Supplementary Material: Unifying Framework for Fast Learning Rate of Non-Sparse Multiple Kernel Learning

Taiji Suzuki
Department of Mathematical Informatics
The University of Tokyo
Tokyo 113-8656, Japan
s-taiji@stat.t.u-tokyo.ac.jp

A Outline of the proof of Theorem 1

Before we go into the rigorous proof of Theorem 1, we describe a brief (but mathematically incorrect) outline of the proof. For simplicity, we assume that the infinity norms of $\hat{f}$ and f^* are bounded from above by a constant: $\|\hat{f}\|_\infty, \|f^*\|_\infty \leq C$. Write $\Delta f = \hat{f} - f^*$. Let P_n and P be operators that give the empirical mean and the population means of a function respectively: $P_n f = \frac{1}{n}\sum_{i=1}^n f(x_i, y_i)$ and $Pf = \mathrm{E}[f(X,Y)]$. Then by the definition of $\hat{f}$, we have

$$P_n(Y-\hat{f})^2 + \lambda_1^{(n)}\|\hat{f}\|_\psi^2 \leq P_n(Y-f^*)^2 + \lambda_1^{(n)}\|f^*\|_\psi^2. \tag{S-1}$$

On the other hand, we have the following equation:

$$P(Y-f^*)^2 + P(\hat{f}-f^*)^2 = P(Y-\hat{f})^2. \tag{S-2}$$

Summing up Eq. (S-1) and Eq. (S-2), we obtain

$$\begin{aligned} P(f^*-\hat{f})^2 + \lambda_1^{(n)}\|\hat{f}\|_\psi^2 &\leq (P_n - P)\left\{(Y-f^*)^2 - (Y-\hat{f})^2\right\} + \lambda_1^{(n)}\|f^*\|_\psi^2 \\ &\leq (P_n - P)\left\{(-2\epsilon - f^* + \hat{f})(f^*-\hat{f})\right\} + \lambda_1^{(n)}\|f^*\|_\psi^2. \end{aligned} \tag{S-3}$$

Note that $|-2\epsilon - f^* + \hat{f}| \leq 2(L+C)$. Therefore, using the contraction inequality for the Rademacher complexity [2, Theorem 4.12] and Talagrand's concentration inequality (Proposition 6), we have the following upper bound of the first term in the RHS of the above inequality (S-3):

$$\begin{aligned} &(P_n - P)\left\{(-2\epsilon - f^* + \hat{f})(f^*-\hat{f})\right\} \\ \leq &\mathcal{O}_p\left(\sum_{m=1}^M \frac{\|\Delta f_m\|_{L_2(\Pi)}^{1-s_m}\|\Delta f_m\|_{\mathcal{H}_m}^{s_m}}{\sqrt{n}} \vee \frac{\|\Delta f_m\|_{L_2(\Pi)}^{\frac{(1-s_m)^2}{1+s_m}}\|\Delta f_m\|_{\mathcal{H}_m}^{\frac{s_m(3-s_m)}{1+s_m}}}{n^{\frac{1}{1+s_m}}} + \sum_{m=1}^M \sqrt{\frac{\log(M)}{n}}\|\Delta f_m\|_{L_2(\Pi)}\right) \\ \leq &\mathcal{O}_p\Bigg(\sum_{m=1}^M \frac{r_m^{-s_m}}{\sqrt{n}}\left(\|\Delta f_m\|_{L_2(\Pi)} + s_m r_m\|\Delta f_m\|_{\mathcal{H}_m}\right) \\ &\quad + \sum_{m=1}^M \frac{r_m^{-\frac{s_m(3-s_m)}{1+s_m}}}{n^{\frac{1}{1+s_m}}}\left(\|\Delta f_m\|_{L_2(\Pi)} + s_m r_m\|\Delta f_m\|_{\mathcal{H}_m}\right) \\ &\quad + \sum_{m=1}^M \sqrt{\frac{\log(M)}{n}}\|\Delta f_m\|_{L_2(\Pi)}\Bigg) \qquad (\because \text{Eq. (S-11), Eq. (S-12)}) \end{aligned} \tag{S-4}$$

$$
\begin{aligned}
\leq & \mathcal{O}_p\Bigg\{ \left(\sum_{m=1}^M \frac{r_m^{-2s_m}}{n}\right)^{\frac{1}{2}} \left(\sum_{m=1}^M \|\Delta f_m\|_{L_2(\Pi)}^2\right)^{\frac{1}{2}} + \left\| \left(\frac{s_m r_m^{1-s_m}}{\sqrt{n}}\right)_{m=1}^M \right\|_{\psi^*} \|\Delta f\|_\psi \\
& + \left(\sum_{m=1}^M \frac{r_m^{-\frac{2s_m(3-s_m)}{1+s_m}}}{n^{\frac{2}{1+s_m}}}\right)^{\frac{1}{2}} \left(\sum_{m=1}^M \|\Delta f_m\|_{L_2(\Pi)}^2\right)^{\frac{1}{2}} + \left\| \left(\frac{s_m r_m^{1-\frac{s_m(3-s_m)}{1+s_m}}}{n^{\frac{1}{1+s_m}}}\right)_{m=1}^M \right\|_{\psi^*} \|\Delta f\|_\psi \\
& + \sqrt{\frac{M\log(M)}{n}} \left(\sum_{m=1}^M \|\Delta f_m\|_{L_2(\Pi)}^2\right)^{\frac{1}{2}} \Bigg\},
\end{aligned} \tag{S-5}
$$

where we used Cauchy-Schwarz inequality and the duality of the ψ-norm in the last line. Utilizing the relation $\sum_{m=1}^M \|\Delta f_m\|_{L_2(\Pi)}^2 \leq \frac{1}{\kappa_M}\|\Delta f\|_{L_2(\Pi)}^2$, Eq. (S-5) implies

$$
\begin{aligned}
& (P_n - P)\left\{(-2\epsilon - f^* + \hat{f})(f^* - \hat{f})\right\} \\
\leq & \mathcal{O}_p\left(\alpha_1\|\Delta f\|_{L_2(\Pi)} + \beta_1\|\Delta f\|_\psi + \alpha_2\|\Delta f\|_{L_2(\Pi)} + \beta_2\|\Delta f\|_\psi + \sqrt{\frac{M\log(M)}{n}}\|\Delta f\|_{L_2(\Pi)}\right) \\
= & \mathcal{O}_p\left(\alpha_1\|\Delta f\|_{L_2(\Pi)} + \alpha_1\frac{\beta_1}{\alpha_1}\|\Delta f\|_\psi + \alpha_2\|\Delta f\|_{L_2(\Pi)} + \alpha_2\frac{\beta_2}{\alpha_2}\|\Delta f\|_\psi + \sqrt{\frac{M\log(M)}{n}}\|\Delta f\|_{L_2(\Pi)}\right) \\
\leq & \mathcal{O}_p\left(\alpha_1^2 + \alpha_2^2 + \frac{M\log(M)}{n}\right) + \frac{1}{2}\|\Delta f\|_{L_2(\Pi)}^2 + \frac{1}{2}\left[\left(\frac{\beta_2}{\alpha_2}\right)^2 + \left(\frac{\beta_2}{\alpha_2}\right)^2\right]\|\Delta f\|_\psi^2.
\end{aligned} \tag{S-6}
$$

Here substitute the relation $\|\Delta f\|_\psi^2 \leq (\|\hat{f}\|_\psi + \|f^*\|_\psi)^2 \leq 2(\|\hat{f}\|_\psi^2 + \|f^*\|_\psi^2)$ into Eq. (S-6), combine Eq. (S-3) and Eq. (S-6), move the terms $\frac{1}{2}\|\Delta f\|_{L_2(\Pi)}^2$ and $\left[\left(\frac{\beta_2}{\alpha_2}\right)^2 + \left(\frac{\beta_2}{\alpha_2}\right)^2\right]\|\hat{f}\|_\psi^2$ to the right hand side, then we obtain the assertion.

B Relation between Entropy Number and Spectral Condition

Associated with the ϵ-covering number, the *i-th entropy number* $e_i(\mathcal{H}_m \to L_2(\Pi))$ is defined as the infimum over all $\varepsilon > 0$ for which $N(\varepsilon, \mathcal{B}_{\mathcal{H}_m}, L_2(\Pi)) \leq 2^{i-1}$. If the spectral assumption (A3) holds, the relation (2) implies that the i-th entropy number is bounded as

$$
e_i(\mathcal{H}_m \to L_2(\Pi)) \leq C i^{-\frac{1}{2s}}, \tag{S-7}
$$

where C is a constant. To bound empirical process a bound of the entropy number with respect to the empirical distribution is needed. The following proposition gives an upper bound of that (see Corollary 7.31 of [5], for example).

Proposition 4. *If there exists constants $0 < s < 1$ and $C \geq 1$ such that $e_i(\mathcal{H}_m \to L_2(\Pi)) \leq C i^{-\frac{1}{2s}}$, then there exists a constant $c_s > 0$ only depending on s such that*

$$
\mathrm{E}_{D_n \sim \Pi^n}[e_i(\mathcal{H}_m \to L_2(D_n))] \leq c_s C (\min(i,n))^{\frac{1}{2s}} i^{-\frac{1}{s}},
$$

in particular $\mathrm{E}_{D_n \sim \Pi^n}[e_i(\mathcal{H}_m \to L_2(D_n))] \leq c_s C i^{-\frac{1}{2s}}$.

C Basic Propositions

The following two propositions are keys to prove Theorem 1. Let $\{\sigma_i\}_{i=1}^n$ be i.i.d. Rademacher random variables, i.e., $\sigma_i \in \{\pm 1\}$ and $P(\sigma_i = 1) = P(\sigma_i = -1) = \frac{1}{2}$.

Proposition 5. [5, Theorem 7.16] *Let $\mathcal{B}_{\sigma,a,b} \subset \mathcal{H}_m$ be a set such that $\mathcal{B}_{\sigma,a,b} = \{f_m \in \mathcal{H}_m \mid \|f_m\|_{L_2(\Pi)} \leq \sigma, \|f_m\|_{\mathcal{H}_m} \leq a, \|f_m\|_\infty \leq b\}$. Assume that there exist constants $0 < s < 1$ and $0 < \tilde{c}_s$ such that*

$$
\mathrm{E}_{D_n}[e_i(\mathcal{H}_m \to L_2(D_n))] \leq \tilde{c}_s i^{-\frac{1}{2s}}.
$$

Then there exists a constant C'_s depending only s such that

$$\mathrm{E}\left[\sup_{f_m\in\mathcal{B}_{\sigma,a,b}}\left|\frac{1}{n}\sum_{i=1}^n\sigma_i f_m(x_i)\right|\right]\le C'_s\left(\frac{\sigma^{1-s}(\tilde{c}_s a)^s}{\sqrt{n}}\vee(\tilde{c}_s a)^{\frac{2s}{1+s}}b^{\frac{1-s}{1+s}}n^{-\frac{1}{1+s}}\right). \tag{S-8}$$

Proposition 6. (Talagrand's Concentration Inequality [6, 1]) *Let $\mathcal{G}$ be a function class on $\mathcal{X}$ that is separable with respect to ∞-norm, and $\{x_i\}_{i=1}^n$ be i.i.d. random variables with values in $\mathcal{X}$. Furthermore, let $B\ge 0$ and $U\ge 0$ be $B:=\sup_{g\in\mathcal{G}}\mathrm{E}[(g-\mathrm{E}[g])^2]$ and $U:=\sup_{g\in\mathcal{G}}\|g\|_\infty$, then there exists a universal constant K such that, for $Z:=\sup_{g\in\mathcal{G}}\left|\frac{1}{n}\sum_{i=1}^n g(x_i)-\mathrm{E}[g]\right|$, we have*

$$P\left(Z\ge K\left[\mathrm{E}[Z]+\sqrt{\frac{Bt}{n}}+\frac{Ut}{n}\right]\right)\le e^{-t}.$$

D Proof of Theorem 1

Let $r_m>0$ $(m=1,\dots,M)$ be arbitrary positive reals. Given $\{r_m\}_{m=1}^M$, we determine $U_{n,s_m}^{(m)}(f_m)$ as follows:

$$U_{n,s_m}^{(m)}(f_m):=3\left(\frac{r_m^{-s_m}}{\sqrt{n}}\vee\frac{r_m^{-\frac{s_m(3-s_m)}{1+s_m}}}{n^{\frac{1}{1+s_m}}}\right)\left(\|f_m\|_{L_2(\Pi)}+s_m r_m\|f_m\|_{\mathcal{H}_m}\right)+\sqrt{\frac{\log(M)}{n}}\|f_m\|_{L_2(\Pi)}.$$

It is easy to see $U_{n,s_m}^{(m)}(f_m)$ is an upper bound of the quantity $\frac{\|f_m\|_{L_2(\Pi)}^{1-s_m}\|f_m\|_{\mathcal{H}_m}^{s_m}}{\sqrt{n}}\vee\frac{\|f_m\|_{L_2(\Pi)}^{\frac{(1-s_m)^2}{1+s_m}}\|f_m\|_{\mathcal{H}_m}^{\frac{s_m(3-s_m)}{1+s_m}}}{n^{\frac{1}{1+s_m}}}$ (this corresponds to the RHS of Eq. (S-8)) because

$$\begin{aligned}\frac{\|f_m\|_{L_2(\Pi)}^{1-s_m}\|f_m\|_{\mathcal{H}_m}^{s_m}}{\sqrt{n}}&=\frac{r_m^{1-s_m}}{\sqrt{n}}\left(\frac{\|f_m\|_{L_2(\Pi)}}{r_m}\right)^{1-s_m}\|f_m\|_{\mathcal{H}_m}^{s_m} &\text{(S-9)}\\&\overset{\text{(Young)}}{\le}\frac{r_m^{1-s_m}}{\sqrt{n}}\left((1-s_m)\frac{\|f_m\|_{L_2(\Pi)}}{r_m}+s_m\|f_m\|_{\mathcal{H}_m}\right) &\text{(S-10)}\\&\le\frac{r_m^{-s_m}}{\sqrt{n}}\left(\|f_m\|_{L_2(\Pi)}+s_m r_m\|f_m\|_{\mathcal{H}_m}\right), &\text{(S-11)}\end{aligned}$$

where we used Young's inequality $a^{1-s_m}b^{s_m}\le(1-s_m)a+s_m b$ in the second line, and similarly we obtain

$$\begin{aligned}\frac{\|f_m\|_{L_2(\Pi)}^{\frac{(1-s_m)^2}{1+s_m}}\|f_m\|_{\mathcal{H}_m}^{\frac{s_m(3-s_m)}{1+s_m}}}{n^{\frac{1}{1+s_m}}}&\le\frac{r_m^{-\frac{s_m(3-s_m)}{1+s_m}}}{n^{\frac{1}{1+s_m}}}\left(\|f_m\|_{L_2(\Pi)}+\frac{s_m(3-s_m)}{1+s_m}r_m\|f_m\|_{\mathcal{H}_m}\right)\\&\le 3\frac{r_m^{-\frac{s_m(3-s_m)}{1+s_m}}}{n^{\frac{1}{1+s_m}}}\left(\|f_m\|_{L_2(\Pi)}+s_m r_m\|f_m\|_{\mathcal{H}_m}\right), &\text{(S-12)}\end{aligned}$$

where we used $\frac{s_m(3-s_m)}{1+s_m}\le 3s_m$ in the last inequality.

Now we define

$$\phi:=\max\left(KL\left[2\tilde{C}_*+1+C_1\right],K\left[2C_1\tilde{C}_*+C_1+C_1^2\right]\right),$$

where $\tilde{C}^*$ is a constant defined later in Lemma 11, C_1 is the one introduced in Assumption 4, K is the universal constant appeared in Talagrand's concentration inequality (Proposition 6) and L is the one introduced in Assumption 1 to bound the magnitude of noise. Remind the definition of $\eta(t)$:

$$\eta(t):=\eta_n(t)=\max(1,\sqrt{t},t/\sqrt{n}).$$

We define events $\mathscr{E}_1(t)$ and $\mathscr{E}_2(t')$ as

$$\mathscr{E}_1(t)=\left\{\left|\frac{1}{n}\sum_{i=1}^n\epsilon_i f_m(x_i)\right|\le\phi U_{n,s_m}^{(m)}(f_m)\eta(t),\ \forall f_m\in\mathcal{H}_m\ (m=1,\dots,M)\right\}, \tag{S-13}$$

$$\mathscr{E}_2(t') = \Bigg\{ \left| \left\|\sum_{m=1}^M f_m\right\|_n^2 - \left\|\sum_{m=1}^M f_m\right\|_{L_2(\Pi)}^2 \right| \leq \phi\sqrt{n}\left(\sum_{m=1}^M U_{n,s_m}^{(m)}(f_m)\right)^2 \eta(t'), \\ \forall f_m \in \mathcal{H}_m \ (m=1,\dots,M)\Bigg\}. \tag{S-14}$$

Using Lemmas 12 and 13 that will be shown in Appendix E, we see that the events $\mathscr{E}_1(t)$ and $\mathscr{E}_2(t')$ occur with probability no less than $1-\exp(-t)$ and $1-\exp(-t')$ respectively as in the following Lemma.

Lemma 7. *Under the Basic Assumption (Assumption 1), the Spectral Assumption (Assumption 2) and the Embedded Assumption (Assumption 4), the probabilities of $\mathscr{E}_1(t)$ and $\mathscr{E}_2$ are bounded as*

$$P(\mathscr{E}_1(t)) \geq 1-\exp(-t), \quad P(\mathscr{E}_2(t')) \geq 1-\exp(-t').$$

Proof. Lemma 13 immediately gives $P(\mathscr{E}_1(t)) \geq 1-\exp(-t)$ by noticing $\bar{\phi}$ in the statement of Lemma 13 satisfies $\bar{\phi} \leq \phi$. Moreover, since $\bar{\phi}'$ in the statement of Lemma 12 satisfies $\bar{\phi}' \leq \phi$, we have $P(\mathscr{E}_2(t')) \geq 1-\exp(-t')$ by Lemma 12. □

Remind the definition (4) of $\alpha_1, \alpha_2, \beta_1, \beta_2$:

$$\alpha_1 = 3\left(\sum_{m=1}^M \frac{r_m^{-2s_m}}{n}\right)^{\frac{1}{2}}, \ \ \alpha_2 = 3\left\|\left(\frac{s_m r_m^{1-s_m}}{\sqrt{n}}\right)_{m=1}^M\right\|_{\psi^*},$$
$$\beta_1 = 3\left(\sum_{m=1}^M \frac{r_m^{-\frac{2s_m(3-s_m)}{1+s_m}}}{n^{\frac{2}{1+s_m}}}\right)^{\frac{1}{2}}, \ \ \beta_2 = 3\left\|\left(\frac{s_m r_m^{\frac{(1-s_m)^2}{1+s_m}}}{n^{\frac{1}{1+s_m}}}\right)_{m=1}^M\right\|_{\psi^*}, \tag{S-15}$$

for given reals $\{r_m\}_{m=1}^M$. The following theorem immediately gives Theorem 1.

Theorem 8. *Suppose Assumptions 1-4 are satisfied. Let $\{r_m\}_{m=1}^M$ be arbitrary positive reals that can depend on n, and assume $\lambda_1^{(n)} \geq \left(\frac{\alpha_2}{\alpha_1}\right)^2 + \left(\frac{\beta_2}{\beta_1}\right)^2$. Then for all n and t' that satisfy $\frac{\log(M)}{\sqrt{n}} \leq 1$ and $\frac{4\phi\sqrt{n}}{\kappa_M}\max\{\alpha_1^2, \beta_1^2, \frac{M\log(M)}{n}\}\eta(t') \leq \frac{1}{12}$ and for all $t \geq 1$, we have*

$$\|\hat{f} - f^*\|_{L_2(\Pi)}^2 \leq \frac{24\eta(t)^2\phi^2}{\kappa_M}\left(\alpha_1^2 + \beta_1^2 + \frac{M\log(M)}{n}\right) + 4\lambda_1^{(n)}\|f^*\|_\psi^2.$$

with probability $1-\exp(-t)-\exp(-t')$.

Proof of Theorem 8. By the assumption of the theorem, we can assume Lemma 7 holds, that is, the event $\mathscr{E}_1(t)\cap\mathscr{E}_2(t')$ occurs with probability $1-\exp(-t)-\exp(-t')$. Below we discuss on the event $\mathscr{E}_1(t)\cap\mathscr{E}_2(t')$.

Since $y_i = f^*(x_i) + \epsilon_i$, we have

$$\|\hat{f} - f^*\|_{L_2(\Pi)}^2 + \lambda_1^{(n)}\|\hat{f}\|_\psi^2$$
$$\leq (\|\hat{f} - f^*\|_{L_2(\Pi)}^2 - \|\hat{f} - f^*\|_n^2) + \frac{2}{n}\sum_{i=1}^n\sum_{m=1}^M \epsilon_i(\hat{f}_m(x_i) - f_m^*(x_i)) + \lambda_1^{(n)}\|f^*\|_\psi^2.$$

Here on the event $\mathscr{E}_2(t')$, the above inequality gives

$$\|\hat{f} - f^*\|_{L_2(\Pi)}^2 + \lambda_1^{(n)}\|\hat{f}\|_\psi^2$$
$$\leq \phi\sqrt{n}\left(\sum_{m=1}^M U_{n,s_m}^{(m)}(\hat{f}_m - f_m^*)\right)^2\eta(t') + \frac{2}{n}\sum_{i=1}^n\sum_{m=1}^M \epsilon_i(\hat{f}_m(x_i) - f_m^*(x_i)) + \lambda_1^{(n)}\|f^*\|_\psi^2. \tag{S-16}$$

Before we prove the statements, we show an upper bound of $\sum_{m=1}^M U_{n,s_m}^{(m)}(f_m)$ required in the proof. By definition, we have

$$
\begin{aligned}
&U_{n,s_m}^{(m)}(f_m) \\
=&3\left(\frac{r_m^{-s_m}}{\sqrt{n}} \vee \frac{r_m^{-\frac{s_m(3-s_m)}{1+s_m}}}{n^{\frac{1}{1+s_m}}}\right)\left(\|f_m\|_{L_2(\Pi)} + s_m r_m \|f_m\|_{\mathcal{H}_m}\right) + \sqrt{\frac{\log(M)}{n}}\|f_m\|_{L_2(\Pi)} \\
\leq&3\frac{r_m^{-s_m}}{\sqrt{n}}\left(\|f_m\|_{L_2(\Pi)} + s_m r_m \|f_m\|_{\mathcal{H}_m}\right) + 3\frac{r_m^{-\frac{s_m(3-s_m)}{1+s_m}}}{n^{\frac{1}{1+s_m}}}\left(\|f_m\|_{L_2(\Pi)} + s_m r_m \|f_m\|_{\mathcal{H}_m}\right) \qquad \text{(S-17)}\\
&+\sqrt{\frac{\log(M)}{n}}\|f_m\|_{L_2(\Pi)}. \qquad \text{(S-18)}
\end{aligned}
$$

Now the sum of the first term is bounded as

$$
\begin{aligned}
&\sum_{m=1}^M 3\frac{r_m^{-s_m}}{\sqrt{n}}\left(\|f_m\|_{L_2(\Pi)} + s_m r_m \|f_m\|_{\mathcal{H}_m}\right) \\
=&3\sum_{m=1}^M \frac{r_m^{-s_m}}{\sqrt{n}}\|f_m\|_{L_2(\Pi)} + 3\sum_{m=1}^M \frac{s_m r_m^{1-s_m}}{\sqrt{n}}\|f_m\|_{\mathcal{H}_m} \\
\leq&3\left(\sum_{m=1}^M \frac{r_m^{-2s_m}}{n}\right)^{\frac{1}{2}}\left(\sum_{m=1}^M \|f_m\|_{L_2(\Pi)}^2\right)^{\frac{1}{2}} + 3\left\|\left(\frac{s_m r_m^{1-s_m}}{\sqrt{n}}\right)_{m=1}^M\right\|_{\psi^*}\|f\|_\psi,
\end{aligned}
$$

where we used Cauchy-Schwarz inequality and the duality of the norm in the last inequality. The sum of the second term of the RHS of Eq. (S-18) is bounded as

$$
\begin{aligned}
&\sum_{m=1}^M 3\frac{r_m^{-\frac{s_m(3-s_m)}{1+s_m}}}{n^{\frac{1}{1+s_m}}}\left(\|f_m\|_{L_2(\Pi)} + s_m r_m \|f_m\|_{\mathcal{H}_m}\right) \\
=&3\sum_{m=1}^M \frac{r_m^{-\frac{s_m(3-s_m)}{1+s_m}}}{n^{\frac{1}{1+s_m}}}\|f_m\|_{L_2(\Pi)} + 3\sum_{m=1}^M \frac{s_m r_m^{\frac{(1-s_m)^2}{1+s_m}}}{n^{\frac{1}{1+s_m}}}\|f_m\|_{\mathcal{H}_m} \\
\leq&3\left(\sum_{m=1}^M \frac{r_m^{-\frac{2s_m(3-s_m)}{1+s_m}}}{n^{\frac{2}{1+s_m}}}\right)^{\frac{1}{2}}\left(\sum_{m=1}^M \|f_m\|_{L_2(\Pi)}^2\right)^{\frac{1}{2}} + 3\left\|\left(\frac{s_m r_m^{\frac{(1-s_m)^2}{1+s_m}}}{n^{\frac{1}{1+s_m}}}\right)_{m=1}^M\right\|_{\psi^*}\|f\|_\psi,
\end{aligned}
$$

where we used Cauchy-Schwarz inequality and the duality of the norm in the last inequality. Finally we have the following bound of the third term of the RHS of Eq. (S-18):

$$
\sum_{m=1}^M \sqrt{\frac{\log(M)}{n}}\|f_m\|_{L_2(\Pi)} \leq \sqrt{\frac{M\log(M)}{n}}\left(\sum_{m=1}^M \|f_m\|_{L_2(\Pi)}^2\right)^{\frac{1}{2}}.
$$

Combine these inequalities and the relation $\sum_{m=1}^M \|f_m\|_{L_2(\Pi)}^2 \leq \frac{1}{\kappa_M}\|f\|_{L_2(\Pi)}^2$ (Assumption 3) to obtain

$$
\begin{aligned}
&\sum_{m=1}^M U_{n,s_m}^{(m)}(f_m) \\
\leq&3\left(\sum_{m=1}^M \frac{r_m^{-2s_m}}{n}\right)^{\frac{1}{2}}\frac{\|f\|_{L_2(\Pi)}}{\sqrt{\kappa_M}} + 3\left\|\left(\frac{s_m r_m^{1-s_m}}{\sqrt{n}}\right)_{m=1}^M\right\|_{\psi^*}\|f\|_\psi \\
&+3\left(\sum_{m=1}^M \frac{r_m^{-\frac{2s_m(3-s_m)}{1+s_m}}}{n^{\frac{2}{1+s_m}}}\right)^{\frac{1}{2}}\frac{\|f\|_{L_2(\Pi)}}{\sqrt{\kappa_M}} + 3\left\|\left(\frac{s_m r_m^{\frac{(1-s_m)^2}{1+s_m}}}{n^{\frac{1}{1+s_m}}}\right)_{m=1}^M\right\|_{\psi^*}\|f\|_\psi
\end{aligned}
$$

$$+\sqrt{\frac{M\log(M)}{n}}\frac{\|f\|_{L_2(\Pi)}}{\sqrt{\kappa_M}}. \tag{S-19}$$

Then by the definition (4) of $\alpha_1, \alpha_2, \beta_1, \beta_2$, we have

$$\begin{aligned}
&\sum_{m=1}^{M} U^{(m)}_{n,s_m}(f_m) \\
\leq&\alpha_1 \frac{\|f\|_{L_2(\Pi)}}{\sqrt{\kappa_M}} + \alpha_2\|f\|_\psi + \beta_1\frac{\|f\|_{L_2(\Pi)}}{\sqrt{\kappa_M}} + \beta_2\|f\|_\psi + \sqrt{\frac{M\log(M)}{n}}\frac{\|f\|_{L_2(\Pi)}}{\sqrt{\kappa_M}}.
\end{aligned} \tag{S-20}$$

Step 1.

By Eq. (S-20), the first term on the RHS of Eq. (S-16) can be upper bounded as

$$\begin{aligned}
&\phi\sqrt{n}\left(\sum_{m=1}^{M} U^{(m)}_{n,s_m}(\hat{f}_m - f^*_m)\right)^2 \eta(t') \\
\leq&4\phi\sqrt{n}\Big(\alpha_1^2\frac{\|\hat{f}-f^*\|^2_{L_2(\Pi)}}{\kappa_M} + \alpha_2^2\|\hat{f}-f^*\|^2_\psi + \beta_1^2\frac{\|\hat{f}-f^*\|^2_{L_2(\Pi)}}{\kappa_M} + \\
&\qquad \beta_2^2\|\hat{f}-f^*\|^2_\psi + \frac{M\log(M)}{n}\frac{\|\hat{f}-f^*\|^2_{L_2(\Pi)}}{\kappa_M}\Big)\eta(t') \\
\leq&\frac{4\phi\sqrt{n}}{\kappa_M}\alpha_1^2\eta(t')\left(\|\hat{f}-f^*\|^2_{L_2(\Pi)} + \left(\frac{\alpha_2}{\alpha_1}\right)^2\|\hat{f}-f^*\|^2_\psi\right) \\
&+\frac{4\phi\sqrt{n}}{\kappa_M}\beta_1^2\eta(t')\left(\|\hat{f}-f^*\|^2_{L_2(\Pi)} + \left(\frac{\beta_2}{\beta_1}\right)^2\|\hat{f}-f^*\|^2_\psi\right) \\
&+\frac{4\phi\sqrt{n}}{\kappa_M}\frac{M\log(M)}{n}\eta(t')\|\hat{f}-f^*\|^2_{L_2(\Pi)}.
\end{aligned}$$

By assumption, we have $\frac{4\phi\sqrt{n}}{\kappa_M}\max\{\alpha_1^2, \beta_1^2, \frac{M\log(M)}{n}\}\eta(t') \leq \frac{1}{12}$. Hence the RHS of the above inequality is bounded by

$$\begin{aligned}
&\phi\sqrt{n}\left(\sum_{m=1}^{M} U^{(m)}_{n,s_m}(\hat{f}_m - f^*_m)\right)^2 \eta(t') \\
\leq&\frac{1}{4}\left\{\|\hat{f}-f^*\|^2_{L_2(\Pi)} + \left[\left(\frac{\alpha_2}{\alpha_1}\right)^2 + \left(\frac{\beta_2}{\beta_1}\right)^2\right]\|\hat{f}-f^*\|^2_\psi\right\}.
\end{aligned} \tag{S-21}$$

Step 2. On the event $\mathscr{E}_1(t)$, we have

$$\begin{aligned}
&\frac{2}{n}\sum_{i=1}^{n}\sum_{m=1}^{M}\epsilon_i(\hat{f}_m(x_i) - f^*_m(x_i)) \leq 2\sum_{m=1}^{M}\eta(t)\phi U^{(m)}_{n,s_m}(\hat{f}_m - f^*_m) \\
&\leq 2\eta(t)\phi\Bigg[\alpha_1\frac{\|\hat{f}-f^*\|_{L_2(\Pi)}}{\sqrt{\kappa_M}} + \alpha_2\|\hat{f}-f^*\|_\psi + \beta_1\frac{\|\hat{f}-f^*\|_{L_2(\Pi)}}{\sqrt{\kappa_M}} + \beta_2\|\hat{f}-f^*\|_\psi \\
&\quad + \sqrt{\frac{M\log(M)}{n}}\frac{\|\hat{f}-f^*\|_{L_2(\Pi)}}{\sqrt{\kappa_M}}\Bigg] \qquad (\because \text{Eq.(S-19)}) \\
&\leq 2\frac{\eta(t)\phi\alpha_1}{\sqrt{\kappa_M}}\left(\|\hat{f}-f^*\|_{L_2(\Pi)} + \frac{\alpha_2}{\alpha_1}\|\hat{f}-f^*\|_\psi\right) + 2\frac{\eta(t)\phi\beta_1}{\sqrt{\kappa_M}}\left(\|\hat{f}-f^*\|_{L_2(\Pi)} + \frac{\beta_2}{\beta_1}\|\hat{f}-f^*\|_\psi\right) \\
&\quad + 2\frac{\eta(t)\phi}{\sqrt{\kappa_M}}\sqrt{\frac{M\log(M)}{n}}\|\hat{f}-f^*\|_{L_2(\Pi)}
\end{aligned}$$

$$
\begin{aligned}
&\leq \frac{12\eta(t)^2\phi^2\alpha_1^2}{\kappa_M} + \frac{1}{24}\left(\|\hat{f}-f^*\|_{L_2(\Pi)} + \frac{\alpha_2}{\alpha_1}\|\hat{f}-f^*\|_{\psi}\right)^2 \\
&\quad + \frac{12\eta(t)^2\phi^2\beta_1^2}{\kappa_M} + \frac{1}{24}\left(\|\hat{f}-f^*\|_{L_2(\Pi)} + \frac{\beta_2}{\beta_1}\|\hat{f}-f^*\|_{\psi}\right)^2 \\
&\quad + \frac{6\eta(t)^2\phi^2}{\kappa_M}\frac{M\log(M)}{n} + \frac{1}{12}\|\hat{f}-f^*\|_{L_2(\Pi)}^2 \\
&\leq \frac{12\eta(t)^2\phi^2\alpha_1^2}{\kappa_M} + \frac{1}{12}\left[\|\hat{f}-f^*\|_{L_2(\Pi)}^2 + \left(\frac{\alpha_2}{\alpha_1}\right)^2\|\hat{f}-f^*\|_{\psi}^2\right] \\
&\quad + \frac{12\eta(t)^2\phi^2\beta_1^2}{\kappa_M} + \frac{1}{12}\left[\|\hat{f}-f^*\|_{L_2(\Pi)}^2 + \left(\frac{\beta_2}{\beta_1}\right)^2\|\hat{f}-f^*\|_{\psi}^2\right] \\
&\quad + \frac{6\eta(t)^2\phi^2}{\kappa_M}\frac{M\log(M)}{n} + \frac{1}{12}\|\hat{f}-f^*\|_{L_2(\Pi)}^2 \\
&\leq \frac{12\eta(t)^2\phi^2}{\kappa_M}\left(\alpha_1^2+\beta_1^2+\frac{M\log(M)}{n}\right) + \frac{1}{4}\left\{\|\hat{f}-f^*\|_{L_2(\Pi)}^2 + \left[\left(\frac{\alpha_2}{\alpha_1}\right)^2 + \left(\frac{\beta_2}{\beta_1}\right)^2\right]\|\hat{f}-f^*\|_{\psi}^2\right\}.
\end{aligned} \tag{S-22}
$$

Step 3.

Substituting the inequalities (S-21) and (S-22) to Eq. (S-16), we obtain

$$
\begin{aligned}
&\|\hat{f}-f^*\|_{L_2(\Pi)}^2 + \lambda_1^{(n)}\|\hat{f}\|_{\psi}^2 \\
\leq &\frac{12\eta(t)^2\phi^2}{\kappa_M}\left(\alpha_1^2+\beta_1^2+\frac{M\log(M)}{n}\right) + \frac{1}{2}\left\{\|\hat{f}-f^*\|_{L_2(\Pi)}^2 + \left[\left(\frac{\alpha_2}{\alpha_1}\right)^2 + \left(\frac{\beta_2}{\beta_1}\right)^2\right]\|\hat{f}-f^*\|_{\psi}^2\right\} \\
&+ \lambda_1^{(n)}\|f^*\|_{\psi}^2.
\end{aligned} \tag{S-23}
$$

Now the term $\|\hat{f}-f^*\|_{\psi}^2$ can be bounded as

$$
\|\hat{f}-f^*\|_{\psi}^2 \leq \left(\|\hat{f}\|_{\psi} + \|f^*\|_{\psi}\right)^2 \leq 2\left(\|\hat{f}\|_{\psi}^2 + \|f^*\|_{\psi}^2\right),
$$

where we used the triangular inequality for the mixed-norm with respect to ψ-norm $\|\cdot\|_{\psi}$. Thus, when $\lambda_1^{(n)} \geq \left(\frac{\alpha_2}{\alpha_1}\right)^2 + \left(\frac{\beta_2}{\beta_1}\right)^2$, Eq. (S-23) yields

$$
\frac{1}{2}\|\hat{f}-f^*\|_{L_2(\Pi)}^2 \leq \frac{12\eta(t)^2\phi^2}{\kappa_M}\left(\alpha_1^2+\beta_1^2+\frac{M\log(M)}{n}\right) + 2\lambda_1^{(n)}\|f^*\|_{\psi}^2.
$$

Therefore by multiplying 2 to both sides, we have

$$
\|\hat{f}-f^*\|_{L_2(\Pi)}^2 \leq \frac{24\eta(t)^2\phi^2}{\kappa_M}\left(\alpha_1^2+\beta_1^2+\frac{M\log(M)}{n}\right) + 4\lambda_1^{(n)}\|f^*\|_{\psi}^2.
$$

This gives the assertion. □

E Bounding the Probabilities of $\mathscr{E}_1(t)$ and $\mathscr{E}_2(t')$

Here we derive bounds of the probabilities of the events $\mathscr{E}_1(t)$ and $\mathscr{E}_2(t')$ (see Eq. (S-13) and Eq. (S-14) for their definitions). The goal of this section is to derive Lemmas 12 and 13.

Using Propositions 6 and 5, we obtain the following ratio type uniform bound.

Lemma 9. *Under the Spectral Assumption (Assumption 2) and the Embedded Assumption (Assumption 4), there exists a constant C_{s_m} depending only on s_m, c and C_1 such that*

$$
\mathrm{E}\left[\sup_{f_m\in\mathcal{H}_m:\|f_m\|_{\mathcal{H}_m}=1}\frac{|\frac{1}{n}\sum_{i=1}^n\sigma_i f_m(x_i)|}{U_{n,s_m}^{(m)}(f_m)}\right] \leq C_{s_m}.
$$

Proof of Lemma 9. Let $\mathcal{H}_m(\delta) := \{f_m \in \mathcal{H}_m \mid \|f_m\|_{\mathcal{H}_m} = 1, \|f_m\|_{L_2(\Pi)} \leq \delta\}$ and $z = 2^{1/s_m} > 1$. Define $\tau := s_m r_m$. Then by combining Propositions 4 and 5 with Assumption 4, we have

$$
\begin{aligned}
&\mathrm{E}\left[\sup_{f_m \in \mathcal{H}_m : \|f_m\|_{\mathcal{H}_m}=1} \frac{|\frac{1}{n}\sum_{i=1}^n \sigma_i f_m(x_i)|}{U^{(m)}_{n,s_m}(f_m)}\right] \\
\leq &\mathrm{E}\left[\sup_{f_m \in \mathcal{H}_m(\tau)} \frac{|\frac{1}{n}\sum_{i=1}^n \sigma_i f_m(x_i)|}{U^{(m)}_{n,s_m}(f_m)}\right] + \sum_{k=1}^{\infty} \mathrm{E}\left[\sup_{f_m \in \mathcal{H}_m(\tau z^k)\setminus \mathcal{H}_m(\tau z^{k-1})} \frac{|\frac{1}{n}\sum_{i=1}^n \sigma_i f_m(x_i)|}{U^{(m)}_{n,s_m}(f_m)}\right] \\
\leq &C'_{s_m} \frac{\frac{\tau^{1-s_m}\tilde{c}_{s_m}^{s_m}}{\sqrt{n}}}{3\frac{r_m^{-s_m}}{\sqrt{n}} s_m r_m} \vee \frac{\frac{C_1^{\frac{1-s_m}{1+s_m}} \tau^{\frac{(1-s_m)^2}{1+s_m}} \tilde{c}_{s_m}^{\frac{2s_m}{1+s_m}}}{n^{\frac{1}{1+s_m}}}}{3\frac{r_m^{-\frac{s_m(3-s_m)}{1+s_m}}}{n^{\frac{1}{1+s_m}}} s_m r_m} \\
&+ \sum_{k=1}^{\infty} C'_{s_m} \frac{\frac{z^{k(1-s_m)}\tau^{1-s_m}\tilde{c}_{s_m}^{s_m}}{\sqrt{n}}}{3\frac{r_m^{-s_m}}{\sqrt{n}} \tau z^{k-1}} \vee \frac{\frac{C_1^{\frac{1-s_m}{1+s_m}} z^{k\frac{(1-s_m)^2}{1+s_m}} \tau^{\frac{(1-s_m)^2}{1+s_m}} \tilde{c}_{s_m}^{\frac{2s_m}{1+s_m}}}{n^{\frac{1}{1+s_m}}}}{3\frac{r_m^{-\frac{s_m(3-s_m)}{1+s_m}}}{n^{\frac{1}{1+s_m}}} \tau z^{k-1}} \\
&\leq \frac{C'_{s_m}}{3}\left(s_m^{-s_m}\tilde{c}_{s_m}^{s_m} \vee s_m^{-3s_m} C_1^{\frac{1-s_m}{1+s_m}} \tilde{c}_{s_m}^{\frac{2s_m}{1+s_m}}\right)\left(1 + \sum_{k=1}^{\infty} z^{1-ks_m} \vee z^{1-k\frac{s_m(3-s_m)}{1+s_m}}\right) \\
&= \frac{C'_{s_m} s_m^{-3s_m}}{3}\left(\tilde{c}_{s_m}^{s_m} \vee C_1^{\frac{1-s_m}{1+s_m}} \tilde{c}_{s_m}^{\frac{2s_m}{1+s_m}}\right)\left(1 + \frac{z^{1-s_m}}{1-z^{-s_m}} \vee \frac{z^{1-\frac{s_m(3-s_m)}{1+s_m}}}{1-z^{-\frac{s_m(3-s_m)}{1+s_m}}}\right) \\
&\leq 9C'_{s_m}\left(\tilde{c}_{s_m}^{s_m} \vee C_1^{\frac{1-s_m}{1+s_m}} \tilde{c}_{s_m}^{\frac{2s_m}{1+s_m}}\right)\left(1 + \frac{z^{1-s_m}}{1-z^{-s_m}} \vee \frac{z^{1-\frac{s_m(3-s_m)}{1+s_m}}}{1-z^{-\frac{s_m(3-s_m)}{1+s_m}}}\right),
\end{aligned}
$$

where we used $s_m^{-s_m} \leq 3$ for $0 < s_m$ in the last line. Thus by setting, $C_{s_m} = 9C'_{s_m}\left(\tilde{c}_{s_m}^{s_m} \vee C_1^{\frac{1-s_m}{1+s_m}} \tilde{c}_{s_m}^{\frac{2s_m}{1+s_m}}\right)\left(1 + \frac{z^{1-s_m}}{1-z^{-s_m}} \vee \frac{z^{1-\frac{s_m(3-s_m)}{1+s_m}}}{1-z^{-\frac{s_m(3-s_m)}{1+s_m}}}\right)$, we obtain the assertion. □

This lemma immediately gives the following corollary.

Corollary 10. *Under the Spectral Assumption (Assumption 2) and the Embedded Assumption (Assumption 4), there exists a constant C_{s_m} depending only on s_m, c and C_1 such that*

$$
\mathrm{E}\left[\sup_{f_m \in \mathcal{H}_m} \frac{|\frac{1}{n}\sum_{i=1}^n \sigma_i f_m(x_i)|}{U^{(m)}_{n,s_m}(f_m)}\right] \leq C_{s_m}.
$$

Proof. By dividing the denominator and the numerator by the RKHS norm $\|f_m\|_{\mathcal{H}_m}$, we have

$$
\begin{aligned}
&\mathrm{E}\left[\sup_{f_m \in \mathcal{H}_m} \frac{|\frac{1}{n}\sum_{i=1}^n \sigma_i f_m(x_i)|}{U^{(m)}_{n,s_m}(f_m)}\right] \\
=&\mathrm{E}\left[\sup_{f_m \in \mathcal{H}_m} \frac{|\frac{1}{n}\sum_{i=1}^n \sigma_i f_m(x_i)|/\|f_m\|_{\mathcal{H}_m}}{U^{(m)}_{n,s_m}(f_m)/\|f_m\|_{\mathcal{H}_m}}\right] \\
=&\mathrm{E}\left[\sup_{f_m \in \mathcal{H}_m} \frac{|\frac{1}{n}\sum_{i=1}^n \sigma_i f_m(x_i)/\|f_m\|_{\mathcal{H}_m}|}{U^{(m)}_{n,s_m}(f_m/\|f_m\|_{\mathcal{H}_m})}\right] \\
=&\mathrm{E}\left[\sup_{f_m \in \mathcal{H}_m : \|f_m\|_{\mathcal{H}_m}=1} \frac{|\frac{1}{n}\sum_{i=1}^n \sigma_i f_m(x_i)|}{U^{(m)}_{n,s_m}(f_m)}\right] \\
\leq &C_{s_m}. \qquad (\because \text{Lemma 9})
\end{aligned}
$$

□

Lemma 11. *If* $\frac{\log(M)}{\sqrt{n}} \leq 1$*, then under the Spectral Assumption (Assumption 2) and the Embedded Assumption (Assumption 4) there exists a constant* $\tilde{C}_*$ *depending only on* $\{s_m\}_{m=1}^M$*,* c*,* C_1 *such that*

$$\mathrm{E}\left[\max_m \sup_{f_m \in \mathcal{H}_m} \frac{|\frac{1}{n}\sum_{i=1}^n \sigma_i f_m(x_i)|}{U_{n,s_m}^{(m)}(f_m)}\right] \leq \tilde{C}_*.$$

Proof of Lemma 11. First notice that the $L_2(\Pi)$-norm and the ∞-norm of $\frac{\sigma_i f_m(x_i)}{U_{n,s_m}^{(m)}(f_m)}$ can be evaluated by

$$\left\|\frac{\sigma_i f_m(x_i)}{U_{n,s_m}^{(m)}(f_m)}\right\|_{L_2(\Pi)} = \frac{\|f_m\|_{L_2(\Pi)}}{U_{n,s_m}^{(m)}(f_m)} \leq \frac{\|f_m\|_{L_2(\Pi)}}{\sqrt{\frac{\log(M)}{n}}\|f_m\|_{L_2(\Pi)}} \leq \sqrt{\frac{n}{\log(M)}}, \tag{S-24}$$

$$\left\|\frac{\sigma_i f_m(x_i)}{U_{n,s_m}^{(m)}(f_m)}\right\|_{\infty} = \frac{\|f_m\|_\infty}{U_{n,s_m}^{(m)}(f_m)} \leq \frac{C_1\|f_m\|_{L_2(\Pi)}^{1-s_m}\|f_m\|_{\mathcal{H}_m}^{s_m}}{U_{n,s_m}^{(m)}(f_m)} \leq \frac{C_1}{3}\sqrt{n} \leq C_1\sqrt{n}, \tag{S-25}$$

where the second line is shown by using the relation (S-11). Let $C_* := \max_m C_{s_m}$ where C_{s_m} is the constant appeared in Lemma 9. Thus Talagrand's inequality and Corollary 10 imply

$$\begin{aligned}
&P\left(\max_m \sup_{f_m \in \mathcal{H}_m} \frac{|\frac{1}{n}\sum_{i=1}^n \sigma_i f_m(x_i)|}{U_{n,s_m}^{(m)}(f_m)} \geq K\left[C_* + \sqrt{\frac{t}{\log(M)}} + \frac{C_1 t}{\sqrt{n}}\right]\right)\\
\leq&\sum_{m=1}^M P\left(\sup_{f_m \in \mathcal{H}_m} \frac{|\frac{1}{n}\sum_{i=1}^n \sigma_i f_m(x_i)|}{U_{n,s_m}^{(m)}(f_m)} \geq K\left[C_* + \sqrt{\frac{t}{\log(M)}} + \frac{C_1 t}{\sqrt{n}}\right]\right)\\
\leq&\sum_{m=1}^M P\left(\sup_{f_m \in \mathcal{H}_m} \frac{|\frac{1}{n}\sum_{i=1}^n \sigma_i f_m(x_i)|}{U_{n,s_m}^{(m)}(f_m)} \geq K\left[C_{s_m} + \sqrt{\frac{t}{\log(M)}} + \frac{C_1 t}{\sqrt{n}}\right]\right)\\
\leq&Me^{-t}.
\end{aligned}$$

By setting $t \leftarrow t + \log(M)$, we obtain

$$P\left(\max_m \sup_{f_m \in \mathcal{H}_m} \frac{|\frac{1}{n}\sum_{i=1}^n \sigma_i f_m(x_i)|}{U_{n,s_m}^{(m)}(f_m)} \geq K\left[C_* + \sqrt{\frac{t+\log(M)}{\log(M)}} + \frac{C_1(t+\log(M))}{\sqrt{n}}\right]\right) \leq e^{-t}$$

for all $t \geq 0$. Consequently the expectation of the max-sup term can be bounded as

$$\begin{aligned}
&\mathrm{E}\left[\max_m \sup_{f_m \in \mathcal{H}_m} \frac{|\frac{1}{n}\sum_{i=1}^n \sigma_i f_m(x_i)|}{U_{n,s_m}^{(m)}(f_m)}\right]\\
\leq&K\left[C_* + 1 + \frac{C_1\log(M)}{\sqrt{n}}\right] + \int_0^\infty K\left[C_* + \sqrt{\frac{t+1+\log(M)}{\log(M)}} + \frac{C_1(t+1+\log(M))}{\sqrt{n}}\right]e^{-t}\mathrm{d}t\\
\leq&2K\left[C_* + \sqrt{2} + \sqrt{\frac{\pi}{4\log(M)}} + \frac{C_1(2+\log(M))}{\sqrt{n}}\right] \leq \tilde{C}_*,
\end{aligned}$$

where we used $\sqrt{t+1+\log(M)} \leq \sqrt{t} + \sqrt{1+\log(M)}$ and $\int_0^\infty \sqrt{t}e^{-t}\mathrm{d}t = \sqrt{\frac{\pi}{4}}$, $\frac{\log(M)}{\sqrt{n}} \leq 1$, and $\tilde{C}_* = 2K[C_* + \sqrt{2} + \sqrt{\frac{\pi}{4}} + 3C_1]$. □

Lemma 12. *Suppose the Basic Assumption (Assumption 1), the Spectral Assumption (Assumption 2) and the Embedded Assumption (Assumption 4) hold. Define* $\bar{\phi} = KL\left[2\tilde{C}_* + 1 + C_1\right]$*. If* $\frac{\log(M)}{\sqrt{n}} \leq 1$*, then the following holds*

$$P\left(\max_m \sup_{f_m \in \mathcal{H}_m} \frac{|\frac{1}{n}\sum_{i=1}^n \epsilon_i f_m(x_i)|}{U_{n,s_m}^{(m)}(f_m)} \geq \bar{\phi}\eta(t)\right) \leq e^{-t}.$$

Proof of Lemma 12. By the contraction inequality [2, Theorem 4.12] and Lemma 11, we have

$$\mathrm{E}\left[\max_m \sup_{f_m\in\mathcal{H}_m} \frac{|\frac{1}{n}\sum_{i=1}^n \epsilon_i f_m(x_i)|}{U^{(m)}_{n,s_m}(f_m)}\right] \le 2\mathrm{E}\left[\max_m \sup_{f_m\in\mathcal{H}_m} \frac{|\frac{1}{n}\sum_{i=1}^n \sigma_i\epsilon_i f_m(x_i)|}{U^{(m)}_{n,s_m}(f_m)}\right] \le 2L\tilde{C}_*,$$

where we used $\epsilon_i \le L$ (Basic Assumption). Using this and Eq. (S-24) and Eq. (S-25), Talgrand's inequality gives

$$P\left(\max_m \sup_{f_m\in\mathcal{H}_m} \frac{|\frac{1}{n}\sum_{i=1}^n \epsilon_i f_m(x_i)|}{U^{(m)}_{n,s_m}(f_m)} \ge KL\left[2\tilde{C}_* + \sqrt{t} + \frac{C_1 t}{\sqrt{n}}\right]\right) \le e^{-t}.$$

Thus we have

$$P\left(\max_m \sup_{f_m\in\mathcal{H}_m} \frac{|\frac{1}{n}\sum_{i=1}^n \epsilon_i f_m(x_i)|}{U^{(m)}_{n,s_m}(f_m)} \ge KL\left[2\tilde{C}_* + 1 + C_1\right]\max\left(1,\sqrt{t},\frac{t}{\sqrt{n}}\right)\right) \le e^{-t}.$$

Therefore by the definition of $\bar{\phi}$ and $\eta(t)$, we obtain the assertion.

$\square$

Lemma 13. *Suppose the Basic Assumption (Assumption 1), the Spectral Assumption (Assumption 2) and the Embedded Assumption (Assumption 4) hold. Let $\bar{\phi}' = K[2C_1\tilde{C}_* + C_1 + C_1^2]$. Then, if $\frac{\log(M)}{\sqrt{n}} \le 1$, we have for all $t \ge 0$*

$$\left|\left\|\sum_{m=1}^M f_m\right\|_n^2 - \left\|\sum_{m=1}^M f_m\right\|_{L_2(\Pi)}^2\right| \le \phi'\sqrt{n}\left(\sum_{m=1}^M U^{(m)}_{n,s_m}(f_m)\right)^2 \eta(t),$$

for all $f_m \in \mathcal{H}_m$ $(m = 1,\dots,M)$ with probability $1-\exp(-t)$.

Proof of Lemma 13.

$$\begin{aligned}
&\mathrm{E}\left[\sup_{f_m\in\mathcal{H}_m} \frac{\left|\left\|\sum_{m=1}^M f_m\right\|_n^2 - \left\|\sum_{m=1}^M f_m\right\|_{L_2(\Pi)}^2\right|}{\left(\sum_{m=1}^M U^{(m)}_{n,s_m}(f_m)\right)^2}\right] \\
\le& 2\mathrm{E}\left[\sup_{f_m\in\mathcal{H}_m} \frac{\left|\frac{1}{n}\sum_{i=1}^n \sigma_i(\sum_{m=1}^M f_m(x_i))^2\right|}{\left(\sum_{m=1}^M U^{(m)}_{n,s_m}(f_m)\right)^2}\right] \\
\le& \sup_{f_m\in\mathcal{H}_m} \frac{\left\|\sum_{m=1}^M f_m\right\|_\infty}{\sum_{m=1}^M U^{(m)}_{n,s_m}(f_m)} \times 2\mathrm{E}\left[\sup_{f_m\in\mathcal{H}_m} \frac{\left|\frac{1}{n}\sum_{i=1}^n \sigma_i(\sum_{m=1}^M f_m(x_i))\right|}{\sum_{m=1}^M U^{(m)}_{n,s_m}(f_m)}\right],
\end{aligned} \tag{S-26}$$

where we used the contraction inequality in the last line [2, Theorem 4.12]. Thus using Eq. (S-25), the RHS of the inequality (S-26) can be bounded as

$$\begin{aligned}
&2C_1\sqrt{n}\mathrm{E}\left[\sup_{f_m\in\mathcal{H}_m} \frac{\left|\frac{1}{n}\sum_{i=1}^n \sigma_i(\sum_{m=1}^M f_m(x_i))\right|}{\sum_{m=1}^M U^{(m)}_{n,s_m}(f_m)}\right] \\
\le& 2C_1\sqrt{n}\mathrm{E}\left[\sup_{f_m\in\mathcal{H}_m} \max_m \frac{\left|\frac{1}{n}\sum_{i=1}^n \sigma_i f_m(x_i)\right|}{U^{(m)}_{n,s_m}(f_m)}\right],
\end{aligned}$$

where we used the relation

$$\frac{\sum_m a_m}{\sum_m b_m} \le \max_m \left(\frac{a_m}{b_m}\right) \tag{S-27}$$

for all $a_m \geq 0$ and $b_m \geq 0$ with a convention $\frac{0}{0} = 0$. By Lemma 11, the right hand side is upper bounded by $2C_1\sqrt{n}\tilde{C}_*$. Here we again apply Talagrand's concentration inequality, then we have

$$P\left(\sup_{f_m \in \mathcal{H}_m} \frac{\left|\left\|\sum_{m=1}^M f_m\right\|_n^2 - \left\|\sum_{m=1}^M f_m\right\|_{L_2(\Pi)}^2\right|}{\left(\sum_{m=1}^M U_{n,s_m}^{(m)}(f_m)\right)^2} \geq K\left[2C_1\tilde{C}_*\sqrt{n} + \sqrt{tn}C_1 + C_1^2 t\right]\right) \leq e^{-t},$$

where we substituted the following upper bounds of B and U.

$$\begin{aligned}
B &\leq \sup_{f_m \in \mathcal{H}_m} \mathrm{E}\left[\left(\frac{(\sum_{m=1}^M f_m)^2}{\left(\sum_{m=1}^M U_{n,s_m}^{(m)}(f_m)\right)^2}\right)^2\right] \\
&\leq \sup_{f_m \in \mathcal{H}_m} \mathrm{E}\left[\frac{(\sum_{m=1}^M f_m)^2}{\left(\sum_{m=1}^M U_{n,s_m}^{(m)}(f_m)\right)^2} \frac{(\|\sum_{m=1}^M f_m\|_\infty)^2}{\left(\sum_{m=1}^M U_{n,s_m}^{(m)}(f_m)\right)^2}\right] \\
&\overset{\text{(S-25)}}{\leq} \sup_{f_m \in \mathcal{H}_m} \frac{\left(\sum_{m=1}^M \|f_m\|_{L_2(\Pi)}\right)^2}{\left(\sum_{m=1}^M U_{n,s_m}^{(m)}(f_m)\right)^2} \frac{(\sum_{m=1}^M C_1\sqrt{n}U_{n,s_m}^{(m)}(f_m))^2}{\left(\sum_{m=1}^M U_{n,s_m}^{(m)}(f_m)\right)^2} \\
&\overset{\text{(S-24)}}{\leq} C_1^2 n^2 \frac{1}{\log(M)} \leq C_1^2 n^2,
\end{aligned}$$

where in the second inequality we used the relation

$$\mathrm{E}[(\sum_{m=1}^M f_m)^2] = \mathrm{E}[\sum_{m,m'=1}^M f_m f_{m'}] \leq \sum_{m,m'=1}^M \|f_m\|_{L_2(\Pi)}\|f_{m'}\|_{L_2(\Pi)} = (\sum_{m=1}^M \|f_m\|_{L_2(\Pi)})^2$$

and in the third and forth inequality we used Eq. (S-25) and Eq. (S-24) with Eq.(S-27) respectively. Here we again use Eq. (S-24) with Eq.(S-27) to obtain

$$U = \sup_{f_m \in \mathcal{H}_m} \left\|\frac{(\sum_{m=1}^M f_m)^2}{\left(\sum_{m=1}^M U_{n,s_m}^{(m)}(f_m)\right)^2}\right\|_\infty \leq C_1^2 n.$$

Therefore the above inequality implies the following inequality

$$\sup_{f_m \in \mathcal{H}_m} \frac{\left|\left\|\sum_{m=1}^M f_m\right\|_n^2 - \left\|\sum_{m=1}^M f_m\right\|_{L_2(\Pi)}^2\right|}{\left(\sum_{m=1}^M U_{n,s_m}^{(m)}(f_m)\right)^2} \leq K\left[2C_1\tilde{C}_s + C_1 + C_1^2\right]\sqrt{n}\max(1, \sqrt{t}, t/\sqrt{n}),$$

with probability $1 - \exp(-t)$. Remind $\bar{\phi}' = K\left[2C_1\tilde{C}_* + C_1 + C_1^2\right]$, then we obtain the assertion. □

F Proof of Theorem 3 (minimax learning rate)

Let *the δ-packing number* $Q(\delta, \mathcal{H}, L_2(\Pi))$ of a function class $\mathcal{H}$ be the largest number of functions $\{f_1, \ldots, f_Q\} \subseteq \mathcal{H}$ such that $\|f_i - f_j\|_{L_2(\Pi)} \geq \delta$ for all $i \neq j$.

Proof of Theorem 3. The proof utilizes the techniques developed by [3, 4] that applied the information theoretic technique developed by [7] to the MKL settings. To simplify the notation, we write $\mathcal{F} := \mathcal{H}_\psi(R)$, $N(\varepsilon, \mathcal{H}) := N(\varepsilon, \mathcal{H}, L_2(\Pi))$ and $Q(\varepsilon, \mathcal{H}) := Q(\varepsilon, \mathcal{H}, L_2(\Pi))$. It can be easily shown that $Q(2\varepsilon, \mathcal{F}) \leq N(2\varepsilon, \mathcal{F}) \leq Q(\varepsilon, \mathcal{F})$. Here due to Theorem 15 of [29], Assumption 5 yields

$$\log N(\varepsilon, \tilde{\mathcal{H}}(1)) \sim \varepsilon^{-2s}. \tag{S-28}$$

We utilize the following inequality given by Lemma 3 of [3]:

$$\min_{\hat f} \max_{f^*\in\mathcal{H}_\psi(R_p)} \mathrm{E}\|\hat f - f^*\|^2_{L_2(\Pi)} \geq \frac{\delta_n^2}{4}\left(1-\frac{\log N(\varepsilon_n,\mathcal{F}) + n\varepsilon_n^2/2\sigma^2 + \log 2}{\log Q(\delta_n,\mathcal{F})}\right).$$

First we show the assertion for the ℓ_∞-norm ball: $\mathcal{H}_\psi(R) = \mathcal{H}_{\ell_\infty}(R) := \left\{f=\sum_{m=1}^M f_m \;\middle|\; \max_{1\leq m\leq M}\|f_m\|_{\mathcal{H}_m}\leq R\right\}$. In this situation, there is a constant C that depends only s such that

$$\log Q(\delta,\mathcal{F}) \geq CM\log Q(\delta/\sqrt{M},\widetilde{\mathcal{H}}(R)),\quad \log N(\varepsilon,\mathcal{F}) \leq M\log N(\varepsilon/\sqrt{M},\widetilde{\mathcal{H}}(R)),$$

(this is shown in Lemma 5 of [4], but we give the proof in Lemma 14 for completeness). Using this expression, the minimax-learning rate is bounded as

$$\min_{\hat f} \max_{f^*\in\mathcal{H}_{\ell_p}(R_p)} \mathrm{E}\|\hat f - f^*\|^2_{L_2(\Pi)} \geq \frac{\delta_n^2}{4}\left(1-\frac{M\log N(\varepsilon_n/\sqrt{M},\widetilde{\mathcal{H}}(R)) + n\varepsilon_n^2/2\sigma^2 + \log 2}{CM\log Q(\delta_n/\sqrt{M},\widetilde{\mathcal{H}}(R))}\right).$$

Here we choose ε_n and δ_n to satisfy the following relations:

$$\frac{n}{2\sigma^2}\varepsilon_n^2 \leq M\log N\left(\varepsilon_n/\sqrt{M},\widetilde{\mathcal{H}}(R)\right), \tag{S-29}$$

$$M\log N\left(\varepsilon_n/\sqrt{M},\widetilde{\mathcal{H}}(R)\right) \geq \log 2, \tag{S-30}$$

$$4\log N\left(\epsilon_n/\sqrt{M},\widetilde{\mathcal{H}}(R)\right) \leq C\log Q\left(\delta_n/\sqrt{M},\widetilde{\mathcal{H}}(R)\right). \tag{S-31}$$

With ε_n and δ_n that satisfy the above relations (S-29) and (S-31), we have

$$\min_{\hat f} \max_{f^*\in\mathcal{H}_{\ell_p}(R_p)} \mathrm{E}\|\hat f - f^*\|^2_{L_2(\Pi)} \geq \frac{\delta_n^2}{16}. \tag{S-32}$$

By Eq. (S-28), the relation (S-29) can be rewritten as

$$\frac{n}{2\sigma^2}\varepsilon_n^2 \leq CM\left(\frac{\varepsilon_n}{R\sqrt{M}}\right)^{-2s}.$$

It is sufficient to impose

$$\varepsilon_n^2 \leq Cn^{-\frac{1}{1+s}}MR^{\frac{2s}{1+s}}, \tag{S-33}$$

with a constant C. Since we have assumed that $n > \frac{\bar c^2 M^2}{R^2\|\mathbf{1}\|^2_{\psi^*}}$ $(= \frac{1}{R^2}$ for $\|\cdot\|_\psi = \|\cdot\|_{\ell_\infty})$, the conditions (S-30) can be satisfied if the constant C in Eq. (S-33) is taken sufficiently small so that we have

$$\log 2 \leq \log N(\varepsilon_n/\sqrt{M},\widetilde{\mathcal{H}}(R)) \sim \left(\frac{\varepsilon_n}{R\sqrt{M}}\right)^{-2s}. \tag{S-34}$$

The relation (S-31) can be satisfied by taking $\delta_n = c\varepsilon_n$ with an appropriately chosen constant c. Thus Eq. (S-32) gives

$$\min_{\hat f} \max_{f^*\in\mathcal{H}_{\ell_p}(R_p)} \mathrm{E}\|\hat f - f^*\|^2_{L_2(\Pi)} \geq Cn^{-\frac{1}{1+s}}MR^{\frac{2s}{1+s}}, \tag{S-35}$$

with a constant C. This gives the assertion for $p=\infty$.

Finally we show the assertion for general isotropic ψ-norm $\|\cdot\|_\psi$. To show that, we prove that $\mathcal{H}_{\ell_\infty}(R\|\mathbf{1}\|_{\psi^*}/(\bar c M)) \subset \mathcal{H}_\psi(R)$. This is true if $\frac{R\|\mathbf{1}\|_{\psi^*}}{\bar c M}\mathbf{1} \in \mathcal{H}_\psi(R)$ because of the second condition of the definition (11) of isotropic property. By the isotropic property, the ψ-norm of $\frac{R\|\mathbf{1}\|_{\psi^*}}{\bar c M}\mathbf{1}$ is bounded as

$$\left\|\frac{R\|\mathbf{1}\|_{\psi^*}}{\bar c M}\mathbf{1}\right\|_\psi = \frac{R\|\mathbf{1}\|_{\psi^*}}{\bar c M}\|\mathbf{1}\|_\psi \overset{\text{isotropic}}{\leq} \frac{R}{\bar c M}\bar c M = R.$$

Thus we have $\frac{R\|\mathbf{1}\|_{\psi^*}}{\bar c M}\mathbf{1} \in \mathcal{H}_\psi(R)$ and thus $\mathcal{H}_{\ell_\infty}(R\|\mathbf{1}\|_{\psi^*}/(\bar c M)) \subset \mathcal{H}_\psi(R)$. Therefore we have

$$\min_{\hat f} \max_{f^*\in\mathcal{H}_\psi(R)} \mathrm{E}\|\hat f - f^*\|^2_{L_2(\Pi)} \geq \min_{\hat f} \max_{f^*\in\mathcal{H}_{\ell_\infty}(R\|\mathbf{1}\|_{\psi^*}/(\bar c M))} \mathrm{E}\|\hat f - f^*\|^2_{L_2(\Pi)}$$

$$\geq Cn^{-\frac{1}{1+s}} M \left(\frac{R\|\mathbf{1}\|_{\psi^*}}{\bar{c}M} \right)^{\frac{2s}{1+s}}, \qquad (\because \text{Eq. (S-35)}).$$

Note that due to the condition $n > \frac{\bar{c}^2 M^2}{R^2 \|\mathbf{1}\|_{\psi^*}^2}$, Eq. (S-35) is still valid under the condition that $\frac{R\|\mathbf{1}\|_{\psi^*}}{\bar{c}M}$ is substituted into R in Eq. (S-35) (more precisely, Eq. (S-34) is valid). Resetting $C \leftarrow C\bar{c}^{-\frac{2s}{1+s}}$, we obtain the assertion. □

Lemma 14. *There is a constant C such that*

$$\log Q(\delta, \mathcal{H}_{\ell_\infty}(R)) \geq CM \log Q(\delta/\sqrt{M}, \widetilde{\mathcal{H}}(R)),$$

for sufficiently small δ.

Proof. The proof is analogous to that of Lemma 5 in [4]. We describe the outline of the proof. Let $N = Q(\sqrt{2}\delta/\sqrt{M}, \widetilde{\mathcal{H}}(R))$ and $\{f_m^1, \dots, f_m^N\}$ be a $\sqrt{2}\delta/\sqrt{M}$-packing of $\mathcal{H}_m(R)$. Then we can construct a function class Υ as

$$\Upsilon = \left\{ f^{\boldsymbol{j}} = \sum_{m=1}^M f_m^{j_m} \mid \boldsymbol{j} = (j_1, \dots, j_M) \in \{1, \dots, N\}^M \right\}.$$

We denote by $[N] := \{1, \dots, N\}$. For two functions $f^{\boldsymbol{j}}, f^{\boldsymbol{j}'} \in \Upsilon$, we have by the construction

$$\|f^{\boldsymbol{j}} - f^{\boldsymbol{j}'}\|_{L_2(\Pi)}^2 = \sum_{m=1}^M \|f_m^{j_m} - f_m^{j'_m}\|_{L_2(\Pi)}^2 \geq \frac{2\delta^2}{M} \sum_{m=1}^M \mathbf{1}[j_m \neq j'_m].$$

Thus, it suffices to construct a sufficiently large subset $A \subset [N]^M$ such that all different pairs $\boldsymbol{j}, \boldsymbol{j}' \in A$ have at least $M/2$ of Hamming distance $d_H(\boldsymbol{j}, \boldsymbol{j}') := \sum_{m=1}^M \mathbf{1}[j_m \neq j'_m]$.

Now we define $d_H(A, \boldsymbol{j}) := \min_{\boldsymbol{j}' \in A} d_H(\boldsymbol{j}', \boldsymbol{j})$. If $|A|$ satisfies

$$\left| \left\{ \boldsymbol{j} \in [N]^M \;\middle|\; d_H(A, \boldsymbol{j}) \leq \frac{M}{2} \right\} \right| < |[N]^M| = N^M, \tag{S-36}$$

then there exists a member $\boldsymbol{j}' \in [N]^M$ such that $\boldsymbol{j}'$ is more than $\frac{M}{2}$ away from A with respect to d_H, i.e. $d_H(A, \boldsymbol{j}') > \frac{M}{2}$. That is, we can add $\boldsymbol{j}'$ to A as long as Eq. (S-36) holds. Now since

$$\left| \left\{ \boldsymbol{j} \in [N]^M \;\middle|\; d_H(A, \boldsymbol{j}) \leq \frac{M}{2} \right\} \right| \leq |A| \binom{M}{M/2} N^{M/2}, \tag{S-37}$$

Eq. (S-36) holds as long as A satisfies

$$|A| \leq \frac{1}{2} \frac{N^M}{\binom{M}{M/2} N^{M/2}} =: Q^*.$$

The logarithm of Q^* can be evaluated as follows

$$\begin{aligned} \log Q^* &= \log \left(\frac{1}{2} \frac{N^M}{\binom{M}{M/2} N^{M/2}} \right) = M \log N - \log 2 - \log \binom{M}{M/2} - \frac{M}{2} \log N \\ &\geq \frac{M}{2} \log N - \log 2 - \log 2^M \geq \frac{M}{2} \log \frac{N}{16}. \end{aligned}$$

There exists a constant C such that $N = Q(\sqrt{2}\delta/\sqrt{M}, \widetilde{\mathcal{H}}(R)) \geq CQ(\delta/\sqrt{M}, \widetilde{\mathcal{H}}(R))$ because $\log Q(\delta, \widetilde{\mathcal{H}}(R)) \sim \left(\frac{\delta}{R}\right)^{-2s}$. Thus we obtain the assertion for sufficiently large N. □

G Proof of Technical Lemmas

G.1 Proof of Lemma 2

Remind that Eq. (6) gives

$$\begin{aligned}&\|\hat{f}-f^*\|_{L_2(\Pi)}^2\\ &=O_p\left(\min_{\substack{\{r_m\}_{m=1}^M:\\ r_m>0}}\left\{\alpha_1^2+\beta_1^2+\left[\left(\frac{\alpha_2}{\alpha_1}\right)^2+\left(\frac{\beta_2}{\beta_1}\right)^2\right]\|f^*\|_\psi^2+\frac{M\log(M)}{n}\right\}\right).\end{aligned} \tag{S-38}$$

We derive an upper bound of the right hand side by adding a constraint $r_m=r$ $(\forall m)$. Since $s_m=s$ $(\forall m)$, under the constraint $r_m=r$ $(\forall m)$ we have

$$\frac{\alpha_2}{\alpha_1}=\frac{3\frac{sr^{1-s}}{\sqrt{n}}\|\mathbf{1}\|_{\psi^*}}{3\sqrt{M\frac{r^{-2s}}{n}}}=\frac{1}{\sqrt{M}}sr\|\mathbf{1}\|_{\psi^*},$$

$$\frac{\beta_2}{\beta_1}=\frac{3\frac{sr^{\frac{(1-s)^2}{1+s}}}{n^{\frac{1}{1+s}}}\|\mathbf{1}\|_{\psi^*}}{3\sqrt{M\frac{r^{-\frac{2s(3-s)}{1+s}}}{n^{\frac{2}{1+s}}}}}=\frac{1}{\sqrt{M}}sr\|\mathbf{1}\|_{\psi^*},$$

Thus $\frac{\alpha_2}{\alpha_1}=\frac{\beta_2}{\beta_1}$, and Eq. (S-38) becomes

$$\|\hat{f}-f^*\|_{L_2(\Pi)}^2=O_p\left(\min_{\substack{r>0,\\ r_m=r}}\left\{\alpha_1^2+\beta_1^2+2\frac{1}{M}s^2r^2\|\mathbf{1}\|_{\psi^*}^2\|f^*\|_\psi^2+\frac{M\log(M)}{n}\right\}\right). \tag{S-39}$$

By the definition, we see that the first two terms are monotonically decreasing function with respect to r and the third term is monotonically increasing function. The minimum of the right hand side is attained by balancing $\alpha_1^2+\beta_1^2$ and $2\frac{1}{M}s^2r^2\|\mathbf{1}\|_{\psi^*}^2\|f^*\|_\psi^2$. Since $\alpha_1^2+\beta_1^2\le 2\max\left(\alpha_1^2,\beta_1^2\right)$, Eq. (S-39) indicates that

$$\|\hat{f}-f^*\|_{L_2(\Pi)}^2\le O_p\left(\min_{\substack{r>0,\\ r_m=r}}\left\{2\max\left(\alpha_1^2,\beta_1^2\right)+2\frac{1}{M}s^2r^2\|\mathbf{1}\|_{\psi^*}^2\|f^*\|_\psi^2+\frac{M\log(M)}{n}\right\}\right). \tag{S-40}$$

To balance the first term and the second term, we need to consider two situations: $\alpha_1^2=\frac{1}{M}s^2r^2\|\mathbf{1}\|_{\psi^*}^2\|f^*\|_\psi^2$ or $\beta_1^2=\frac{1}{M}s^2r^2\|\mathbf{1}\|_{\psi^*}^2\|f^*\|_\psi^2$.

First we balance the terms α_1^2 and $\frac{1}{M}s^2r^2\|\mathbf{1}\|_{\psi^*}^2\|f^*\|_\psi^2$ under the restriction that $r_m=r$ $(\forall m)$:

$$\begin{aligned}&\alpha_1^2=\frac{1}{M}s^2r^2\|\mathbf{1}\|_{\psi^*}^2\|f^*\|_\psi^2\\ \Leftrightarrow\quad&9M\frac{r^{-2s}}{n}=\frac{1}{M}s^2r^2\|\mathbf{1}\|_{\psi^*}^2\|f^*\|_\psi^2\\ \Leftrightarrow\quad&r^{-1}=(s/3)^{\frac{1}{1+s}}M^{-\frac{1}{1+s}}n^{\frac{1}{2(1+s)}}(\|\mathbf{1}\|_{\psi^*}\|f^*\|_\psi)^{\frac{1}{1+s}}.\end{aligned}$$

For this r, we obtain

$$\begin{aligned}&\alpha_1^2=9M\frac{r^{-2s}}{n}\\ =&9^{\frac{1}{1+s}}s^{\frac{2s}{1+s}}M^{1-\frac{2s}{1+s}}n^{-\frac{1}{1+s}}(\|\mathbf{1}\|_{\psi^*}\|f^*\|_\psi)^{\frac{2s}{1+s}}\le 9M^{1-\frac{2s}{1+s}}n^{-\frac{1}{1+s}}(\|\mathbf{1}\|_{\psi^*}\|f^*\|_\psi)^{\frac{2s}{1+s}},\end{aligned}$$

where we used $s^{\frac{2s}{1+s}}\le 1$ and $9^{\frac{1}{1+s}}\le 9$ in the last inequality.

Next we balance the terms β_1^2 and $\frac{1}{M}s^2r^2\|\mathbf{1}\|_{\psi^*}^2\|f^*\|_\psi^2$ under the restriction that $r_m=r$ $(\forall m)$:

$$\beta_1^2=\frac{1}{M}s^2r^2\|\mathbf{1}\|_{\psi^*}^2\|f^*\|_\psi^2$$

$$\Leftrightarrow \quad 9M\frac{r^{-\frac{2s(3-s)}{1+s}}}{n^{\frac{2}{1+s}}} = \frac{1}{M}s^2r^2\,\|\mathbf{1}\|_{\psi^*}^2\,\|f^*\|_\psi^2$$
$$\Leftrightarrow \quad r^{-1} = (s/3)^{\frac{1+s}{1+4s-s^2}} M^{-\frac{1+s}{1+4s-s^2}} n^{\frac{1}{1+4s-s^2}} \left(\|\mathbf{1}\|_{\psi^*}\|f^*\|_\psi\right)^{\frac{1+s}{1+4s-s^2}}.$$

For this r, we obtain

$$\begin{aligned}
\beta_1^2 &= 9M\frac{r^{-\frac{2s(3-s)}{1+s}}}{n^{\frac{2}{1+s}}} \\
&= 9^{\frac{1+s}{1+4s-s^2}} s^{\frac{2s(3-s)}{1+4s-s^2}} M^{-\frac{1-2s+s^2}{1+4s-s^2}} n^{-\frac{2}{1+4s-s^2}} \left(\|\mathbf{1}\|_{\psi^*}\|f^*\|_\psi\right)^{\frac{2s(3-s)}{1+4s-s^2}} \\
&\leq 9M^{\frac{1-2s+s^2}{1+4s-s^2}} n^{-\frac{2}{1+4s-s^2}} \left(\|\mathbf{1}\|_{\psi^*}\|f^*\|_\psi\right)^{\frac{2s(3-s)}{1+4s-s^2}},
\end{aligned}$$

where we used $s^{\frac{2s(3-s)}{1+4s-s^2}} \leq 1$ and $9^{\frac{1+s}{1+4s-s^2}} \leq 9$ in the last inequality.

Therefore the right hand side of Eq. (S-40) is further bounded as

$$\begin{aligned}
&\|\hat{f} - f^*\|_{L_2(\Pi)}^2 \\
\leq\, & O_p\Bigg(4\max\Bigg\{9M^{1-\frac{2s}{1+s}} n^{-\frac{1}{1+s}} \left(\|\mathbf{1}\|_{\psi^*}\|f^*\|_\psi\right)^{\frac{2s}{1+s}}, \\
&\qquad 9M^{\frac{1-2s+s^2}{1+4s-s^2}} n^{-\frac{2}{1+4s-s^2}} \left(\|\mathbf{1}\|_{\psi^*}\|f^*\|_\psi\right)^{\frac{2s(3-s)}{1+4s-s^2}}\Bigg\} + \frac{M\log(M)}{n}\Bigg) \\
=\, & O_p\Bigg(M^{1-\frac{2s}{1+s}} n^{-\frac{1}{1+s}} \left(\|\mathbf{1}\|_{\psi^*}\|f^*\|_\psi\right)^{\frac{2s}{1+s}} + \\
&\qquad M^{\frac{(1-s)^2}{1+4s-s^2}} n^{-\frac{2}{1+4s-s^2}} \left(\|\mathbf{1}\|_{\psi^*}\|f^*\|_\psi\right)^{\frac{2s(3-s)}{1+4s-s^2}} + \frac{M\log(M)}{n}\Bigg).
\end{aligned}$$

Finally, if $n \geq (\|\mathbf{1}\|_{\psi^*}\|f^*\|_\psi/M)^{\frac{4s}{1-s}}$, the first term of the right hand side of this bound is not less than the second term:

$$M^{1-\frac{2s}{1+s}} n^{-\frac{1}{1+s}} \left(\|\mathbf{1}\|_{\psi^*}\|f^*\|_\psi\right)^{\frac{2s}{1+s}} \geq M^{\frac{(1-s)^2}{1+4s-s^2}} n^{-\frac{2}{1+4s-s^2}} \left(\|\mathbf{1}\|_{\psi^*}\|f^*\|_\psi\right)^{\frac{2s(3-s)}{1+4s-s^2}}.$$

Thus we obtain the assertion.

References

[1] O. Bousquet. A Bennett concentration inequality and its application to suprema of empirical process. *C. R. Acad. Sci. Paris Ser. I Math.*, 334:495–500, 2002.

[2] M. Ledoux and M. Talagrand. *Probability in Banach Spaces. Isoperimetry and Processes*. Springer, New York, 1991. MR1102015.

[3] G. Raskutti, M. Wainwright, and B. Yu. Lower bounds on minimax rates for nonparametric regression with additive sparsity and smoothness. In *Advances in Neural Information Processing Systems 22*, pages 1563–1570. MIT Press, Cambridge, MA, 2009.

[4] G. Raskutti, M. Wainwright, and B. Yu. Minimax-optimal rates for sparse additive models over kernel classes via convex programming. Technical report, 2010. arXiv:1008.3654.

[5] I. Steinwart. *Support Vector Machines*. Springer, 2008.

[6] M. Talagrand. New concentration inequalities in product spaces. *Inventiones Mathematicae*, 126:505–563, 1996.

[7] Y. Yang and A. Barron. Information-theoretic determination of minimax rates of convergence. *The Annals of Statistics*, 27(5):1564–1599, 1999.