Supplementary: Optimal Rates of Sketched-regularized Algorithms for Least-squares Regression over Hilbert Spaces

In this appendix, we first prove the lemmas stated in Section 4 and Corollary 5. We then review how the regression setting considered in this paper covers non-parametric regression with kernel methods. Finally, we construct a simple example for the non-attainable case, showing that the target function f_H is not necessarily to be in H_ρ.

A. Proofs for Lemmas in Section 4 and Corollary 5

For notational simplicity, we denote

$$\mathcal{R}_\lambda(u) = 1 - \mathcal{G}_\lambda(u)u, \tag{43}$$

and

$$\mathcal{N}(\lambda) = \mathrm{tr}(\mathcal{T}(\mathcal{T}+\lambda)^{-1}).$$

To proceed the proof, we need some basic operator inequalities.

Lemma 16. *(Fujii et al., 1993) Let A and B be two positive bounded linear operators on a separable Hilbert space. Then*

$$\|A^s B^s\| \le \|AB\|^s, \quad \textit{when } 0 \le s \le 1.$$

Lemma 17. *Let H_1, H_2 be two separable Hilbert spaces and $\mathcal{S} : H_1 \to H_2$ a compact operator. Then for any function $f : [0, \|\mathcal{S}\|] \to [0, \infty[$,*

$$f(\mathcal{S}\mathcal{S}^*)\mathcal{S} = \mathcal{S} f(\mathcal{S}^*\mathcal{S}).$$

Proof. The result can be proved using singular value decomposition of a compact operator. □

Lemma 18. *Let A and B be two non-negative bounded linear operators on a separable Hilbert space with $\max(\|A\|, \|B\|) \le \kappa^2$ for some non-negative κ^2. Then for any $\zeta > 0$,*

$$\|A^\zeta - B^\zeta\| \le C_{\zeta,\kappa} \|A - B\|^{\zeta \wedge 1}, \tag{44}$$

where

$$C_{\zeta,\kappa} = \begin{cases} 1 & \textit{when } \zeta \le 1, \\ 2\zeta\kappa^{2\zeta-2} & \textit{when } \zeta > 1. \end{cases} \tag{45}$$

Proof. The proof is based on the fact that u^ζ is operator monotone if $0 < \zeta \le 1$. While for $\zeta \ge 1$, the proof can be found in, e.g., (Dicker et al., 2016). □

Lemma 19. *Let X and A be bounded linear operators on a separable Hilbert space. Suppose that $X \succeq 0$ and $\|A\| \le 1$. Then for any $s \in [0, 1]$,*

$$X^* A^s X \le (X^* A X)^s.$$

Proof. Following from (Hansen, 1980) and the fact that the function u^s with $s \in [0, 1]$ is operator monotone. □

A.1. Proof of Proposition 7

Adding and subtracting with the same term, and using the triangle inequality, we have

$$\|\mathcal{L}^{-a}(\mathcal{S}_\rho \omega_\lambda^{\mathbf{z}} - f_H)\|_\rho \le \|\mathcal{L}^{-a}\mathcal{S}_\rho(\omega_\lambda^{\mathbf{z}} - \omega_\lambda)\|_\rho + \|\mathcal{L}^{-a}(\mathcal{S}_\rho \omega_\lambda - f_H)\|_\rho.$$

Applying Part 1) of Lemma 6 to bound the last term, with $0 \le a \le \zeta$,

$$\begin{aligned}\|\mathcal{L}^{-a}(\mathcal{S}_\rho \omega_\lambda^{\mathbf{z}} - f_H)\|_\rho &\le \|\mathcal{L}^{-a}\mathcal{S}_\rho(\omega_\lambda^{\mathbf{z}} - \omega_\lambda)\|_\rho + R\lambda^{\zeta - a} \\ &\le \|\mathcal{L}^{-a}\mathcal{S}_\rho \mathcal{T}^{a-\frac{1}{2}}\| \|\mathcal{T}^{\frac{1}{2}-a}(\omega_\lambda^{\mathbf{z}} - \omega_\lambda)\|_H + R\lambda^{\zeta - a}.\end{aligned}$$

Using the spectral theorem for compact operators, $\mathcal{L} = \mathcal{S}_\rho \mathcal{S}_\rho^*$, and $\mathcal{T} = \mathcal{S}_\rho^* \mathcal{S}_\rho$, we have

$$\|\mathcal{L}^{-a}\mathcal{S}_\rho \mathcal{T}^{a-\frac{1}{2}}\| \leq 1,$$

and thus

$$\|\mathcal{L}^{-a}(\mathcal{S}_\rho \omega_\lambda^{\mathbf{z}} - f_H)\|_\rho \leq \|\mathcal{T}^{\frac{1}{2}-a}(\omega_\lambda^{\mathbf{z}} - \omega_\lambda)\|_H + R\lambda^{\zeta-a}.$$

Adding and subtracting with the same term, and using the triangle inequality,

$$\|\mathcal{L}^{-a}(\mathcal{S}_\rho \omega_\lambda^{\mathbf{z}} - f_H)\|_\rho \leq \|\mathcal{T}^{\frac{1}{2}-a}(\omega_\lambda^{\mathbf{z}} - P\omega_\lambda)\|_H + \|\mathcal{T}^{\frac{1}{2}-a}(I-P)\omega_\lambda\|_H + R\lambda^{\zeta-a}.$$

Since P is an orthogonal projected operator and $a \in [0, \frac{1}{2}]$, we have

$$\begin{aligned}
&\|\mathcal{T}^{\frac{1}{2}-a}(I-P)\omega_\lambda\|_H \\
=&\|\mathcal{T}^{\frac{1}{2}(1-2a)}(I-P)^{1-2a}(I-P)\omega_\lambda\|_H \\
\leq&\|\mathcal{T}^{\frac{1}{2}(1-2a)}(I-P)^{1-2a}\|\|(I-P)\mathcal{T}^{\frac{1}{2}}\|\|\mathcal{T}^{-\frac{1}{2}}\omega_\lambda\|_H \\
\leq&\|\mathcal{T}^{\frac{1}{2}}(I-P)\|^{1-2a}\|(I-P)\mathcal{T}^{\frac{1}{2}}\|\tau R\kappa^{2(\zeta-1)_+}\lambda^{(\zeta-1)_-} \\
=&\Delta_5^{1-a}\tau R\kappa^{2(\zeta-1)_+}\lambda^{(\zeta-1)_-},
\end{aligned}$$

(where for the last second inequality, we used Lemma 16 and Part 2) of Lemma 6), and we subsequently get that

$$\|\mathcal{L}^{-a}(\mathcal{S}_\rho \omega_\lambda^{\mathbf{z}} - f_H)\|_\rho \leq \|\mathcal{T}^{\frac{1}{2}-a}(\omega_\lambda^{\mathbf{z}} - P\omega_\lambda)\|_H + \tau R\kappa^{2(\zeta-1)_+}\lambda^{(\zeta-1)_-}\Delta_5^{1-a} + R\lambda^{\zeta-a}.$$

Since for all $\omega \in H$, and $a \in [0, \frac{1}{2}]$,

$$\begin{aligned}
\|\mathcal{T}^{\frac{1}{2}-a}\omega\|_H \leq&\|\mathcal{T}_\lambda^{\frac{1}{2}-a}\mathcal{T}_{\mathbf{x}\lambda}^{a-\frac{1}{2}}\|\|\mathcal{T}_{\mathbf{x}\lambda}^{\frac{1}{2}-a}\omega\|_H \\
\leq&\lambda^{-a}\|\mathcal{T}_\lambda^{\frac{1}{2}-a}\mathcal{T}_{\mathbf{x}\lambda}^{a-\frac{1}{2}}\|\|\mathcal{T}_{\mathbf{x}\lambda}^{\frac{1}{2}}\omega\|_H \\
\leq&\lambda^{-a}\|\mathcal{T}_\lambda^{\frac{1}{2}}\mathcal{T}_{\mathbf{x}\lambda}^{-\frac{1}{2}}\|^{1-2a}\|\mathcal{T}_{\mathbf{x}\lambda}^{\frac{1}{2}}\omega\|_H \\
\leq&\lambda^{-a}\Delta_1^{\frac{1}{2}-a}\|\mathcal{T}_{\mathbf{x}\lambda}^{\frac{1}{2}}\omega\|_H
\end{aligned}$$

(where we used Lemma 16 for the last second inequality), we get

$$\|\mathcal{L}^{-a}(\mathcal{S}_\rho \omega_\lambda^{\mathbf{z}} - f_H)\|_\rho \leq \lambda^{-a}\Delta_1^{\frac{1}{2}-a}\|\mathcal{T}_{\mathbf{x}\lambda}^{\frac{1}{2}}(\omega_\lambda^{\mathbf{z}} - P\omega_\lambda)\|_H + \tau R\kappa^{2(\zeta-1)_+}\lambda^{(\zeta-1)_-}\Delta_5^{1-a} + R\lambda^{\zeta-a}. \tag{46}$$

In what follows, we estimate $\|\mathcal{T}_{\mathbf{x}\lambda}^{\frac{1}{2}}(\omega_\lambda^{\mathbf{z}} - P\omega_\lambda)\|_H$.

Introducing with (11), with $P^2 = P$,

$$\|\mathcal{T}_{\mathbf{x}\lambda}^{\frac{1}{2}}(\omega_\lambda^{\mathbf{z}} - P\omega_\lambda)\|_H = \|\mathcal{T}_{\mathbf{x}\lambda}^{\frac{1}{2}}P(\mathcal{G}_\lambda(P\mathcal{T}_{\mathbf{x}}P)P\mathcal{S}_{\mathbf{x}}^*\mathbf{y} - P\omega_\lambda)\|_H.$$

Since for any $\omega \in H$,

$$\|\mathcal{T}_{\mathbf{x}\lambda}^{\frac{1}{2}}P\omega\|_H^2 = \langle P\mathcal{T}_{\mathbf{x}\lambda}P\omega, \omega\rangle_H \leq \langle (P\mathcal{T}_{\mathbf{x}}P + \lambda)\omega, \omega\rangle_H = \|(P\mathcal{T}_{\mathbf{x}}P + \lambda)^{\frac{1}{2}}\omega\|_H^2,$$

and we thus get

$$\|\mathcal{T}_{\mathbf{x}\lambda}^{\frac{1}{2}}(\omega_\lambda^{\mathbf{z}} - P\omega_\lambda)\|_H \leq \|\mathcal{U}_\lambda^{\frac{1}{2}}(\mathcal{G}_\lambda(\mathcal{U})P\mathcal{S}_{\mathbf{x}}^*\mathbf{y} - P\omega_\lambda)\|_H,$$

where we denote

$$\mathcal{U} = P\mathcal{T}_{\mathbf{x}}P, \quad \mathcal{U}_\lambda = \mathcal{U} + \lambda. \tag{47}$$

Subtracting and adding with the same term, and applying the triangle inequality, with the notation $\mathcal{R}_\lambda$ given by (43) and $P^2 = P$, we have

$$\|\mathcal{T}_{\mathbf{x}\lambda}^{\frac{1}{2}}(\omega_\lambda^{\mathbf{z}} - P\omega_\lambda)\|_H \leq \|\underbrace{\mathcal{U}_\lambda^{\frac{1}{2}}\mathcal{G}_\lambda(\mathcal{U})P(\mathcal{S}_{\mathbf{x}}^*\mathbf{y} - \mathcal{T}_{\mathbf{x}}P\omega_\lambda)}_{\textbf{Term.A}}\|_H + \|\underbrace{\mathcal{U}_\lambda^{\frac{1}{2}}\mathcal{R}_\lambda(\mathcal{U})P\omega_\lambda}_{\textbf{Term.B}}\|_H. \tag{48}$$

We will estimate the above two terms of the right-hand side.
Estimating $\|\mathbf{Term.A}\|_H$:
Note that

$$\begin{aligned}&(\mathcal{U}_\lambda^{\frac{1}{2}}\mathcal{G}_\lambda(\mathcal{U})P\mathcal{T}_{\mathbf{x}\lambda}^{\frac{1}{2}})(\mathcal{U}_\lambda^{\frac{1}{2}}\mathcal{G}_\lambda(\mathcal{U})P\mathcal{T}_{\mathbf{x}\lambda}^{\frac{1}{2}})^*\\&=\mathcal{U}_\lambda^{\frac{1}{2}}\mathcal{G}_\lambda(\mathcal{U})(\mathcal{U}+\lambda P^2)\mathcal{G}_\lambda(\mathcal{U})\mathcal{U}_\lambda^{\frac{1}{2}}\\&\preceq\left[\mathcal{U}_\lambda\mathcal{G}_\lambda(\mathcal{U})\right]^2,\end{aligned}$$

where we used $P^2=P\preceq I$ for the last inequality. Thus, combing with $\|A\|=\|A^*A\|^{\frac{1}{2}}$,

$$\|\mathcal{U}_\lambda^{\frac{1}{2}}\mathcal{G}_\lambda(\mathcal{U})P\mathcal{T}_{\mathbf{x}\lambda}^{\frac{1}{2}}\|\leq\|\mathcal{U}_\lambda\mathcal{G}_\lambda(\mathcal{U})\|.$$

Using the spectral theorem, with $\|\mathcal{U}\|\leq\|\mathcal{T}_{\mathbf{x}}\|\leq\kappa^2$ (implied by (6)), and then applying (12),

$$\|\mathcal{U}_\lambda^{\frac{1}{2}}\mathcal{G}_\lambda(\mathcal{U})P\mathcal{T}_{\mathbf{x}\lambda}^{\frac{1}{2}}\|\leq\sup_{u\in[0,\kappa^2]}|(u+\lambda)\mathcal{G}_\lambda(u)|\leq\tau.$$

Using the above inequality, and by a simple calculation,

$$\|\mathbf{Term.A}\|_H\leq\|\mathcal{U}_\lambda^{\frac{1}{2}}\mathcal{G}_\lambda(\mathcal{U})P\mathcal{T}_{\mathbf{x}\lambda}^{\frac{1}{2}}\|\|\mathcal{T}_{\mathbf{x}\lambda}^{-\frac{1}{2}}(\mathcal{S}_{\mathbf{x}}^*\mathbf{y}-\mathcal{T}_{\mathbf{x}}P\omega_\lambda)\|\leq\tau\|\mathcal{T}_{\mathbf{x}\lambda}^{-\frac{1}{2}}(\mathcal{S}_{\mathbf{x}}^*\mathbf{y}-\mathcal{T}_{\mathbf{x}}P\omega_\lambda)\|.$$

Adding and subtracting with the same terms, and using the triangle inequality,

$$\begin{aligned}\|\mathbf{Term.A}\|_H\leq&\tau\|\mathcal{T}_{\mathbf{x}\lambda}^{-\frac{1}{2}}(\mathcal{S}_{\mathbf{x}}^*\mathbf{y}-\mathcal{T}_{\mathbf{x}}\omega_\lambda)\|_H+\tau\|\mathcal{T}_{\mathbf{x}\lambda}^{-\frac{1}{2}}\mathcal{T}_{\mathbf{x}}(I-P)\omega_\lambda)\|_H\\\leq&\tau\|\mathcal{T}_{\mathbf{x}\lambda}^{-\frac{1}{2}}\mathcal{T}_\lambda^{\frac{1}{2}}\|\|\mathcal{T}_\lambda^{-\frac{1}{2}}(\mathcal{S}_{\mathbf{x}}^*\mathbf{y}-\mathcal{T}_{\mathbf{x}}\omega_\lambda)\|_H+\tau\|\mathcal{T}_{\mathbf{x}\lambda}^{-\frac{1}{2}}\mathcal{T}_{\mathbf{x}}(I-P)\omega_\lambda)\|_H\\\leq&\tau\Delta_1^{\frac{1}{2}}\|\mathcal{T}_\lambda^{-\frac{1}{2}}(\mathcal{S}_{\mathbf{x}}^*\mathbf{y}-\mathcal{T}_{\mathbf{x}}\omega_\lambda)\|_H+\tau\|\mathcal{T}_{\mathbf{x}\lambda}^{-\frac{1}{2}}\mathcal{T}_{\mathbf{x}}(I-P)\omega_\lambda)\|_H\\\leq&\tau\Delta_1^{\frac{1}{2}}(\Delta_2+\|\mathcal{T}_\lambda^{-\frac{1}{2}}(\mathcal{T}\omega_\lambda-\mathcal{S}_\rho^*f_H)\|_H)+\tau\|\mathcal{T}_{\mathbf{x}\lambda}^{-\frac{1}{2}}\mathcal{T}_{\mathbf{x}}(I-P)\omega_\lambda)\|_H\\\leq&\tau\Delta_1^{\frac{1}{2}}(\Delta_2+\|\mathcal{T}_\lambda^{-\frac{1}{2}}\mathcal{S}_\rho^*\|\|\mathcal{S}_\rho\omega_\lambda-f_H\|_\rho)+\tau\|\mathcal{T}_{\mathbf{x}\lambda}^{-\frac{1}{2}}\mathcal{T}_{\mathbf{x}}(I-P)\omega_\lambda)\|_H,\end{aligned}$$

where we used $\mathcal{T}=\mathcal{S}_\rho^*\mathcal{S}_\rho$ for the last inequality. Applying Part 1) of Lemma 6 and $\|\mathcal{T}_\lambda^{-\frac{1}{2}}\mathcal{S}_\rho^*\|\leq1$,

$$\|\mathbf{Term.A}\|_H\leq\tau\Delta_1^{\frac{1}{2}}(\Delta_2+R\lambda^\zeta)+\tau\|\mathcal{T}_{\mathbf{x}\lambda}^{-\frac{1}{2}}\mathcal{T}_{\mathbf{x}}(I-P)\omega_\lambda)\|_H. \tag{49}$$

In what follows, we estimate $\|\mathcal{T}_{\mathbf{x}\lambda}^{-\frac{1}{2}}\mathcal{T}_{\mathbf{x}}(I-P)\omega_\lambda\|_H$, considering two different cases.
Case $\zeta\leq1$.
We have

$$\|\mathcal{T}_{\mathbf{x}\lambda}^{-\frac{1}{2}}\mathcal{T}_{\mathbf{x}}(I-P)\omega_\lambda\|_H\leq\|\mathcal{T}_{\mathbf{x}\lambda}^{-\frac{1}{2}}\mathcal{T}_{\mathbf{x}}\mathcal{T}_{\mathbf{x}\lambda}^{\frac{-1}{2}}\|\|\mathcal{T}_{\mathbf{x}\lambda}^{\frac{1}{2}}\mathcal{T}_\lambda^{-\frac{1}{2}}\|\|\mathcal{T}_\lambda^{\frac{1}{2}}(I-P)\omega_\lambda\|_H\leq\Delta_1^{\frac{1}{2}}\|\mathcal{T}_\lambda^{\frac{1}{2}}(I-P)\omega_\lambda\|_H.$$

Since P is a projection operator, $(I-P)^2=I-P$, and we thus have

$$\|\mathcal{T}_{\mathbf{x}\lambda}^{-\frac{1}{2}}\mathcal{T}_{\mathbf{x}}(I-P)\omega_\lambda\|_H\leq\Delta_1^{\frac{1}{2}}\|\mathcal{T}_\lambda^{\frac{1}{2}}(I-P)\|\|(I-P)\mathcal{T}^{\frac{1}{2}}\|\|\mathcal{T}^{-\frac{1}{2}}\omega_\lambda\|_H\leq\tau\Delta_1^{\frac{1}{2}}\|\mathcal{T}_\lambda^{\frac{1}{2}}(I-P)\|\Delta_5^{\frac{1}{2}}R\lambda^{\zeta-1},$$

where for the last inequality, we used Part 2) of Lemma 6. Note that for any $\omega\in H$ with $\|\omega\|_H=1$,

$$\|\mathcal{T}_\lambda^{\frac{1}{2}}(I-P)\omega\|_H^2=\langle\mathcal{T}_\lambda(I-P)\omega,(I-P)\omega\rangle_H=\|\mathcal{T}^{\frac{1}{2}}(I-P)\omega\|_H^2+\lambda\|(I-P)\omega\|_H^2\leq\|\mathcal{T}^{\frac{1}{2}}(I-P)\|^2+\lambda\leq\Delta_5+\lambda.$$

It thus follows that

$$\|\mathcal{T}_\lambda^{\frac{1}{2}}(I-P)\|_H\leq(\Delta_5+\lambda)^{\frac{1}{2}}, \tag{50}$$

and thus

$$\|\mathcal{T}_{\mathbf{x}\lambda}^{-\frac{1}{2}}\mathcal{T}_{\mathbf{x}}(I-P)\omega_\lambda\|_H\leq\Delta_1^{\frac{1}{2}}(\Delta_5+\lambda)\tau R\lambda^{\zeta-1}.$$

Introducing the above into (49), we know that $\mathbf{Term.A}$ can be estimated as ($\zeta \leq 1$)

$$\|\mathbf{Term.A}\|_H \leq \tau \Delta_1^{\frac{1}{2}} \left(\Delta_2 + (\tau+1) R \lambda^{\zeta} + \tau R \lambda^{\zeta-1} \Delta_5 \right). \tag{51}$$

Case $\zeta \geq 1$.
We first have

$$\begin{aligned}
\|\mathcal{T}_{\mathbf{x}\lambda}^{-\frac{1}{2}} \mathcal{T}_{\mathbf{x}} (I-P)\omega_\lambda\|_H \leq & \Delta_1^{\frac{1}{2}} \|\mathcal{T}_{\lambda}^{-\frac{1}{2}} \mathcal{T}_{\mathbf{x}} (I-P)\omega_\lambda)\|_H \\
\leq & \Delta_1^{\frac{1}{2}} \left(\|\mathcal{T}_{\lambda}^{-\frac{1}{2}} (\mathcal{T}_{\mathbf{x}} - \mathcal{T})(I-P)\omega_\lambda\|_H + \|\mathcal{T}_{\lambda}^{-\frac{1}{2}} \mathcal{T}(I-P)\omega_\lambda\|_H \right) \\
\leq & \Delta_1^{\frac{1}{2}} \left(\Delta_4 \|(I-P)\omega_\lambda\|_H + \|\mathcal{T}^{\frac{1}{2}}(I-P)\omega_\lambda\|_H \right).
\end{aligned}$$

Since P is a projection operator, $(I-P)^2 = I-P$, we thus have

$$\begin{aligned}
\|\mathcal{T}_{\mathbf{x}\lambda}^{-\frac{1}{2}} \mathcal{T}_{\mathbf{x}} (I-P)\omega_\lambda\|_H \leq & \Delta_1^{\frac{1}{2}} \left(\Delta_4 \|I-P\| \|\mathcal{T}^{\frac{1}{2}}\| \|\mathcal{T}^{-\frac{1}{2}}\omega_\lambda\|_H + \|\mathcal{T}^{\frac{1}{2}}(I-P)\| \|(I-P)\mathcal{T}^{\frac{1}{2}}\| \|\mathcal{T}^{-\frac{1}{2}}\omega_\lambda\|_H \right) \\
\leq & \Delta_1^{\frac{1}{2}} \left(\kappa \Delta_4 + \Delta_5 \right) \|\mathcal{T}^{-\frac{1}{2}}\omega_\lambda\|_H,
\end{aligned}$$

where we used (3) for the last inequality. Applying Part 2) of Lemma 6, we get

$$\|\mathcal{T}_{\mathbf{x}\lambda}^{-\frac{1}{2}} \mathcal{T}_{\mathbf{x}} (I-P)\omega_\lambda\|_H \leq \Delta_1^{\frac{1}{2}} \left(\kappa \Delta_4 + \Delta_5 \right) \tau \kappa^{2(\zeta-1)} R.$$

Introducing the above into (49), we get for $\zeta \geq 1$,

$$\|\mathbf{Term.A}\|_H \leq \tau \Delta_1^{\frac{1}{2}} \left(\Delta_2 + R\lambda^{\zeta} + \left(\kappa \Delta_4 + \Delta_5 \right) \tau \kappa^{2(\zeta-1)} R \right). \tag{52}$$

Estimating $\|\mathbf{Term.B}\|_H$:
We estimate $\|\mathbf{Term.B}\|_H$, considering two different cases.
Case I: $\zeta \leq 1$.
We first have

$$\begin{aligned}
\mathcal{U}_\lambda^{\frac{1}{2}} \mathcal{R}_\lambda(\mathcal{U}) P \mathcal{T}_{\mathbf{x}\lambda}^{\frac{1}{2}} (\mathcal{U}_\lambda^{\frac{1}{2}} \mathcal{R}_\lambda(\mathcal{U}) P \mathcal{T}_{\mathbf{x}\lambda}^{\frac{1}{2}})^* = & \mathcal{U}_\lambda^{\frac{1}{2}} \mathcal{R}_\lambda(\mathcal{U}) (\mathcal{U} + \lambda P^2) \mathcal{R}_\lambda(\mathcal{U}) \mathcal{U}_\lambda^{\frac{1}{2}} \\
\preceq & \left(\mathcal{R}_\lambda(\mathcal{U}) \mathcal{U}_\lambda \right)^2,
\end{aligned}$$

where we used $P^2 = P \preceq I$ for the last inequality. Thus, according to $\|A\| = \|AA^*\|^{\frac{1}{2}}$,

$$\|\mathcal{U}_\lambda^{\frac{1}{2}} \mathcal{R}_\lambda(\mathcal{U}) P \mathcal{T}_{\mathbf{x}\lambda}^{\frac{1}{2}}\| \leq \|\mathcal{R}_\lambda(\mathcal{U}) \mathcal{U}_\lambda\|.$$

Using the spectral theorem and (13), and noting that $\|\mathcal{U}\| \leq \|P\|^2 \|\mathcal{T}_{\mathbf{x}}\| \leq \kappa^2$ by (6), we get

$$\|\mathcal{U}_\lambda^{\frac{1}{2}} \mathcal{R}_\lambda(\mathcal{U}) P \mathcal{T}_{\mathbf{x}\lambda}^{\frac{1}{2}}\| \leq \sup_{u \in [0,\kappa^2]} |\mathcal{R}_\lambda(u)(u+\lambda)| \leq \lambda.$$

Using the above inequality and by a direct calculation,

$$\|\mathbf{Term.B}\|_H \leq \|\mathcal{U}_\lambda^{\frac{1}{2}} \mathcal{R}_\lambda(\mathcal{U}) P \mathcal{T}_{\mathbf{x}\lambda}^{\frac{1}{2}}\| \|\mathcal{T}_{\mathbf{x}\lambda}^{\frac{-1}{2}} \mathcal{T}_{\lambda}^{\frac{1}{2}}\| \|\mathcal{T}^{-\frac{1}{2}}\omega_\lambda\|_H \leq \lambda \Delta_1^{\frac{1}{2}} \|\mathcal{T}^{-\frac{1}{2}}\omega_\lambda\|_H.$$

Applying Part 2) of Lemma 6, we get

$$\|\mathbf{Term.B}\|_H \leq \tau R \lambda^{\zeta} \Delta_1^{\frac{1}{2}}. \tag{53}$$

Applying the above and (51) into (48), we know that for any $\zeta \in [0,1]$,

$$\|\mathcal{T}_{\mathbf{x}\lambda}^{\frac{1}{2}} (\omega_\lambda^{\mathbf{z}} - P\omega_\lambda)\|_H \leq \tau \Delta_1^{\frac{1}{2}} \left(\Delta_2 + (2\tau+1) R \lambda^{\zeta} + \tau R \Delta_5 \lambda^{\zeta-1} \right).$$

Using the above into (46), we can prove the first desired result.
Case II: $\zeta \geq 1$
We denote

$$\mathcal{V} = \mathcal{T}_{\mathbf{x}}^{\frac{1}{2}} P \mathcal{T}_{\mathbf{x}}^{\frac{1}{2}}, \quad \mathcal{V}_\lambda = \mathcal{V} + \lambda. \tag{54}$$

Noting that $\mathcal{U} = P\mathcal{T}_{\mathbf{x}}P = P\mathcal{T}_{\mathbf{x}}^{\frac{1}{2}}(P\mathcal{T}_{\mathbf{x}}^{\frac{1}{2}})^*$, thus following from Lemma 17 (with $f(u) = (u+\lambda)^{\frac{1}{2}}\mathcal{R}_\lambda(u)$) and $P^2 = P$,

$$\|\mathcal{U}_\lambda^{\frac{1}{2}}\mathcal{R}_\lambda(\mathcal{U})P\mathcal{T}_{\mathbf{x}}^{\zeta-\frac{1}{2}}\| = \|\mathcal{U}_\lambda^{\frac{1}{2}}\mathcal{R}_\lambda(\mathcal{U})(P\mathcal{T}_{\mathbf{x}}^{\frac{1}{2}})\mathcal{T}_{\mathbf{x}}^{\zeta-1}\| = \|(P\mathcal{T}_{\mathbf{x}}^{\frac{1}{2}})\mathcal{V}_\lambda^{\frac{1}{2}}\mathcal{R}_\lambda(\mathcal{V})\mathcal{T}_{\mathbf{x}}^{\zeta-1}\|.$$

Adding and subtracting with the same term, using the triangle inequality,

$$\begin{aligned}\|\mathcal{U}_\lambda^{\frac{1}{2}}\mathcal{R}_\lambda(\mathcal{U})P\mathcal{T}_{\mathbf{x}}^{\zeta-\frac{1}{2}}\| \leq& \|P\mathcal{T}_{\mathbf{x}}^{\frac{1}{2}}\mathcal{V}_\lambda^{\frac{1}{2}}\mathcal{R}_\lambda(\mathcal{V})\mathcal{V}^{\zeta-1}\| + \|P\mathcal{T}_{\mathbf{x}}^{\frac{1}{2}}\mathcal{V}_\lambda^{\frac{1}{2}}\mathcal{R}_\lambda(\mathcal{V})(\mathcal{T}_{\mathbf{x}}^{\zeta-1} - \mathcal{V}^{\zeta-1})\| \\ \leq& \|P\mathcal{T}_{\mathbf{x}}^{\frac{1}{2}}\mathcal{V}_\lambda^{\frac{1}{2}}\mathcal{R}_\lambda(\mathcal{V})\mathcal{V}^{\zeta-1}\| + \|P\mathcal{T}_{\mathbf{x}}^{\frac{1}{2}}\mathcal{V}_\lambda^{\frac{1}{2}}\mathcal{R}_\lambda(\mathcal{V})\|\|\mathcal{T}_{\mathbf{x}}^{\zeta-1} - \mathcal{V}^{\zeta-1}\|.\end{aligned}$$

Using Lemma 18, with (6) and $\|\mathcal{V}\| \leq \|\mathcal{T}_{\mathbf{x}}\| \leq \kappa^2$,

$$\|\mathcal{U}_\lambda^{\frac{1}{2}}\mathcal{R}_\lambda(\mathcal{U})P\mathcal{T}_{\mathbf{x}}^{\zeta-\frac{1}{2}}\| \leq \|P\mathcal{T}_{\mathbf{x}}^{\frac{1}{2}}\mathcal{V}_\lambda^{\frac{1}{2}}\mathcal{R}_\lambda(\mathcal{V})\mathcal{V}^{\zeta-1}\| + \|P\mathcal{T}_{\mathbf{x}}^{\frac{1}{2}}\mathcal{V}_\lambda^{\frac{1}{2}}\mathcal{R}_\lambda(\mathcal{V})\|\kappa^{2(\zeta-2)_+}\|\mathcal{T}_{\mathbf{x}} - \mathcal{V}\|^{(\zeta-1)\wedge 1}.$$

Using $\|A\| = \|A^*A\|^{\frac{1}{2}}$, $P^2 = P$, the spectral theorem, and (13), for any $s \in [1, \tau]$,

$$\begin{aligned}\|P\mathcal{T}_{\mathbf{x}}^{\frac{1}{2}}\mathcal{V}_\lambda^{\frac{1}{2}}\mathcal{R}_\lambda(\mathcal{V})\mathcal{V}^{s-1}\| =& \|\mathcal{V}^{s-1}\mathcal{R}_\lambda(\mathcal{V})\mathcal{V}_\lambda\mathcal{V}\mathcal{R}_\lambda(\mathcal{V})\mathcal{V}^{s-1}\|^{\frac{1}{2}} \\ &\leq \sup_{u\in[0,\kappa^2]}|\mathcal{R}_\lambda(u)u^{s-\frac{1}{2}}(u+\lambda)^{\frac{1}{2}}| \leq \lambda^s,\end{aligned}$$

and thus we get

$$\|\mathcal{U}_\lambda^{\frac{1}{2}-a}\mathcal{R}_\lambda(\mathcal{U})P\mathcal{T}_{\mathbf{x}}^{\zeta-\frac{1}{2}}\| \leq \lambda^\zeta + \lambda\kappa^{2(\zeta-2)_+}\|\mathcal{T}_{\mathbf{x}} - \mathcal{V}\|^{(\zeta-1)\wedge 1}.$$

Using Lemma 14, $(I-P)^2 = I - P$ and $\|A^*A\| = \|A\|^2$, we have

$$\|\mathcal{T}_{\mathbf{x}} - \mathcal{V}\| = \|\mathcal{T}_{\mathbf{x}}^{\frac{1}{2}}(I-P)\mathcal{T}_{\mathbf{x}}^{\frac{1}{2}}\| \leq \|\mathcal{T}_{\mathbf{x}} - \mathcal{T}\| + \|\mathcal{T}^{\frac{1}{2}}(I-P)\mathcal{T}^{\frac{1}{2}}\| \leq \Delta_3 + \Delta_5,$$

and we thus get

$$\|\mathcal{U}_\lambda^{\frac{1}{2}}\mathcal{R}_\lambda(\mathcal{U})P\mathcal{T}_{\mathbf{x}}^{\zeta-\frac{1}{2}}\| \leq \lambda^\zeta + \lambda\kappa^{2(\zeta-2)_+}(\Delta_3 + \Delta_5)^{(\zeta-1)\wedge 1}. \tag{55}$$

Now we are ready to estimate $\|\mathbf{Term.B}\|_H$. By some direct calculations and Part 2) of Lemma 6,

$$\|\mathbf{Term.B}\|_H \leq \|\mathcal{U}_\lambda^{\frac{1}{2}}\mathcal{R}_\lambda(\mathcal{U})P\mathcal{T}^{\zeta-\frac{1}{2}}\|\|\mathcal{T}^{\frac{1}{2}-\zeta}\omega_\lambda\|_H \leq \|\mathcal{U}_\lambda^{\frac{1}{2}}\mathcal{R}_\lambda(\mathcal{U})P\mathcal{T}^{\zeta-\frac{1}{2}}\|\tau R.$$

Adding and subtracting with the same term, and using the triangle inequality,

$$\|\mathbf{Term.B}\|_H \leq \tau R\left(\|\mathcal{U}_\lambda^{\frac{1}{2}}\mathcal{R}_\lambda(\mathcal{U})P\mathcal{T}_{\mathbf{x}}^{\zeta-\frac{1}{2}}\| + \|\mathcal{U}_\lambda^{\frac{1}{2}}\mathcal{R}_\lambda(\mathcal{U})\|\|\mathcal{T}^{\zeta-\frac{1}{2}} - \mathcal{T}_{\mathbf{x}}^{\zeta-\frac{1}{2}}\|\right).$$

Using the spectral theorem, with $\|\mathcal{U}\| \leq \|\mathcal{T}_{\mathbf{x}}\| \leq \kappa^2$ by (6) and (13),

$$\|\mathcal{U}_\lambda^{\frac{1}{2}}\mathcal{R}_\lambda(\mathcal{U})\| = \sup_{u\in]0,\kappa^2]}|\mathcal{R}_\lambda(u)(u+\lambda)^{\frac{1}{2}}| \leq \lambda^{\frac{1}{2}},$$

and we thus get

$$\|\mathbf{Term.B}\|_H \leq \tau R\left(\|\mathcal{U}_\lambda^{\frac{1}{2}}\mathcal{R}_\lambda(\mathcal{U})P\mathcal{T}_{\mathbf{x}}^{\zeta-\frac{1}{2}}\| + \lambda^{\frac{1}{2}}\|\mathcal{T}^{\zeta-\frac{1}{2}} - \mathcal{T}_{\mathbf{x}}^{\zeta-\frac{1}{2}}\|\right).$$

Applying Lemma 18, with (3) and (6),

$$\|\mathbf{Term.B}\|_H \leq \tau R\left(\|\mathcal{U}_\lambda^{\frac{1}{2}}\mathcal{R}_\lambda(\mathcal{U})P\mathcal{T}_{\mathbf{x}}^{\zeta-\frac{1}{2}}\| + \lambda^{\frac{1}{2}}\kappa^{(2\zeta-3)_+}\Delta_3^{(\zeta-\frac{1}{2})\wedge 1}\right).$$

Introducing with (55),

$$\|\mathbf{Term.B}\|_H \leq \tau R\left(\lambda^\zeta + \kappa^{2(\zeta-2)_+}\lambda(\Delta_3 + \Delta_5)^{(\zeta-1)\wedge 1} + \kappa^{(2\zeta-3)_+}\lambda^{\frac{1}{2}}\Delta_3^{(\zeta-\frac{1}{2})\wedge 1}\right).$$

Introducing the above inequality and (52) into (48), noting that $\Delta_1 \geq 1$ and $\kappa^2 \geq 1$, we know that for any $\zeta \geq 1$,

$$\|\mathcal{T}_{\mathbf{x}\lambda}^{\frac{1}{2}}(\omega_\lambda^{\mathbf{z}} - P\omega_\lambda)\|_H \leq \tau\Delta_1^{\frac{1}{2}}\left(\Delta_2 + 2R\lambda^\zeta + \kappa^{2(\zeta-1)}R(\kappa\tau\Delta_4 + \tau\Delta_5 + \lambda(\Delta_3 + \Delta_5)^{(\zeta-1)\wedge 1} + \lambda^{\frac{1}{2}}\Delta_3^{(\zeta-\frac{1}{2})\wedge 1})\right).$$

Using the above into (46), and by a simple calculation, we can prove the second desired result.

A.2. Proofs of Lemma 12

We first introduce the following basic probabilistic estimate.

Lemma 20. *Let $\mathcal{X}_1, \cdots, \mathcal{X}_m$ be a sequence of independently and identically distributed self-adjoint Hilbert-Schmidt operators on a separable Hilbert space. Assume that $\mathbb{E}[\mathcal{X}_1] = 0$, and $\|\mathcal{X}_1\| \leq B$ almost surely for some $B > 0$. Let $\mathcal{V}$ be a positive trace-class operator such that $\mathbb{E}[\mathcal{X}_1^2] \preccurlyeq \mathcal{V}$. Then with probability at least $1-\delta$, ($\delta \in]0,1[$), there holds*

$$\left\| \frac{1}{m} \sum_{i=1}^m \mathcal{X}_i \right\| \leq \frac{2B\beta}{3m} + \sqrt{\frac{2\|\mathcal{V}\|\beta}{m}}, \qquad \beta = \log \frac{4 \operatorname{tr} \mathcal{V}}{\|\mathcal{V}\|\delta}.$$

The above lemma was first proved in (Hsu et al., 2014; Tropp, 2012) for the matrix case, and it was later extended to the general operator case in (Minsker, 2011), see also (Rudi et al., 2015; Bach, 2015; Dicker et al., 2017). We refer to (Rudi et al., 2015; Dicker et al., 2017) for the proof.

Using the above lemma, we can prove Lemma 12.

Proof of Lemma 12. We use Lemma 20 to prove the result. Let $W = m^{-\frac{1}{2}} \mathbf{G} \mathcal{S}_{\mathbf{x}}$. Denote the i-th row of $\mathbf{G}$ by $\mathbf{a}_i^*$ for all $i \in [m]$. Using $\mathcal{T}_{\mathbf{x}} = \mathcal{S}_{\mathbf{x}}^* \mathcal{S}_{\mathbf{x}}$, we have

$$\mathcal{T}_{\mathbf{x}\lambda}^{-\frac{1}{2}} (\mathcal{T}_{\mathbf{x}} - W^* W) \mathcal{T}_{\mathbf{x}\lambda}^{-\frac{1}{2}} = \mathcal{T}_{\mathbf{x}\lambda}^{-\frac{1}{2}} \mathcal{S}_{\mathbf{x}}^* (I - m^{-1} \mathbf{G}^* \mathbf{G}) \mathcal{S}_{\mathbf{x}} \mathcal{T}_{\mathbf{x}\lambda}^{-\frac{1}{2}} = \frac{1}{m} \sum_{i=1}^m \mathcal{X}_i,$$

where we let

$$\mathcal{X}_i = \mathcal{T}_{\mathbf{x}\lambda}^{-\frac{1}{2}} \mathcal{S}_{\mathbf{x}}^* (I - \mathbf{a}_i \mathbf{a}_i^*) \mathcal{S}_{\mathbf{x}} \mathcal{T}_{\mathbf{x}\lambda}^{-\frac{1}{2}}.$$

Since $\mathbf{a}_1 \sim F$, according to the isotropy property (26) of F,

$$\mathbb{E}[\mathcal{X}_1] = \mathcal{T}_{\mathbf{x}\lambda}^{-\frac{1}{2}} \mathcal{S}_{\mathbf{x}}^* (I - \mathbb{E}[\mathbf{a}_i \mathbf{a}_i^*]) \mathcal{S}_{\mathbf{x}} \mathcal{T}_{\mathbf{x}\lambda}^{-\frac{1}{2}} = 0.$$

Note that

$$\|\mathcal{T}_{\mathbf{x}\lambda}^{-\frac{1}{2}} \mathcal{S}_{\mathbf{x}}^* \mathbf{a}_1\|_H = \frac{1}{n} \left\| \sum_{j=1}^n \mathbf{a}_1(j) \mathcal{T}_{\mathbf{x}\lambda}^{-\frac{1}{2}} x_j \right\|_H \leq \frac{1}{n} \sum_{j=1}^n |\mathbf{a}_1(j)| \|\mathcal{T}_{\mathbf{x}\lambda}^{-\frac{1}{2}} x_j\|_H.$$

Using Cauchy-Schwarz inequality and the bounded assumption (27),

$$\|\mathcal{T}_{\mathbf{x}\lambda}^{-\frac{1}{2}} \mathcal{S}_{\mathbf{x}}^* \mathbf{a}_1\|_H \leq \frac{1}{n} \|\mathbf{a}_1\|_2 \left(\sum_{j=1}^n \|\mathcal{T}_{\mathbf{x}\lambda}^{-\frac{1}{2}} x_j\|_H^2 \right)^{\frac{1}{2}} \leq \left(\frac{1}{n} \sum_{j=1}^n \|\mathcal{T}_{\mathbf{x}\lambda}^{-\frac{1}{2}} x_j\|_H^2 \right)^{\frac{1}{2}}.$$

According to $\operatorname{tr}(x \otimes x) = \|x\|_H^2$ and the definition of $\mathcal{T}_{\mathbf{x}}$, we know that the left-hand side is $\sqrt{\operatorname{tr}(\mathcal{T}_{\mathbf{x}\lambda}^{-1} \mathcal{T}_{\mathbf{x}})}$, and thus

$$\|\mathcal{T}_{\mathbf{x}\lambda}^{-\frac{1}{2}} \mathcal{S}_{\mathbf{x}}^* \mathbf{a}_1\|_H \leq \sqrt{\operatorname{tr}(\mathcal{T}_{\mathbf{x}\lambda}^{-1} \mathcal{T}_{\mathbf{x}})}.$$

Therefore,

$$\|\mathcal{T}_{\mathbf{x}\lambda}^{-\frac{1}{2}} \mathcal{S}_{\mathbf{x}}^* \mathbf{a}_1 \mathbf{a}_1^* \mathcal{S}_{\mathbf{x}} \mathcal{T}_{\mathbf{x}\lambda}^{-\frac{1}{2}}\| \leq \operatorname{tr}(\mathcal{T}_{\mathbf{x}\lambda}^{-\frac{1}{2}} \mathcal{S}_{\mathbf{x}}^* \mathbf{a}_1 \mathbf{a}_1^* \mathcal{S}_{\mathbf{x}} \mathcal{T}_{\mathbf{x}\lambda}^{-\frac{1}{2}}) \leq \|\mathcal{T}_{\mathbf{x}\lambda}^{-\frac{1}{2}} \mathcal{S}_{\mathbf{x}}^* \mathbf{a}_1\|_H^2 \leq \operatorname{tr}(\mathcal{T}_{\mathbf{x}\lambda}^{-1} \mathcal{T}_{\mathbf{x}}),$$

and by $\|a - \mathbb{E}[a]\| \leq \|a\| + \mathbb{E}\|a\|$,

$$\|\mathcal{X}_1\| \leq 2 \operatorname{tr}(\mathcal{T}_{\mathbf{x}\lambda}^{-1} \mathcal{T}_{\mathbf{x}}).$$

Moreover, using $\mathbb{E}[a - \mathbb{E}[a]]^2 \preceq \mathbb{E} a^2$,

$$\begin{aligned} \mathbb{E}[\mathcal{X}_1^2] \preceq \mathbb{E}[\mathcal{T}_{\mathbf{x}\lambda}^{-\frac{1}{2}} \mathcal{S}_{\mathbf{x}}^* \mathbf{a}_1 \mathbf{a}_1^* \mathcal{S}_{\mathbf{x}} \mathcal{T}_{\mathbf{x}\lambda}^{-\frac{1}{2}}]^2 &= \mathbb{E}[\|\mathcal{T}_{\mathbf{x}\lambda}^{-\frac{1}{2}} \mathcal{S}_{\mathbf{x}}^* \mathbf{a}_1\|_H^2 \mathcal{T}_{\mathbf{x}\lambda}^{-\frac{1}{2}} \mathcal{S}_{\mathbf{x}}^* \mathbf{a}_1 \mathbf{a}_1^* \mathcal{S}_{\mathbf{x}} \mathcal{T}_{\mathbf{x}\lambda}^{-\frac{1}{2}}] \\ &\preceq \operatorname{tr}(\mathcal{T}_{\mathbf{x}\lambda}^{-1} \mathcal{T}_{\mathbf{x}}) \mathcal{T}_{\mathbf{x}\lambda}^{-\frac{1}{2}} \mathcal{S}_{\mathbf{x}}^* \mathbb{E}[\mathbf{a}_1 \mathbf{a}_1^*] \mathcal{S}_{\mathbf{x}} \mathcal{T}_{\mathbf{x}\lambda}^{-\frac{1}{2}} \\ &= \operatorname{tr}(\mathcal{T}_{\mathbf{x}\lambda}^{-1} \mathcal{T}_{\mathbf{x}}) \mathcal{T}_{\mathbf{x}\lambda}^{-1} \mathcal{T}_{\mathbf{x}}. \end{aligned}$$

Letting $\mathcal{V} = \mathrm{tr}(\mathcal{T}_{\mathbf{x}\lambda}^{-1}\mathcal{T}_{\mathbf{x}})\mathcal{T}_{\mathbf{x}\lambda}^{-1}\mathcal{T}_{\mathbf{x}}$, a simple calculation shows that

$$\|\mathcal{V}\| = \mathrm{tr}(\mathcal{T}_{\mathbf{x}\lambda}^{-1}\mathcal{T}_{\mathbf{x}})\|\mathcal{T}_{\mathbf{x}\lambda}^{-1}\mathcal{T}_{\mathbf{x}}\| \leq \mathrm{tr}(\mathcal{T}_{\mathbf{x}\lambda}^{-1}\mathcal{T}_{\mathbf{x}}).$$

Also, $\|\mathcal{T}_{\mathbf{x}\lambda}^{-1}\mathcal{T}_{\mathbf{x}}\| = \frac{\|\mathcal{T}_{\mathbf{x}}\|}{\|\mathcal{T}_{\mathbf{x}}\|+\lambda}$,

$$\frac{\mathrm{tr}(\mathcal{V})}{\|\mathcal{V}\|} = \frac{\mathrm{tr}(\mathcal{T}_{\mathbf{x}\lambda}^{-1}\mathcal{T}_{\mathbf{x}})}{\|\mathcal{T}_{\mathbf{x}\lambda}^{-1}\mathcal{T}_{\mathbf{x}}\|} = \mathrm{tr}(\mathcal{T}_{\mathbf{x}\lambda}^{-1}\mathcal{T}_{\mathbf{x}})\left(1 + \frac{\lambda}{\|\mathcal{T}_{\mathbf{x}}\|}\right).$$

Applying Lemma 20, one can prove the desired result. □

A.3. Proof of Lemma 13

If $\lambda \geq \|\mathcal{T}_{\mathbf{x}}\|$, then the result follows trivially,

$$\|(I-P)\mathcal{T}_{\mathbf{x}}^{\frac{1}{2}}\|^2 \leq \|(I-P)\|^2\|\mathcal{T}_{\mathbf{x}}\| \leq \frac{1}{n^{\theta}}.$$

We thus only need to consider the case $\lambda \leq \|\mathcal{T}_{\mathbf{x}}\|$. Let $M = m^{-1}\mathcal{S}_{\mathbf{x}}^*\mathbf{G}^*\mathbf{G}\mathcal{S}_{\mathbf{x}}$ and $M_\lambda = M + \lambda I$. Applying Lemma 12, we know that there exists a subset $U_{\mathbf{x}}$ of $\mathbb{R}^{m\times n}$ with measure at least $1-\delta$, such that

$$\left\|\mathcal{T}_{\mathbf{x}\lambda}^{-1/2}(\mathcal{T}_{\mathbf{x}} - M)\mathcal{T}_{\mathbf{x}\lambda}^{-1/2}\right\| \leq \frac{4\mathcal{N}_{\mathbf{x}}(\lambda)\beta}{3m} + \sqrt{\frac{2\mathcal{N}_{\mathbf{x}}(\lambda)\beta}{m}}, \quad \forall \mathbf{G} \in U_{\mathbf{x}}. \tag{56}$$

Using Condition (39),

$$\mathcal{N}_{\mathbf{x}}(\lambda) \leq b_\gamma n^{\theta\gamma}.$$

With $\lambda \leq \|\mathcal{T}_{\mathbf{x}}\|$, we have

$$\beta \leq \log\frac{4b_\gamma n^{\theta\gamma}(1+\lambda/\|\mathcal{T}_{\mathbf{x}}\|)}{\delta} \leq \log\frac{8b_\gamma n^{\theta\gamma}}{\delta},$$

and, combining with (40),

$$\frac{4\mathcal{N}_{\mathbf{x}}(\lambda)\beta}{3m} + \sqrt{\frac{2\mathcal{N}_{\mathbf{x}}(\lambda)\beta}{m}} \leq \frac{2}{3}.$$

Thus,

$$\left\|\mathcal{T}_{\mathbf{x}\lambda}^{-1/2}(\mathcal{T} - M)\mathcal{T}_{\mathbf{x}\lambda}^{-1/2}\right\| \leq \frac{2}{3}, \quad \forall \mathbf{G} \in U_{\mathbf{x}}.$$

Following from (Caponnetto & De Vito, 2007),

$$\|M_\lambda^{-1/2}\mathcal{T}_{\mathbf{x}\lambda}^{\frac{1}{2}}\|^2 = \|\mathcal{T}_{\mathbf{x}\lambda}^{\frac{1}{2}}M_\lambda^{-1}\mathcal{T}_{\mathbf{x}\lambda}^{\frac{1}{2}}\|^2 = \|(I - \mathcal{T}_{\mathbf{x}\lambda}^{-1/2}(\mathcal{T}_{\mathbf{x}} - M)\mathcal{T}_{\mathbf{x}\lambda}^{-1/2})^{-1/2}\|,$$

we get

$$\|M_{\lambda'}^{-1/2}\mathcal{T}_{\mathbf{x}\lambda}^{\frac{1}{2}}\|^2 \leq 3, \quad \forall \mathbf{G} \in U_{\mathbf{x}}. \tag{57}$$

Let $W = m^{-1/2}\mathbf{G}\mathcal{S}_{\mathbf{x}}$. As P is the projection operator onto $\overline{range\{W^*\}}$,

$$P = W^*(WW^*)^\dagger W \succeq W^*(WW^*+\lambda)^{-1}W = W^*W(W^*W+\lambda)^{-1} = M(M+\lambda)^{-1},$$

where for the last second equality, we used Lemma 17. Thus (Rudi et al., 2015),

$$I - P \preceq I - M(M+\lambda)^{-1} = \lambda(M+\lambda)^{-1}.$$

It thus follows that

$$\mathcal{T}_{\mathbf{x}}^{\frac{1}{2}}(I-P)\mathcal{T}_{\mathbf{x}}^{\frac{1}{2}} \preceq \lambda\mathcal{T}_{\mathbf{x}}^{\frac{1}{2}}(M+\lambda)^{-1}\mathcal{T}_{\mathbf{x}}^{\frac{1}{2}} \preceq \lambda\mathcal{T}_{\mathbf{x}\lambda}^{\frac{1}{2}}(M+\lambda)^{-1}\mathcal{T}_{\mathbf{x}\lambda}^{\frac{1}{2}}.$$

Using $\|A^*A\|^2 = \|A\|^2$ and the above,

$$\|(I-P)\mathcal{T}_{\mathbf{x}}^{\frac{1}{2}}\|^2 = \|\mathcal{T}_{\mathbf{x}}^{\frac{1}{2}}(I-P)\mathcal{T}_{\mathbf{x}}^{\frac{1}{2}}\| \leq \lambda\|\mathcal{T}_{\mathbf{x}\lambda}^{\frac{1}{2}}(M+\lambda)^{-1}\mathcal{T}_{\mathbf{x}\lambda}^{\frac{1}{2}}\| = \lambda\|(M+\lambda)^{-1/2}\mathcal{T}_{\mathbf{x}\lambda}^{\frac{1}{2}}\|^2. \tag{58}$$

Applying (57), one can prove the desired result.

A.4. Proof of Lemma 14

Since P is a projection operator, $(I-P)^2 = I-P$. Then

$$\|A^s(I-P)A^t\| = \|A^s(I-P)(I-P)A^t\| \le \|A^s(I-P)\|\|(I-P)A^t\|.$$

Moreover, by Lemma 16,

$$\|A^s(I-P)\| = \|A^{\frac{1}{2}2s}(I-P)^{2s}\| \le \|A^{\frac{1}{2}}(I-P)\|^{2s}.$$

Similarly, $\|(I-P)A^t\| \le \|(I-P)A^{\frac{1}{2}}\|^{2t}$. Thus,

$$\|A^s(I-P)A^t\| \le \|A^{\frac{1}{2}}(I-P)\|^{2s}\|(I-P)A^{\frac{1}{2}}\|^{2t} = \|(I-P)A^{\frac{1}{2}}\|^{2(t+s)}.$$

Using $\|D\|^2 = \|D^*D\|$,

$$\|A^s(I-P)A^t\| \le \|(I-P)A(I-P)\|^{t+s}.$$

Adding and subtracting with the same term, using the triangle inequality, and noting that $\|I-P\| \le 1$ and $s+t \le 1$,

$$\begin{aligned}\|A^s(I-P)A^t\| \le & \|(I-P)A(I-P)\|^{t+s} \\ \le & (\|(I-P)(A-B)(I-P)\| + \|(I-P)B(I-P)\|)^{t+s} \\ \le & \|A-B\|^{s+t} + \|(I-P)B(I-P)\|^{s+t},\end{aligned}$$

which leads to the desired result using $\|D^*D\| = \|DD^*\|$.

A.5. Proof of Lemma 15

To prove the result, we need the following concentration inequality.

Lemma 21. *Let $w_1, \cdots, w_m$ be i.i.d random variables in a separable Hilbert space with norm $\|\cdot\|$. Suppose that there are two positive constants B and σ^2 such that*

$$\mathbb{E}[\|w_1 - \mathbb{E}[w_1]\|^l] \le \frac{1}{2} l! B^{l-2}\sigma^2, \quad \forall l \ge 2. \tag{59}$$

Then for any $0 < \delta < 1/2$, the following holds with probability at least $1-\delta$,

$$\left\| \frac{1}{m}\sum_{k=1}^m w_m - \mathbb{E}[w_1] \right\| \le 2\left(\frac{B}{m} + \frac{\sigma}{\sqrt{m}}\right)\log\frac{2}{\delta}.$$

In particular, (59) holds if

$$\|w_1\| \le B/2 \ \text{ a.s.}, \quad \text{and} \quad \mathbb{E}[\|w_1\|^2] \le \sigma^2. \tag{60}$$

The above lemma is a reformulation of the concentration inequality for sums of Hilbert-space-valued random variables from (Pinelis & Sakhanenko, 1986). We refer to (Smale & Zhou, 2007; Caponnetto & De Vito, 2007) for the detailed proof.

Proof of Lemma 15. We first use Lemma 21 to estimate $\mathrm{tr}(\mathcal{T}_\lambda^{-\frac{1}{2}}(\mathcal{T}_{\mathbf{x}} - \mathcal{T})\mathcal{T}_\lambda^{-\frac{1}{2}})$. Note that

$$\mathrm{tr}(\mathcal{T}_\lambda^{-\frac{1}{2}}\mathcal{T}_{\mathbf{x}}\mathcal{T}_\lambda^{-\frac{1}{2}}) = \frac{1}{n}\sum_{j=1}^n \|\mathcal{T}_\lambda^{-\frac{1}{2}} x_j\|_H^2 = \frac{1}{n}\sum_{j=1}^n \xi_j,$$

where we let $\xi_j = \|\mathcal{T}_\lambda^{-\frac{1}{2}} x_j\|_H^2$ for all $j \in [n]$. Besides, it is easy to see that

$$\mathrm{tr}(\mathcal{T}_\lambda^{-\frac{1}{2}}(\mathcal{T}_{\mathbf{x}} - \mathcal{T})\mathcal{T}_\lambda^{-\frac{1}{2}}) = \frac{1}{n}\sum_{j=1}^n(\xi_j - \mathbb{E}[\xi_j]).$$

Using Assumption (2),

$$\xi_1 \leq \frac{1}{\lambda}\|x_1\|_H^2 \leq \frac{\kappa^2}{\lambda},$$

and

$$\mathbb{E}[\|\xi_1\|^2] \leq \frac{\kappa^2}{\lambda}\mathbb{E}\|\mathcal{T}_\lambda^{-\frac{1}{2}} x_1\|_H^2 \leq \frac{\kappa^2\mathcal{N}(\lambda)}{\lambda}.$$

Applying Lemma 21, we get that there exists a subset V_1 of Z^n with measure at least $1-\delta$, such that for all $\mathbf{z} \in V_1$,

$$\mathrm{tr}(\mathcal{T}_\lambda^{-\frac{1}{2}}(\mathcal{T}_{\mathbf{x}} - \mathcal{T})\mathcal{T}_\lambda^{-\frac{1}{2}}) \leq 2\left(\frac{2\kappa^2}{n\lambda} + \sqrt{\frac{\kappa^2\mathcal{N}(\lambda)}{n\lambda}}\right)\log\frac{2}{\delta}.$$

Combining with Lemma 8, taking the union bounds, rescaling δ, and noting that

$$\begin{aligned}
\mathrm{tr}(\mathcal{T}_{\mathbf{x}\lambda}^{-1}\mathcal{T}_{\mathbf{x}}) =& \,\mathrm{tr}(\mathcal{T}_{\mathbf{x}\lambda}^{-\frac{1}{2}}\mathcal{T}_\lambda^{\frac{1}{2}}\mathcal{T}_\lambda^{-\frac{1}{2}}\mathcal{T}_{\mathbf{x}}\mathcal{T}_\lambda^{-\frac{1}{2}}\mathcal{T}_\lambda^{\frac{1}{2}}\mathcal{T}_{\mathbf{x}\lambda}^{-\frac{1}{2}}) \\
\leq& \|\mathcal{T}_\lambda^{\frac{1}{2}}\mathcal{T}_{\mathbf{x}\lambda}^{-\frac{1}{2}}\|^2\,\mathrm{tr}(\mathcal{T}_\lambda^{-\frac{1}{2}}\mathcal{T}_{\mathbf{x}}\mathcal{T}_\lambda^{-\frac{1}{2}}) \\
=& \|\mathcal{T}_\lambda^{\frac{1}{2}}\mathcal{T}_{\mathbf{x}\lambda}^{-\frac{1}{2}}\|^2\left(\mathrm{tr}(\mathcal{T}_\lambda^{-\frac{1}{2}}(\mathcal{T}_{\mathbf{x}} - \mathcal{T})\mathcal{T}_\lambda^{-\frac{1}{2}}) + \mathcal{N}(\lambda)\right).
\end{aligned}$$

we get that there exists a subset V of Z^n with measure at least $1-\delta$, such that for all $\mathbf{z} \in V$,

$$\mathrm{tr}((\mathcal{T}_{\mathbf{x}} + \lambda)^{-1}\mathcal{T}_{\mathbf{x}}) \leq 3a_{n,\delta/2,\gamma}(\theta)\left(2\left(\frac{2\kappa^2}{n\lambda} + \sqrt{\frac{\kappa^2\mathcal{N}(\lambda)}{n\lambda}}\right)\log\frac{4}{\delta} + \mathcal{N}(\lambda)\right),$$

which leads to the desired result using $\lambda \leq 1$, $n\lambda \geq 1$ and Assumption 3. □

A.6. Proof for Corollary 5

Proof. Using a similar argument as that for (58), with $W = \mathcal{S}_{\tilde{\mathbf{x}}}$, where $\tilde{\mathbf{x}} = \{x_1, \cdots, x_m\}$, we get for any $\eta > 0$,

$$\|(I-P)\mathcal{T}^{\frac{1}{2}}\|^2 \leq \eta\|(\mathcal{T}_{\tilde{\mathbf{x}}} + \eta)^{-1/2}(\mathcal{T} + \eta)^{1/2}\|^2.$$

Letting $\eta = \frac{1}{m}$, and using Lemma 8, we get that with probability at least $1-\delta$,

$$\|(I-P)\mathcal{T}^{\frac{1}{2}}\|^2 \lesssim \frac{1}{m}\log\frac{3m^\gamma}{\delta}.$$

Combining with Corollary 3, one can prove the desired result. □

B. Learning with Kernel Methods

Let the input space Ξ be a closed subset of Euclidean space $\mathbb{R}^d$, the output space $Y \subseteq \mathbb{R}$. Let μ be an unknown but fixed Borel probability measure on $\Xi \times Y$. Assume that $\{(\xi_i, y_i)\}_{i=1}^m$ are i.i.d. from the distribution μ. A reproducing kernel K is a symmetric function $K : \Xi \times \Xi \to \mathbb{R}$ such that $(K(u_i, u_j))_{i,j=1}^\ell$ is positive semidefinite for any finite set of points $\{u_i\}_{i=1}^\ell$ in Ξ. The kernel K defines a reproducing kernel Hilbert space (RKHS) $(\mathcal{H}_K, \|\cdot\|_K)$ as the completion of the linear span of the set $\{K_\xi(\cdot) := K(\xi, \cdot) : \xi \in \Xi\}$ with respect to the inner product $\langle K_\xi, K_u\rangle_K := K(\xi, u)$. For any $f \in \mathcal{H}_K$, the reproducing property holds: $f(\xi) = \langle K_\xi, f\rangle_K$.

Example B.1 (Sobolev Spaces)**.** *Let $X = [0,1]$ and the kernel*

$$K(x, x') = \begin{cases} (1-y)x, & x \leq y; \\ (1-x)y, & x \geq y. \end{cases}$$

Then the kernel induces a Sobolev Space $H = \{f : X \to \mathbb{R} | f$ is absolutely continuous $, f(0) = f(1) = 0, f \in L^2(X)\}$.

In learning with kernel methods, one considers the following minimization problem

$$\inf_{f\in\mathcal{H}_K}\int_{\Xi\times Y}(f(\xi)-y)^2 d\mu(\xi,y).$$

Since $f(\xi)=\langle K_\xi, f\rangle_K$ by the reproducing property, the above can be rewritten as

$$\inf_{f\in\mathcal{H}_K}\int_{\Xi\times Y}(\langle f, K_\xi\rangle_K-y)^2 d\mu(\xi,y).$$

Letting $X=\{K_\xi:\xi\in\Xi\}$ and defining another probability measure $\rho(K_\xi,y)=\mu(\xi,y)$, the above reduces to the learning setting in Section 2.

C. An Example for the Non-attainable Case

Let H be the usual sequence space l_2 of all infinite sequences of real numbers with its norm given by

$$\|\xi\|_2=\left(\sum_{i=1}^{\infty}\xi_i^2\right)^{\frac{1}{2}},\quad \forall\xi\in l_2.$$

Assume that the random vector $x=(x_1,\cdots,x_j,...,x_\infty)$ is drawn according to the distribution ρ_X, where $x_j=e_j/j$ for all j and $e_1,\cdots,e_\infty$ are independent Bernoulli random variables. Define the regression function $f_\rho(x)=\sum_{j=1}^{\infty}x_j=\langle\mathbf{1},x\rangle_{l_2}$, where $\mathbf{1}=(1,\cdots,1)$. Using a direct computation, it is relatively easy to show that $f_H=f_\rho\in\overline{H_\rho}$, but $f_H\notin H_\rho$.