Supplementary Material

9. Proof of Theorem 1

The kernel function

$$k(x,y) = p(\min\{x,y\})q(\max\{x,y\}),$$

is in fact the Green's function of the Sturm-Liouville operator (Zaitsev & Polyanin, 2002)

$$\mathcal{L} := \frac{d}{dx}\alpha(x)\frac{d}{dx} + \beta(x).$$

Thus, the inner product induced by k is

$$\langle f,g\rangle_k = \int_0^1 f\mathcal{L}g dx.$$

For any $l \in \mathbb{N}$ and $i \neq j$, the supports of $\phi_{l,i}$ and $\phi_{l,j}$ are $[(i-1)2^{-l},(i+1)2^{-l}]$ and $[(j-1)2^{-l},(j+1)2^{-l}]$, respectively. These two supports are disjoint because both i and j are odd so $\langle\phi_{l,i},\phi_{l,j}\rangle_k = 0$ if $i \neq j$. For any $l,n \in \mathbb{N}$ and any i,j, the supports supt$[\phi_{l,i}]$ and supt$[\phi_{n,j}]$ are either disjoint or nested. If they are disjoint, then $\langle\phi_{n,j},\phi_{n,i}\rangle_k = 0$. If they are nested, without loss of generality assume $l > n$ and $i \leq j2^{l-n}$, then because both p and q satisfy

$$\mathcal{L}p = \mathcal{L}q = 0,$$

we have

$$\begin{aligned}
&\langle\phi_{l,i},\phi_{n,j}\rangle_k\\
&= \int_{(i-1)2^{-l}}^{(i+1)2^{-l}} \phi_{n,j}\mathcal{L}\phi_{l,i}dx\\
&= \int_{(i-1)2^{-l}}^{(i+1)2^{-l}} \phi_{n,j}\mathcal{L}\frac{p(x)q_{l,i-1}-q(x)p_{i,l-1}}{p_{l,i}q_{l,i-1}-q_{l,i}p_{l,i-1}}dx\\
&= 0.
\end{aligned}$$

As a result, we have

$$\langle\phi_{l,i},\phi_{n,j}\rangle_k = \lambda_{l,i}\delta_{(l,i),(n,j)},$$

where $\lambda_{l,i}$ is a function of l and i.

10. Proof of Theorem 3

We need the following lemmas.

Lemma 1. *Denote $f_M = \operatorname{argmin}_{f\in\mathcal{F}_M}\|f_0 - f\|_k$. Then we have*

$$R(f_M) - R(f_0) \leq CM^{-2}\log^{4D-4}M\|f_0\|_k^2,$$

for some constant C.

Proof. According to Assumption 2, we can see that

$$R(f_M) - R(f_0) = \mathbb{E}[m_y''(\mathbf{u}^*)(f_M(x) - f_0(x))^2].$$

In view of Assumption 3, it suffices to prove

$$\|f_M - f_0\|_{L^2}^2 = CM^{-2}\log^{4D-4}M\|f_0\|_k^2,$$

for any $f_0 \in \mathcal{H}_k$ we then can finish the proof. Let $M = |\{(\mathbf{l},\mathbf{i}) : |\mathbf{l}| \leq n, \mathbf{i} \in B_{\mathbf{l}}\}|$. According to theorem 2, we have the following expansion:

$$\begin{aligned}
&\|f_M - f_0\|_{L^2}\\
&= \left\|\sum_{|\mathbf{l}|>n}\sum_{\mathbf{i}\in B_{\mathbf{l}}}\left\langle f_0, \frac{\phi_{\mathbf{l},\mathbf{i}}}{\|\phi_{\mathbf{l},\mathbf{i}}\|_k}\right\rangle_k \frac{\phi_{\mathbf{l},\mathbf{i}}(\cdot)}{\|\phi_{\mathbf{l},\mathbf{i}}\|_k}\right\|_{L^2}\\
&= \left\|\sum_{|\mathbf{l}|>n}\sum_{\mathbf{i}\in B_{\mathbf{i}}}\int_{\mathbf{S}_{\mathbf{l},\mathbf{i}}} f_0(\mathbf{s})\mathcal{L}\phi_{\mathbf{l},\mathbf{i}}(\mathbf{s})d\mathbf{s}\frac{\phi_{\mathbf{l},\mathbf{i}}(\cdot)}{\|\phi_{\mathbf{l},\mathbf{i}}\|_k^2}\right\|_{L^2}.
\end{aligned}$$

where $\mathbf{S}_{\mathbf{l},\mathbf{i}}$ is the support of $\phi_{\mathbf{l}.\mathbf{i}}$. We let

$$v(\cdot)_{\mathbf{l}} := \sum_{\mathbf{i}\in B_{\mathbf{i}}}\int_{\mathbf{S}_{\mathbf{l},\mathbf{i}}} f_0(\mathbf{s})\mathcal{L}\phi_{\mathbf{l},\mathbf{i}}(\mathbf{s})d\mathbf{s}\frac{\phi_{\mathbf{l},\mathbf{i}}(\cdot)}{\|\phi_{\mathbf{l},\mathbf{i}}\|_k^2}.$$

Our first goal is to estimate $v_{\mathbf{l}}$. From theorem 2 of (Ding & Zhang, 2018) or direct calculation based on the property of Green's function, we can see that for any $f \in \mathcal{H}_k$:

$$\int_{\mathbf{S}_{\mathbf{l},\mathbf{i}}} f(\mathbf{s})\mathcal{L}\phi_{\mathbf{l},\mathbf{i}}(\mathbf{s})d\mathbf{s} = \left[\bigotimes_{d=1}^{D}\Delta_{l_d,i_d}\right]f,$$

where

$$\begin{aligned}
\Delta_{l_d,i_d}f &:= \alpha_{l_d,i_d}f\big|_{x_d=z_{l_d,i_d}}\\
&\quad - \beta_{l_d,i_d-1}f\big|_{x_d=z_{l_d,i_d-1}} - \beta_{l_d,i_d+1}f\big|_{x_d=z_{l_d,i_d}},\\
\alpha_{l,i} &= \frac{p_{l,i+1}q_{l,i-1} - p_{l,i-1}q_{l,i+1}}{[p_{l,i+1}q_{l,i} - p_{l,i}q_{l,i+1}][p_{l,i1}q_{l,i-1} - p_{l,i-1}q_{l,i}]},\\
\beta_{l,i} &= \frac{1}{p_{l,i+1}q_{l,i} - p_{l,i}q_{l,i+1}},
\end{aligned}$$

and $\bigotimes$ denotes the tensor product of the $\Delta_{l,i}$ operators. Since both q and p are the solution of the SL-equation, therefore, p,q are twice differentiable. We have

$$\begin{aligned}
&\frac{1}{p_{l,i+1}q_{l,i} - p_{l,i}q_{l,i+1}}\\
&= \frac{2^l}{[p_{l,i+1}q_{l,i} - p_{l,i}q_{l,i}]/2^{-l} - [p_{l,i}q_{l,i+1} - p_{l,i}q_{l,i}]/2^{-l}}\\
&\sim \frac{2^l}{p'_{l,i}q_{l,i} - p_{l,i}q'_{l,i}}.
\end{aligned}$$

We notice that $p'_{l,i}q_{l,i} - p_{l,i}q'_{l,i}$ is the Wronskian of the SL-operator, which is bounded away from 0. Therefore, Δ_{l_d,i_d} acting on f has the following approximation:

$$\begin{aligned}\Delta_{l_d,i_d} f &\sim \frac{\left[2f\big|_{x_d=z_{l_d,i_d}} - f\big|_{x_d=z_{l_d,i_d-1}} - f\big|_{x_d=z_{l_d,i_d+1}}\right]}{2^{-l}} \\ &\leq C \max_{j=1,-1} \left\{ \frac{\left|f\big|_{x_d=z_{l_d,i_d+j}} - f\big|_{x_d=z_{l_d,i_d}}\right|}{2^{-l}} \right\}.\end{aligned}$$

As a result, $\bigotimes_{d=1}^D \Delta_{l_d,i_d}$ acting on f has the following approximation:

$$\begin{aligned}&\bigotimes_{d=1}^D \Delta_{l_d,i_d} f \\ &\leq C \prod_{d=1}^D \max_{j=1,-1} \left\{ \frac{\left|f\big|_{x_d=z_{l_d,i_d+j}} - f\big|_{x_d=z_{l_d,i_d}}\right|}{2^{-l}} \right\}.\end{aligned}$$

From the same reasoning, we can see that

$$\|\phi_{\mathbf{l},\mathbf{i}}\|_k^2 = \prod_{d=1}^D \alpha_{l_d,i_d} \sim 2^{|\mathbf{l}|}.$$

We also Taylor expand ϕ_{l_d,i_d} for each $1 \leq d \leq D$ up to second order and from direct calculation, we can have

$$\phi_{l_d,i_d}(x) \sim \max\left\{0, 1 - \frac{|x - z_{l_d,i_d}|}{2^{-l_d}}\right\} + \mathcal{O}(2^{-l_d}).$$

This gives us the approximation up to second order:

$$\begin{aligned}&\|\phi_{\mathbf{l},\mathbf{i}}\|_{L_2}^2 \\ &= \int_{\mathbf{S}_{\mathbf{l},\mathbf{i}}} \prod_{d=1}^D \phi_{l_d,i_d}^2(s_d) d\mathbf{s} \\ &\sim \int_{\mathbf{S}_{\mathbf{l},\mathbf{i}}} \prod_{d=1}^D \left[\max\left\{0, 1 - \frac{|s - z_{l_d,i_d}|}{2^{-l_d}}\right\}\right]^2 d\mathbf{s} \\ &= \big(\frac{2}{3}\big)^D 2^{-|\mathbf{l}|} = \big(\frac{1}{3}\big)^D \mathrm{Vol}(\mathbf{S}_{\mathbf{l},\mathbf{i}}).\end{aligned}$$

Therefore, we can have the following estimate for $v_{\mathbf{l}}$:

$$\begin{aligned}\|v_{\mathbf{l}}\|_{L^2} &= \|\sum_{\mathbf{i}\in B_{\mathbf{i}}} \int_{\mathbf{S}_{\mathbf{l},\mathbf{i}}} f_0(\mathbf{s}) \mathcal{L}\phi_{\mathbf{l},\mathbf{i}}(\mathbf{s}) d\mathbf{s} \frac{\phi_{\mathbf{l},\mathbf{i}}(\cdot)}{\|\phi_{\mathbf{l},\mathbf{i}}\|_k^2}\|_{L^2} \\ &\leq \left| 2^{-2|\mathbf{l}|} C \sum_{\mathbf{i}\in B_{\mathbf{i}}} \left[\bigotimes_{d=1}^D \Delta_{l_d,i_d} f\right]^2 \mathrm{Vol}(\mathbf{S}_{\mathbf{l},\mathbf{i}}) \right|^{\frac{1}{2}} \\ &\sim 2^{-|\mathbf{l}|} \left\| \prod_{d=1}^D \frac{\partial}{\partial x_d} f_0 \right\|_{L^2} \\ &\sim 2^{-|\mathbf{l}|} \|f_0\|_k,\end{aligned}$$

where the second line is from the fact that supports of $\{\phi_{\mathbf{l},\mathbf{i}} : \mathbf{i} \in B_{\mathbf{l}}\}$ are disjoint, the third line is from the Riemann integral approximation and the last line is from the energy estimate assumption of SL-operator (see, for instance, section 6.2.2 of (Evans, 2010)). Finally, we have:

$$\begin{aligned}\|f_0 - f_M\|_{L^2} &\leq \sum_{|\mathbf{l}|>n} \|v_{\mathbf{l}}\|_{L^2} \\ &\sim \|f_0\|_k \sum_{|\mathbf{l}|>n} 2^{-|\mathbf{l}|} \\ &= \|f_0\|_k \sum_{i>n} 2^{-i} \sum_{|\mathbf{l}|=i} 1 \\ &= \|f_0\|_k \sum_{i>n} 2^{-i} \binom{i-1}{d-1} \\ &\sim \|f_0\|_k 2^{-n} n^{D-1},\end{aligned}$$

where the identity of the last line can be verified in (Ding et al., 2019). From (Bungartz & Griebel, 2004) we also have

$$M = \mathcal{O}(2^n n^{D-1}).$$

We can substitute this identity to the previous equation to have the final result. □

The (ϵ, L_∞)-covering number of a function space $\mathcal{F}$, denoted as $N(\epsilon, \mathcal{F}, \|\cdot\|_{L_\infty})$, is defined as the smallest number N_0, so that there exist centers $f_1, \ldots, f_{N_0}$, and for each $f \in \mathcal{F}$, there exists f_i so that $\|f - f_i\|_{L_\infty} < \epsilon$.

Lemma 2. *The covering number of the unit ball of $\mathcal{H}_k$, denoted as $\mathcal{F} := \{f \in \mathcal{H}_k : \|f\|_k \leq 1\}$, is bounded as follows:*

$$\log N(\epsilon, \mathcal{F}, \|\cdot\|_{L_\infty}) = \mathcal{O}\left(\frac{1}{\varepsilon} \log^{D-\frac{1}{2}} \frac{1}{\varepsilon}\right).$$

Proof. When $k(\mathbf{x}, \mathbf{y}) = e^{-\omega\|\mathbf{x}-\mathbf{y}\|_1}$ or $k(\mathbf{x}, \mathbf{y}) = \prod_{d=1}^D \min\{x_d, y_d\}$, $\mathcal{H}_k$ is equivalent to the Sobolev space of mixed first derivative $\mathcal{H}^1_{\text{mix}}([0,1]^D)$ (Ding et al., 2019). According to 6.6 of (Dung et al., 2016), we can immediately derive the result. When kernel k differs from these two, the energy property of an SL-operator requires that

$$\begin{aligned}\langle f, f\rangle_k &= \int_{[0,1]^D} f(\mathbf{x}) \left[\prod_{d=1}^D \mathcal{L}\right] f(\mathbf{x}) d(\mathbf{x}) \\ &\leq C \int_{[0,1]^D} \left|\prod_{d=1}^D \frac{\partial}{\partial x_d} f\right|^2 d\mathbf{x},\end{aligned}$$

which implies that $\mathcal{H}_k$ can be embedded on $\mathcal{H}^1_{\text{mix}}$. Therefore, the covering number of $\mathcal{H}_k$ must be bounded by that of $\mathcal{H}^1_{\text{mix}}$. □

Lemma 3 shows the the function classes associated with the learning problem are Donsker. We refer to (van der Vaart & Wellner, 1996) for the definition and properties of Donsker classes. Let $\mathcal{G}_R := \{L(y, f(\mathbf{x})) : \|f\|_k \le R\}$.

Lemma 3. *Let P be the probability measure of $(\mathbf{x}, y)$. The space $\mathcal{G}_R$ is P-Donsker for each $R > 0$.*

Proof. In view of Theorem 2.5.6 of (van der Vaart & Wellner, 1996), it suffices to prove that

$$\int_0^\infty \sqrt{\log N_{[]}(\epsilon, \mathcal{G}_R, \|\cdot\|_{L_2(P)})} d\epsilon < \infty,$$

where $N_{[]}(\epsilon, \mathcal{G}_R, \|\cdot\|_{L_2(P)})$ is the covering number with bracketing defined as follows. For function $g : \mathbb{R}^D \times \mathbb{R} \to \mathbb{R}$, its $L_2(P)$ norm is defined as $[\mathbb{E}[g(x,y)]^2]^{1/2}$. Given functions g_L, g_U such that $g_L(\mathbf{u}, v) \le g_U(\mathbf{u}, v)$ for each $(\mathbf{u}, v)$, define the bracket $[g_L, g_U]$ as the set of functions $\{g : g_L(\mathbf{u}, v) \le g(\mathbf{u}, v) \le g_U(\mathbf{u}, v)\}$. The covering number with bracketing $N_{[]}(\epsilon, \mathcal{G}_R, \|\cdot\|_{L_2(P)})$ is the smallest number N_0 so that there exist brackets $[g_{L,1}, g_{U,1}], \ldots, [g_{L,N_0}, g_{U,N_0}]$, such that $\cup_{i=1}^{N_0}[g_{L,i}, g_{U,i}] \supset \mathcal{G}_R$, and $\|g_{U,i} - g_{L,i}\|_{L_2(P)} \le \epsilon$ for all i.

Let $\mathcal{F}_R = \{f : \|f\|_k < R\}$. We start with the centers $f_1, \ldots, f_{N_0}$ with $N_0 = N(\epsilon, \mathcal{F}_R, \|\cdot\|_{L_\infty}) = N(\epsilon/R, \mathcal{F}_1, \|\cdot\|_{L_\infty})$ so that for each $f \in \mathcal{F}_R$, there exists $f_i =: \xi(f)$ such that $\|f - f_i\|_{L_\infty} < \epsilon$. To bound the covering number with bracketing, we need to construct the associated brackets. The reproduction property implies that $\|f\|_{L_\infty} \le c\|f\|_k$ with $c := \max_x k(x,x)$. Then for any $f \in \mathcal{F}_R$, by mean value theorem,

$$\begin{aligned} |L(y, f(\mathbf{x})) - L(y, \xi(f)(\mathbf{x}))| &\le \sup_{|\mathbf{u}|<cR} \left|\frac{\partial L}{\partial u}(y, \mathbf{u})\right| \epsilon \\ &=: S(y)\epsilon. \end{aligned} \tag{12}$$

Now we define $g_{L,i}(\mathbf{u}, v) = L(v, f_i(\mathbf{u})) - S(v)\epsilon$ and $g_{U,i}(\mathbf{u}, v) = L(v, f_i(\mathbf{u})) + S(v)\epsilon$. Clearly $g_{L,i} \le g_{U,i}$ and

$$\|g_{U,i} - g_{L,i}\|_{L_2(P)} = 2\epsilon[\mathbb{E}[S(y)]^2]^{1/2},$$

which is a multiple of ϵ according to Assumptions 2-3. Besides, (12) implies that for all f such that $\|f - f_i\|_{L_\infty} < \epsilon$, $L(v, f(\mathbf{u})) \in [g_{L,i}, g_{U,i}]$. Now we invoke Lemma 2 to find that

$$\begin{aligned} &\log N_{[]}(2\epsilon[\mathbb{E}[S(y)]^2]^{1/2}, \mathcal{F}_R, \|\cdot\|_{L_2(P)}) \\ = \ &\mathcal{O}\left(\frac{R}{\varepsilon} \log^{D-\frac{1}{2}} \frac{R}{\varepsilon}\right), \end{aligned}$$

which implies the desired result. □

To bound the generalization error, we observe that

$$\begin{aligned} R(\hat{f}) - R(f_0) &= \left\{R(\hat{f}) - \frac{1}{N}\sum_{i=1}^N L(y_i, \hat{f}(\mathbf{x}_i))\right\} \\ &+ \left\{\frac{1}{N}\sum_{i=1}^N L(y_i, \hat{f}(\mathbf{x}_i)) - \frac{1}{N}\sum_{i=1}^N L(y_i, f_M(\mathbf{x}_i))\right\} \\ &+ \left\{\frac{1}{N}\sum_{i=1}^N L(y_i, f_M(\mathbf{x}_i)) - R(f_M)\right\} \\ &+ \left\{R(f_M) - R(f_0)\right\} =: I_1 + I_2 + I_3 + I_4. \end{aligned}$$

We will bound I_1 and I_3 by applying a uniform error bound of empirical processes. For I_2, we have

$$I_2 \le \lambda\|f_M\|_k^2 - \lambda\|\hat{f}\|_k^2 \le \lambda\|f_0\|_K^2 = \mathcal{O}(N^{-1/2})\|f_0\|_k^2,$$

where the first inequality follows from the optimality condition

$$\begin{aligned} &\frac{1}{N}\sum_{i=1}^N L(y_i, \hat{f}(\mathbf{x}_i)) + \lambda\|\hat{f}\|_k^2 \\ \le &\frac{1}{N}\sum_{i=1}^N L(y_i, f_M(\mathbf{x}_i)) + \lambda\|f_M\|_k^2. \end{aligned} \tag{13}$$

The term I_4 is bounded by Lemma 1.

Now we turn to I_1 and I_3. To show that $I_1 = \mathcal{O}_p(N^{-1/2})$ and $I_3 = \mathcal{O}_p(N^{-1/2})$, it suffices to show that the functions $L(y, \hat{f}(\mathbf{x}))$ and $L(y, f_M(\mathbf{x}))$ fall in a Donsker class (van der Vaart & Wellner, 1996) with probability arbitrarily close to one. For $L(y, f_M(\mathbf{x}))$, this is clearly true in view of Lemma 3 and the fact that $\|f_M\|_k \le \|f_0\|_k$. Therefore, $I_3 = \mathcal{O}_p(N^{-1/2})$. For $L(y, \hat{f}(\mathbf{x}))$, it suffices to prove that $\|\hat{f}\|_k = \mathcal{O}_p(1)$. To show this result, we start with the optimality condition (13). In view of Assumption 2, we can write

$$\begin{aligned} L(y, f(\mathbf{x})) - L(y, f_0(\mathbf{x})) = W \cdot (f(\mathbf{x}) - f_0(\mathbf{x})) \\ + m_y''(u^*)(f(\mathbf{x}) - f_0(\mathbf{x}))^2, \end{aligned}$$

where $W = m_y'(f_0(\mathbf{x}))$, and u^* lies between $f(\mathbf{x})$ and $f_0(\mathbf{x})$. Assumption 1 implies that for any $\delta \in \mathcal{H}_k$,

$$\begin{aligned} 0 &= \frac{\partial}{\partial t} R(f_0 + t\delta)\Big|_{t=0} \\ &= \frac{\partial}{\partial t} \mathbb{E}L(y, f_0(\mathbf{x}) + t\delta(\mathbf{x}))\Big|_{t=0} \\ &= \mathbb{E}\left[m_y'(f_0(\mathbf{x}))\delta(\mathbf{x})\right], \end{aligned} \tag{14}$$

where in the last equation we interchange the partial derivative and the expectation, which is valid because of Assumptions 2 and 3. Let $P_{\mathbf{x}}$ be the probability measure of $\mathbf{x}$. Since $\delta \in \mathcal{H}_k$ is arbitrary, (14) implies

$$0 = \mathbb{E}m_y'(f_0(\mathbf{x})) = \mathbb{E}W.$$

We then invoke (13) and Assumption 2 to find

$$\begin{aligned}
&\lambda\|\hat{f}\|_k^2 \leq -\frac{1}{N}\sum_{i=1}^N W_i(\hat{f}(\mathbf{x}_i) - f_0(\mathbf{x}_i)) \\
+ \quad & \left\{\frac{1}{N}\sum_{i=1}^N L(y_i, f_M(\mathbf{x}_i)) - \frac{1}{N}\sum_{i=1}^N L(y_i, f_0(\mathbf{x}_i))\right\} \\
- \quad & V(\hat{f}(\mathbf{x}_i) - f_0(\mathbf{x}_i))^2 + \lambda\|f_0\|_k^2 \\
=: \quad & J_1 + J_2 + J_3 + J_4, \qquad (15)
\end{aligned}$$

for some $V > 0$ due to the strong convexity of $m_y(\cdot)$. For the first term, we have

$$\begin{aligned}
J_1 \quad &\leq \quad (\|\hat{f}\|_k + 1) \sup_{f \in \mathcal{H}_k} \frac{1}{N}\sum_{i=1}^N -W_i \frac{f(\mathbf{x}_i) - f_0(\mathbf{x}_i)}{\|f\|_k + 1} \\
&= \quad (\|\hat{f}\|_k + 1)\mathcal{O}_p(N^{-1/2}),
\end{aligned}$$

where the last step follows from the fact that $\mathbb{E}W_i = 0$, W_i is bounded, and Lemma 3.4.3 of (van der Vaart & Wellner, 1996) and the fact that $\|f - f_0\|_k/(\|f\|_k + 1) = \mathcal{O}(1)$. Clearly, we have $J_2 = I_3 + \mathcal{O}_p(N^{-1/2}) = \mathcal{O}_p(N^{-1/2})$ according to the central limit theorem. The third term is clearly non-positive. We also have $J_4 = \mathcal{O}_p(N^{-1/2})$ by assumption for λ.

Now we conclude from (15) that

$$\lambda\|\hat{f}\|_k^2 \leq \|\hat{f}\|_k \mathcal{O}_p(N^{-1/2}) + \mathcal{O}_p(N^{-1/2}),$$

which implies $\|\hat{f}\|_k = \mathcal{O}_p(1)$. This completes the proof.