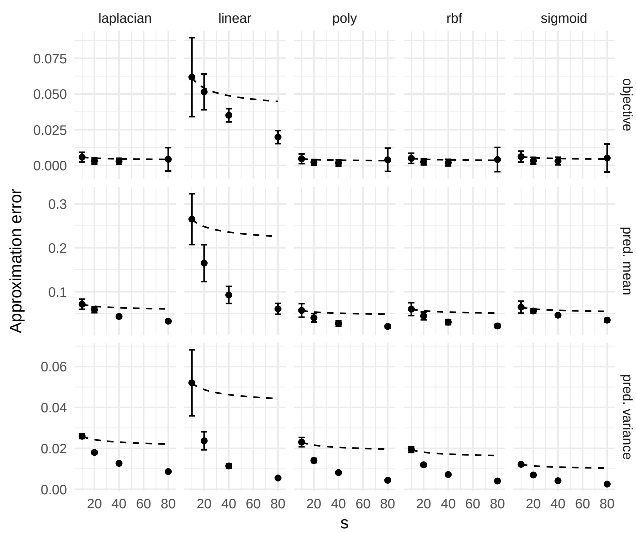

Figure 9: Approximation errors on `space_ga` data set.