KERNEL ADDITIVE SLICED INVERSE REGRESSION

Heng Lian and Qin Wang

University of New South Wales and Virginia Commonwealth University

Supplementary Material

S1 Proof of the main theorem.

In the proofs, C denotes a generic positive constant. We first note that $E_{X^*}[(\hat{f}(X^*)-f(X^*))^2] = \|\Sigma^{1/2}(\hat{f}-f)\|^2_{\mathcal{H}}$. From (2.3), $\Sigma^{1/2}\hat{f}$ satisfies the eigenvalue equation

$$\Sigma^{1/2}(\Sigma_n+sI)^{-1}\Gamma_n\Sigma^{-1/2}(\Sigma^{1/2}\hat{f}) = \lambda\Sigma^{1/2}\hat{f}.$$

Similarly, $\Sigma^{1/2}f$ satisfies the eigenvalue equation

$$\Sigma^{1/2}\Sigma^{-1}\Gamma\Sigma^{-1/2}(\Sigma^{1/2}f) = \lambda\Sigma^{1/2}f.$$

Using the perturbation theory for operators, for example as in Chapter 4 of Kato (1995), we only need to show that $\|\Sigma^{1/2}(\Sigma_n+sI)^{-1}\Gamma_n\Sigma^{-1/2} - \Sigma^{1/2}\Sigma^{-1}\Gamma\Sigma^{-1/2}\|^2 = O_p(c_n n^{-d/(d+1)})$, where $\|A\|$ denotes the operator norm for an operator A defined on $\mathcal{H}_K$.

We write

$$\begin{aligned}
& \|\Sigma^{1/2}(\Sigma_n+sI)^{-1}\Gamma_n\Sigma^{-1/2} - \Sigma^{1/2}\Sigma^{-1}\Gamma\Sigma^{-1/2}\| \\
= & \|\Sigma^{1/2}((\Sigma_n+sI)^{-1}\Gamma_n - (\Sigma+sI)^{-1}\Gamma + (\Sigma+sI)^{-1}\Gamma - \Sigma^{-1}\Gamma)\Sigma^{-1/2}\| \\
= & \|\Sigma^{1/2}((\Sigma+sI)^{-1}(\Gamma_n-\Gamma) + (\Sigma+sI)^{-1}(\Sigma-\Sigma_n)(\Sigma_n+sI)^{-1}\Gamma_n \\
& +(\Sigma+sI)^{-1}\Gamma - \Sigma^{-1}\Gamma)\Sigma^{-1/2}\| \\
\leq & \|\Sigma^{1/2}(\Sigma+sI)^{-1}(\Gamma_n-\Gamma)\Sigma^{-1/2}\| + \|\Sigma^{1/2}(\Sigma+sI)^{-1}(\Sigma-\Sigma_n)(\Sigma_n+sI)^{-1}\Gamma_n\Sigma^{-1/2}\| \\
& +\|\Sigma^{1/2}((\Sigma+sI)^{-1}\Gamma - \Sigma^{-1}\Gamma)\Sigma^{-1/2}\| \\
=: & (I)+(II)+(III). \qquad \text{(S1.1)}
\end{aligned}$$

To simplify the proofs and notations a little bit, we assume $\overline{K(.,X)} = 0$ in the following, since using $\|\overline{K(.,X)}\|_{\mathcal{H}} = O_p(n^{-1/2})$, such terms does not lead to extra difficulties in the proof. By the same reason, we also replace $\hat{p}_h$ by $p_h = P(Y = y_h)$ in the following expression of Γ_n. Obviously we can rewrite Γ_n and Γ as

$$\begin{aligned}
\Gamma_n &= \sum_{h=1}^H \frac{1}{p_h}\left(\frac{1}{n}\sum_{i=1}^n K(.,X_i)I\{Y_i=y_h\}\right) \otimes \left(\frac{1}{n}\sum_{i=1}^n K(.,X_i)I\{Y_i=y_h\}\right), \\
\Gamma &= \sum_{h=1}^H \frac{1}{p_h}E[(K(.,X)I\{Y=y_h\})] \otimes E[K(.,X)I\{Y=y_h\}]. \qquad \text{(S1.2)}
\end{aligned}$$

The term (III) is easy to deal with as follows. Using

$$\begin{aligned}
& \|\Sigma^{1/2}((\Sigma+sI)^{-1}-\Sigma^{-1})(E[K(.,X)I\{Y=y_h\}]\otimes E[K(.,X)I\{Y=y_h\}])\Sigma^{-1/2}\|^2 \\
= & \|s\Sigma^{1/2}(\Sigma+sI)^{-1}(\Sigma^{-1}E[K(.,X)I\{Y=y_h\}])\otimes(\Sigma^{-1/2}E[K(.,X)I\{Y=y_h\}])\|^2 \\
= & O(\|s\Sigma^{1/2}(\Sigma+sI)^{-1}\|^2) \\
= & O(s),
\end{aligned}$$

we have $(III)^2=O(s)$.

For the term (II), writing $\Sigma(x)=K(.,x)\otimes K(.,x)$ and using that

$$\begin{aligned}
& E\|\Sigma^{1/2}(\Sigma+sI)^{-1}\Sigma(x)\|_{HS}^2 \\
= & E\mathrm{tr}(\Sigma(x)^2(\Sigma+sI)^{-1}\Sigma(\Sigma+sI)^{-1}) \\
\leq & CE\mathrm{tr}(\Sigma(x)(\Sigma+sI)^{-1}\Sigma(\Sigma+sI)^{-1}) \\
= & C\mathrm{tr}(\Sigma(\Sigma+sI)^{-1}\Sigma(\Sigma+sI)^{-1}) \\
= & C\sum_j \frac{\lambda_j^2}{(\lambda_j+s)^2},
\end{aligned}$$

where $\|.\|_{HS}$ is the Hilbert-Schmidt norm. In the inequality above we used that $\|\Sigma(x)f\|_{\mathcal{H}}=\|K(.,x)f(x)\|_{\mathcal{H}}=\sqrt{f^2(x)K(x,x)}\leq C\|f\|_\infty\leq C\|f\|_{\mathcal{H}}$ and thus $\|\Sigma(x)\|\leq C$, and the inequality $\mathrm{tr}(AB)\leq\|A\|\mathrm{tr}(B)$.

Thus using the Markov inequality, we have

$$\begin{aligned}
& (II)^2 \\
\leq & \|\Sigma^{1/2}(\Sigma+sI)^{-1}(\Sigma-\Sigma_n)\|^2\cdot\|(\Sigma_n+sI)^{-1}\Gamma_n\Sigma^{-1/2}\|^2 \\
= & O_p\left(\sum_j \frac{\lambda_j^2}{n(\lambda_j+s)^2}\right)\|(\Sigma_n+sI)^{-1}\Gamma_n\Sigma^{-1/2}\|^2.
\end{aligned}$$

We will argue later that actually $\|(\Sigma_n+sI)^{-1}\Gamma_n\Sigma^{-1/2}\|^2=O_p(1)$.

The term (I) is more complicated. Let Γ_{nh} and Γ_h be the terms on the right hand side of the sums in (S1.2) such that $\Gamma_n=\sum_h p_h^{-1}\Gamma_{nh}, \Gamma=\sum_h p_h^{-1}\Gamma_h$. To bridge Γ_{nh} and Γ_h, we further define

$$\Gamma'_{nh}=\left(\frac{1}{n}\sum_{i=1}^n K(.,X_i)I\{Y_i=y_h\}\right)\otimes E[K(.,X)I\{Y=y_h\}].$$

Since

$$
\begin{aligned}
& E\|\Sigma^{1/2}(\Sigma+sI)^{-1}(K(.,X)I\{Y=y_h\}-E[K(.,X)I\{Y=y_h\}]) \\
& \qquad \otimes(\Sigma^{-1/2}E[K(.,X)I\{Y=y_h\}])\|_{HS}^2 \\
\leq\ & CE\|\Sigma^{1/2}(\Sigma+sI)^{-1}(K(.,X)I\{Y=y_h\}-E[K(.,X)I\{Y=y_h\}])\|_{\mathcal{H}}^2 \\
\leq\ & CE\|\Sigma^{1/2}(\Sigma+sI)^{-1}K(.,X)I\{Y=y_h\}\|_{\mathcal{H}}^2 \\
=\ & CE\left[I\{Y=y_j\}\langle\Sigma(\Sigma+sI)^{-2}K(.,X),K(.,X)\rangle_{\mathcal{H}}^2\right] \\
=\ & CE\left[I\{Y=y_j\}\mathrm{tr}(\Sigma(\Sigma+sI)^{-2}\Sigma(X))\right] \\
=\ & CE\left[\mathrm{tr}(\Sigma(\Sigma+sI)^{-2}\Sigma(X))E[I\{Y=y_j\}|X]\right] \\
\leq\ & C\mathrm{tr}(\Sigma^2(\Sigma+sI)^{-2}) \\
=\ & C\sum_j\frac{\lambda_j^2}{(\lambda_j+s)^2},
\end{aligned}
$$

by Markov inequality,

$$
\|\Sigma^{1/2}(\Sigma+sI)^{-1}(\Gamma'_{nh}-\Gamma_h)\Sigma^{-1/2}\|^2=O_p\left(\sum_j\frac{\lambda_j^2}{n(\lambda_j+s)^2}\right).
$$

Similarly one can show

$$
\|\Sigma^{1/2}(\Sigma+sI)^{-1}(\Gamma_n-\Gamma'_{nh})\Sigma^{-1/2}\|^2=O_p\left(\sum_j\frac{\lambda_j^2}{n(\lambda_j+s)^2}\right).
$$

These imply that

$$
(II)^2=O_p\left(\sum_j\frac{\lambda_j^2}{n(\lambda_j+s)^2}\right).
$$

Once we have shown $\|(\Sigma_n+sI)^{-1}\Gamma_n\Sigma^{-1/2}\|=O_p(1)$, the bounds given above will combine to obtain that (S1.1) is bounded by $O_p\left(s+\sum_j\frac{\lambda_j^2}{n(\lambda_j+s)^2}\right)$ and direct calculations by plugging in $\lambda_j\asymp j^{-d}$ and $s=c_nn^{-d/(d+1)}$ shows that this is $O_p(c_nn^{-d/(d+1)})$.

What is left is to show $\|(\Sigma_n+sI)^{-1}\Gamma_n\Sigma^{-1/2}\|=O_p(1)$, which is equivalent to showing $\|(\Sigma_n+sI)^{-1}\Gamma_n\Sigma^{-1/2}-\Sigma^{-1}\Gamma\Sigma^{-1/2}\|=O_p(1)$. Note this equation is actually similar to (S1.1). Following similar lines that are used to upper bound the terms (I)-(III) in (S1.1), we will get

$$
\begin{aligned}
& \|(\Sigma_n+sI)^{-1}\Gamma_n\Sigma^{-1/2}-\Sigma^{-1}\Gamma\Sigma^{-1/2}\|^2 \\
=\ & O_p\left(\sum_j\frac{\lambda_j}{n(\lambda_j+s)^2}\right)\|(\Sigma_n+sI)^{-1}\Gamma_n\Sigma^{-1/2}\|^2+O_p(1+\sum_j\frac{\lambda_j}{n(\lambda_j+s)^2}).
\end{aligned}
$$

When $s=c_nn^{-d/(d+1)}$ with $c_n\to\infty$, by direct calculations we have $\sum_j\frac{\lambda_j}{n(\lambda_j+s)^2}=o(1)$ and thus the above displayed equation implies $\|(\Sigma_n+sI)^{-1}\Gamma_n\Sigma^{-1/2}\|=O_p(1)$. □