A Supplementary Results

A.1 Proof of Lemma 1

Proof. Note that

$$R(P) = \mathbb{E}\,\|(I-P)k(\cdot,X)\|^2_{\mathcal{H}} = \mathbb{E}\,\langle (I-P)k(\cdot,X),(I-P)k(\cdot,X)\rangle_{\mathcal{H}}\,,$$

which in turn is equivalent to

$$\mathbb{E}\,\langle (I-P)k(\cdot,X),k(\cdot,X)\rangle_{\mathcal{H}} = \mathbb{E}\,\langle (I-P),k(\cdot,X)\otimes_{\mathcal{H}} k(\cdot,X)\rangle_{\mathcal{L}^2(\mathcal{H})}\,,$$

where we used $\langle Bf,g\rangle_{\mathcal{H}} = \langle B, f\otimes_{\mathcal{H}} g\rangle_{\mathcal{L}^2(\mathcal{H})}$ and $(I-P)^2=(I-P)$ in the above equivalence. Since k is bounded, it follows that

$$\mathbb{E}\,\langle (I-P),k(\cdot,X)\otimes_{\mathcal{H}} k(\cdot,X)\rangle_{\mathcal{L}^2(\mathcal{H})} = \langle (I-P),\mathbb{E}[k(\cdot,X)\otimes_{\mathcal{H}} k(\cdot,X)]\rangle_{\mathcal{L}^2(\mathcal{H})}\,.$$

The result follows by using the above in $R(P)$ and noting that

$$\langle (I-P),C\rangle_{\mathcal{L}^2(\mathcal{H})} = \mathrm{tr}\left((I-P)C\right) = \mathrm{tr}\left(C^{1/2}(I-P)(I-P)C^{1/2}\right) = \left\|(I-P)C^{1/2}\right\|^2_{\mathcal{L}^2(\mathcal{H})}\,,$$

where we have used the invariance of trace under cyclic permutations. □

Lemma A.1. *For $\delta>0$, suppose $\frac{9\kappa}{n}\log\frac{n}{\delta}\le t\le\lambda_1$. Then the following hold:*

(i) $\mathbb{P}^n\left\{\sqrt{\frac{2}{3}}\le\left\|(C+tI)^{1/2}(C_n+tI)^{-1/2}\right\|_{\mathcal{L}^\infty(\mathcal{H})}\le\sqrt{2}\right\}\ge 1-\delta;$

(ii) $\mathbb{P}^n\left\{\left\|(C+tI)^{-1/2}(C_n+tI)^{1/2}\right\|_{\mathcal{L}^\infty(\mathcal{H})}\le\sqrt{\frac{3}{2}}\right\}\ge 1-\delta;$

(iii) $\mathbb{P}^n\left\{\hat{\lambda}_\ell+t\le\frac{3}{2}(\lambda_\ell+t)\right\}\ge 1-\delta.$

(iv) $\mathbb{P}^n\left\{\lambda_\ell+t\le 2(\widehat{\lambda}_\ell+t)\right\}\ge 1-\delta.$

Proof. (i) The result is quoted from Lemma 3.6 of (Rudi et al., 2013) with $\alpha=\frac{1}{2}$.

(ii) This is a slight variation of (i) and the proof idea follows that of Lemma 3.6 of (Rudi et al., 2013) with $\alpha=\frac{1}{2}$. Note that

$$\left\|(C+tI)^{-1/2}(C_n+tI)^{1/2}\right\|_{\mathcal{L}^\infty(\mathcal{H})} = \left\|(C+tI)^{-1/2}(C_n+tI)(C+tI)^{-1/2}\right\|^{1/2}_{\mathcal{L}^\infty(\mathcal{H})}.$$

By defining $B_n=(C+tI)^{-1/2}(C-C_n)(C+tI)^{-1/2}$, we have

$$I-B_n=(C+tI)^{-1/2}\left((C+tI)-C+C_n\right)(C+tI)^{-1/2}=(C+tI)^{-1/2}(C_n+tI)(C+tI)^{-1/2}$$

and therefore

$$\left\|(C+tI)^{-1/2}(C_n+tI)^{1/2}\right\|_{\mathcal{L}^\infty(\mathcal{H})} = \|I-B_n\|^{1/2}_{\mathcal{L}^\infty(\mathcal{H})}\le\left(1+\|B_n\|_{\mathcal{L}^\infty(\mathcal{H})}\right)^{1/2}. \tag{18}$$

It follow from the proof of Lemma 3.6 of (Rudi et al., 2013) that for $\frac{9\kappa}{n}\log\frac{n}{\delta}\le t$,

$$\mathbb{P}^n\left\{\|B_n\|_{\mathcal{L}^\infty(\mathcal{H})}\le\frac{1}{2}\right\}\ge 1-\delta. \tag{19}$$

Combining (18) and (19) completes the proof.

(iii) Since $\sqrt{\frac{2}{3}}\le\left\|(C+tI)^{1/2}(C_n+tI)^{-1/2}\right\|_{\mathcal{L}^\infty(\mathcal{H})}$ as obtained in (i), it is equivalent (see (Rudi et al., 2013, Lemmas B.2 and 3.5)) to $C_n+tI\preceq\frac{3}{2}(C+tI)$. This implies (see Gohberg et al., 2003) that $\widehat{\lambda}_k+t\le\frac{3}{2}(\lambda_k+t)$ for all $k\ge 1$. (iv) follows similarly. □

Lemma A.2 (Rudi et al. (2015), Lemma 6)**.** *Suppose Assumption 1 holds, and suppose for some $m < n$, the set $\{\tilde{X}_j\}_{j=1}^m$ is drawn uniformly from the set of all partitions of size m of the training data, $\{X_i\}_{i=1}^n$. For $t > 0$ and any $\delta > 0$ such that $m \geq (67 \vee 5\mathcal{N}_{C,\infty}(t)) \log \frac{4\kappa}{t\delta}$, we have*

$$\mathbb{P}^n \left\{ \left\| (I - P_m)(C + tI)^{1/2} \right\|^2_{\mathcal{L}^\infty(\mathcal{H})} \leq 3t \right\} \geq 1 - \delta,$$

where P_m is the orthogonal projector onto $\mathcal{H}_m = \mathrm{span}\{k(\cdot, \tilde{X}_j) | j \in [m]\}$.

Lemma A.3 (Rudi et al. (2015), Lemma 7)**.** *Suppose Assumption 1 holds. Let $(\hat{l}_i(s))_{i=1}^n$ be the collection of approximate leverage scores. Letting $N := \{1, ..., n\}$, for $t > 0$ define p_t as the distribution over N with probabilities $p_t(i) = \hat{l}_i(t) / \sum_{j=1}^n \hat{l}_j(t)$. Let $\mathcal{I}_m = \{i_1, ..., i_m\} \subset N$ be a collection of indices independently sampled from p_t with replacement. Let P_m be the orthogonal projector onto $\mathcal{H}_m = \mathrm{span}\{k(\cdot, \tilde{X}_j) | j \in \mathcal{I}_m\}$. Additionally, for any $\delta > 0$, suppose the following hold:*

1. *There exists $T \geq 1$ and $t_0 > 0$ such that for any $s \geq t_0$, $(\hat{l}_i(s))_{i=1}^n$ are $T-$approximate leverage scores with confidence δ,*
2. *$n \geq 1655\kappa + 223\kappa \log \frac{2\kappa}{\delta}$,*
3. *$t_0 \vee \frac{19\kappa}{n} \log \frac{2n}{\delta} \leq t \leq \lambda_1$,*
4. *$m \geq 334 \log \frac{8n}{\delta} \vee 78T^2 \mathcal{N}_C(t) \log \frac{8n}{\delta}$.*

Then

$$\mathbb{P}^n \left\{ \left\| (I - P_m)(C + tI)^{1/2} \right\|^2_{\mathcal{L}^\infty(\mathcal{H})} \leq 3t \right\} \geq 1 - 2\delta.$$

B Technical Results

Proposition B.1. *Suppose $\underline{A} i^{-\alpha} \leq \lambda_i \leq \bar{A} i^{-\alpha}$ for $\alpha > 1$ and $\underline{A}, \bar{A} \in (0, \infty)$. The following holds:*

$$\mathcal{N}_C(t) \lesssim t^{-1/\alpha}.$$

Proof. We have

$$\mathcal{N}_C(t) = \mathrm{tr}\left((C + tI)^{-1} C\right) = \sum_{i \geq 1} \frac{\lambda_i}{\lambda_i + t} \leq \sum_{i \geq 1} \frac{\bar{A} i^{-\alpha}}{\underline{A} i^{-\alpha} + t} = \frac{\bar{A}}{\underline{A}} \sum_{i \geq 1} \frac{i^{-\alpha}}{i^{-\alpha} + t \underline{A}^{-1}}.$$

Let $u = t^{1/\alpha} \underline{A}^{-1/\alpha} x \implies u^\alpha = t \underline{A}^{-1} x^\alpha$ and $dx = t^{-1/\alpha} \underline{A}^{1/\alpha} du$. Therefore,

$$\sum_{i \geq 1} \frac{i^{-\alpha}}{i^{-\alpha} + t \underline{A}^{-1}} \leq \int_0^\infty \frac{x^{-\alpha}}{x^{-\alpha} + t \underline{A}^{-1}} \, dx = \int_0^\infty \frac{1}{1 + t \underline{A}^{-1} x^\alpha} dx = \left(\frac{\underline{A}}{t}\right)^{1/\alpha} \int_0^\infty \frac{1}{1 + u^\alpha} du.$$

Since $\frac{1}{1+u^\alpha}$ is decreasing in α on $u \in (0, \infty)$, we have

$$\frac{1}{1 + u^\alpha} \leq \frac{1}{1 + u^2}, \quad \text{if} \quad \alpha \geq 2.$$

So for $\alpha \geq 2$,

$$\left(\frac{\underline{A}}{t}\right)^{1/\alpha} \int_0^\infty \frac{1}{1 + u^\alpha} du \lesssim t^{-1/\alpha} \int_0^\infty \frac{1}{1 + u^2} du = t^{-1/\alpha} \left[\tan^{-1}(u)|_0^\infty\right] = \frac{\pi}{2} t^{-1/\alpha},$$

implying $\mathcal{N}_C(t) \lesssim t^{-1/\alpha}$. For $1 < \alpha < 2$, we obtain

$$t^{-1/\alpha} \int_0^\infty \frac{1}{1 + u^\alpha} du \leq t^{-1/\alpha} \sum_{k=0}^\infty \frac{1}{1 + k^\alpha} \leq t^{-1/\alpha} \left(1 + \sum_{k=1}^\infty \frac{1}{k^\alpha}\right).$$

Since $1 + \sum_{k=1}^\infty \frac{1}{k^\alpha}$ converges for $\alpha > 1$, we obtain $\mathcal{N}_C(t) \lesssim t^{-1/\alpha}$. □

Proposition B.2. *Suppose* $\underline{B}e^{-\tau i} \leq \lambda_i \leq \bar{B}e^{-\tau i}$ *for* $\tau > 0$ *and* $\underline{B}, \bar{B} \in (0, \infty)$*. Let* $\ell = \frac{1}{\tau} \log n^\theta$*,* $\theta > 0$*. Then*

$$\mathcal{N}_C(t) \lesssim \log\left(\frac{1}{t}\right).$$

Proof. We have

$$\mathcal{N}_C(t) = \operatorname{tr}\left((C + tI)^{-1}C\right) = \sum_{i \geq 1} \frac{\lambda_i}{\lambda_i + t} \leq \frac{\bar{B}e^{-\tau i}}{\underline{\mathrm{B}}e^{-\tau i} + t} = \frac{\bar{B}}{\underline{\mathrm{B}}} \sum_{i \geq 1} \frac{1}{1 + t\underline{\mathrm{B}}^{-1}e^{\tau i}}$$

$$\lesssim \int_0^\infty \frac{1}{1 + t\underline{\mathrm{B}}^{-1}e^{\tau x}} dx = \left[x - \frac{1}{\tau}\log\left(t\underline{\mathrm{B}}^{-1}\, e^{\tau x} + 1\right)\right]\Big|_0^\infty.$$

Since

$$x - \frac{1}{\tau}\log\left(t\underline{\mathrm{B}}^{-1}\, e^{\tau x} + 1\right) = \frac{1}{\tau}\left(\log(e^{\tau x}) - \log\left(t\underline{\mathrm{B}}^{-1}\, e^{\tau x} + 1\right)\right) = \frac{1}{\tau}\log\left(t^{-1}\underline{\mathrm{B}}\frac{e^{\tau x}}{e^{\tau x} + t^{-1}\underline{\mathrm{B}}}\right),$$

evaluating

$$\frac{1}{\tau}\log\left(t^{-1}\underline{\mathrm{B}}\frac{e^{\tau x}}{e^{\tau x} + t^{-1}\underline{\mathrm{B}}}\right)\Big|_0^\infty$$

yields the result. □