Supplementary Material for "Estimating Density Ridges by Direct Estimation of Density-Derivative-Ratios"

Hiroaki Sasaki[1,4]
hsasaki@is.naist.jp

Takafumi Kanamori[2,4]
kanamori@is.nagoya-u.ac.jp

Masashi Sugiyama[3,4]
sugi@k.u-tokyo.ac.jp

[1] Graduate School of Information Science, Nara Institute of Science and Technology, Nara, Japan
[2] Department of Computer Science and Mathematical Informatics, Nagoya University, Aichi, Japan
[3] Graduate School of Frontier Sciences, The University of Tokyo, Tokyo, Japan
[4] Center for Advanced Intelligence Project, RIKEN, Tokyo, Japan

A Proof of Theorem 1

Let $L(r)$ be the objective function of (9). Suppose that n is sufficiently large and $\|r^*_{k,\boldsymbol{j}}\|_{\mathcal{H}} \leq M_n$ holds. Then, the inequality $L(\widehat{r}_{k,\boldsymbol{j}}) \leq L(r^*_{k,\boldsymbol{j}})$ leads to

$$\frac{1}{2}\|\widehat{r} - r^*_{k,\boldsymbol{j}}\|_P^2 \leq \frac{1}{2}\int\{(\widehat{r})^2 - (r^*_{k,\boldsymbol{j}})^2\}\mathrm{d}(P-P_n) - (-1)^k\int\{\partial^{k,\boldsymbol{j}}\widehat{r} - \partial^{k,\boldsymbol{j}}r^*_{k,\boldsymbol{j}}\}\mathrm{d}(P-P_n), \quad (1)$$

where we used the equality

$$\int (r^*_{k,\boldsymbol{j}})^2\mathrm{d}P = (-1)^k\int \partial^{k,\boldsymbol{j}}r^*_{k,\boldsymbol{j}}\mathrm{d}P$$

that holds under the assumption of $p(\boldsymbol{x})$.

We use the following theorem.

Theorem A (Proposition 4 in Cucker and Smale [2002]). *Let $\mathcal{F} \subset (C^\infty(\mathcal{D}), \|\cdot\|_\infty)$ be a function set defined on a compact set $\mathcal{D}$ endowed with the supremum norm. Suppose that there exists $B > 0$ such that $\|f\|_\infty \leq B$ for any $f \in \mathcal{F}$. Let P_n be the empirical distribution of n i.i.d. samples from P. Then, for all $\varepsilon > 0$*

$$\Pr\left\{\sup_{f\in\mathcal{F}}\left|\int f\,\mathrm{d}(P-P_n)\right| > \varepsilon\right\} \leq 2\mathcal{N}_\infty(\mathcal{F}, \varepsilon/4)\exp\left(-\frac{n\varepsilon^2}{4(2\sigma^2 + B\varepsilon/3)}\right)$$

holds, where $\mathcal{N}_\infty(\mathcal{F}, \varepsilon)$ is the covering number of $\mathcal{F}$ with the radius ε under the supremum norm and $\sigma^2 := \sup_{f\in\mathcal{F}}\mathrm{Var}[f] \leq B^2$.

A.1 Bounds of Covering Entropy

Suppose that $M = M_n$ for a fixed sample size n. Let $\mathcal{H}_M$ be $\mathcal{H}_M = \{r \in \mathcal{H} \,|\, \|r\|_{\mathcal{H}} \leq M\}$. We derive upper bounds for the covering numbers of the function sets $\mathcal{F}_M$ and $\mathcal{G}_M$ defined as

$$\mathcal{F}_M = \{r^2 - (r^*_{k,\boldsymbol{j}})^2 \,|\, r \in \mathcal{H}_M\}$$
$$\mathcal{G}_M = \{\partial^{k,\boldsymbol{j}}r - \partial^{k,\boldsymbol{j}}r^*_{k,\boldsymbol{j}} \,|\, r \in \mathcal{H}_M\}.$$

The supremum norm is bounded above by the RKHS norm, i.e., we have

$$|r(\boldsymbol{x})| \leq \|r\|_{\mathcal{H}}\sqrt{K(\boldsymbol{x},\boldsymbol{x})} \leq Mc_K.$$

Firstly, let us consider $\mathcal{F}_M$. For any $r^2 - (r^*_{k,\boldsymbol{j}})^2 \in \mathcal{F}_M$, the supremum norm is bounded above by $2M^2c_K^2$, since the inequality

$$\begin{aligned}\|r^2 - (r^*_{k,\boldsymbol{j}})^2\|_\infty &\leq \|r^2\|_\infty + \|(r^*_{k,\boldsymbol{j}})^2\|_\infty \\ &= \|r\|_\infty^2 + \|r^*_{k,\boldsymbol{j}}\|_\infty^2 \\ &= 2M^2c_K^2\end{aligned}$$

holds. For $r_1^2 - (r^*_{k,\boldsymbol{j}})^2, r_2^2 - (r^*_{k,\boldsymbol{j}})^2 \in \mathcal{F}_M$, we have

$$\begin{aligned}&\|(r_1^2 - (r^*_{k,\boldsymbol{j}})^2) - (r_2^2 - (r^*_{k,\boldsymbol{j}})^2)\|_\infty \\ &= \sup_{\boldsymbol{x}}|r_1(\boldsymbol{x}) + r_2(\boldsymbol{x})|\cdot|r_1(\boldsymbol{x}) - r_2(\boldsymbol{x})| \\ &\leq 2Mc_K\sup_{\boldsymbol{x}}|r_1(\boldsymbol{x}) - r_2(\boldsymbol{x})| \\ &= 2Mc_K\|r_1 - r_2\|_\infty.\end{aligned}$$

Hence, an upper bound of the covering number of $\mathcal{F}_M$ is given by

$$\ln\mathcal{N}_\infty(\mathcal{F}_M, 2Mc_K\varepsilon) \leq \ln\mathcal{N}_\infty(\mathcal{H}_M, \varepsilon) \leq C\left(\frac{M}{\varepsilon}\right)^{2D/h}$$

for all $h > D$. The second inequality is presented in Cucker and Smale [2002], where C is a constant independent of M and ε. Hence, we have

$$\ln\mathcal{N}_\infty(\mathcal{F}_M, \varepsilon) \leq C\left(\frac{M^2}{\varepsilon}\right)^{2D/h}. \quad (2)$$

Secondly, let us consider $\mathcal{G}_M$. The assumption (10) ensures that

$$\begin{aligned}|\partial^{k,\boldsymbol{j}} r(\boldsymbol{x}) - \partial^{k,\boldsymbol{j}} r^*_{k,\boldsymbol{j}}(\boldsymbol{x})| &= |\partial^{k,\boldsymbol{j}}(r(\boldsymbol{x}) - r^*_{k,\boldsymbol{j}}(\boldsymbol{x}))| \\ &\leq \|r - r^*_{k,\boldsymbol{j}}\|_{\mathcal{H}} \sqrt{\partial^{\boldsymbol{j}} \partial'^{\boldsymbol{j}} K(\boldsymbol{x}, \boldsymbol{x})} \\ &\leq 2Mc_K\end{aligned}$$

holds for any $\partial^{k,\boldsymbol{j}} r - \partial^{k,\boldsymbol{j}} r^*_{k,\boldsymbol{j}} \in \mathcal{G}_M$. The above inequality is shown in Zhou [2008]. In the same way we have

$$|\partial^{k',\boldsymbol{j}'} r(\boldsymbol{x}) - \partial^{k',\boldsymbol{j}'} r^*_{k,\boldsymbol{j}}(\boldsymbol{x})| \leq 2Mc_K$$

for all $\boldsymbol{j}'$ such that $k' = |\boldsymbol{j}'| \leq k + \ell$. This means that $\partial^{k,\boldsymbol{j}}(r - r^*_{k,\boldsymbol{j}})$ is included in the Sobolev space $(W^\ell, \|\cdot\|_\ell)$ endowed with the inner product

$$\int_{\mathcal{D}} \sum_{|\boldsymbol{j}'| \leq \ell} \partial^{k',\boldsymbol{j}'} f(\boldsymbol{x}) \partial^{k',\boldsymbol{j}'} g(\boldsymbol{x}) \mathrm{d}\boldsymbol{x},$$

where $\mathrm{d}\boldsymbol{x}$ denotes the Lebesgue measure on the compact set $\mathcal{D} \subset \mathbb{R}^D$. The Sobolev norm of $\partial^{k,\boldsymbol{j}}(r - r^*_{k,\boldsymbol{j}})$ is bounded above by

$$\begin{aligned}\|\partial^{k,\boldsymbol{j}}(r - r^*_{k,\boldsymbol{j}})\|^2_\ell &= \sum_{\boldsymbol{j}': k \leq |\boldsymbol{j}'| \leq k+\ell} \int_{\mathcal{D}} |\partial^{k',\boldsymbol{j}'}(r - r^*_{k,\boldsymbol{j}})|^2 \mathrm{d}\boldsymbol{x} \\ &\leq 4\mathrm{vol}(\mathcal{D}) {c_K}^2 M^2 \sum_{a=k}^{k+\ell} \binom{D-1+a}{D-1} \\ &\leq 12\mathrm{vol}(\mathcal{D}) {c_K}^2 M^2 (D + k + \ell)^{k+\ell},\end{aligned}$$

where $\mathrm{vol}(\mathcal{D})$ is the volume of $\mathcal{D}$. Hence, $\|\partial^{k,\boldsymbol{j}} r - \partial^{k,\boldsymbol{j}} r^*_{k,\boldsymbol{j}}\|_\ell \leq \sqrt{12\mathrm{vol}(\mathcal{D})} c_K M (D+k+\ell)^{(k+\ell)/2}$ for all $\partial^{k,\boldsymbol{j}} r - \partial^{k,\boldsymbol{j}} r^*_{k,\boldsymbol{j}} \in \mathcal{G}_M$. Let B_R be the ball in W^ℓ with the radius $R > 0$ centered as the origin. Then, $\mathcal{G}_M$ is included in B_R with $R = \sqrt{12\mathrm{vol}(\mathcal{D})} c_K M (D + k + \ell)^{(k+\ell)/2}$ in W^ℓ. Let $I_\ell : W^\ell \to (C(\mathcal{D}), \|\cdot\|_\infty)$ be the embedding map. If $\ell > D/2$ holds, the covering entropy of $\overline{I_\ell(B_R)}$ is bounded above by

$$\ln \mathcal{N}_\infty(\overline{I_\ell(B_R)}, \varepsilon) \leq C(R/\varepsilon)^{D/\ell} + 1$$

as shown in Cucker and Smale [2002], where C is a constant independent of R and ε. Hence, the upper bound of the covering entropy of $\mathcal{G}_M$ is given by

$$\ln \mathcal{N}(\mathcal{G}_M, \varepsilon) \leq C \left(\frac{M}{\varepsilon}\right)^{D/\ell} \tag{3}$$

for $M \geq 1$.

A.2 Uniform Law of Large Numbers

Theorem A is used to derive an upper bound of $\|\widehat{r}_{k,\boldsymbol{j}} - r^*_{k,\boldsymbol{j}}\|^2_P$ in (1). Both $\mathcal{F}_M$ and $\mathcal{G}_M$ satisfy the assumption in Theorem A. Hence, (2) and (3) lead to

$$\begin{aligned}&\sup_{r \in \mathcal{H}_M} \left| \int \{r^2 - (r^*_{k,\boldsymbol{j}})^2\} \mathrm{d}(P - P_n) \right| \\ &\qquad = O_p\left(\frac{M^{2D/(D+h)}}{n^{h/(2h+2D)}}\right), \quad h > D \\ &\sup_{r \in \mathcal{H}_M} \left| \int \{\partial^{k,\boldsymbol{j}} r - \partial^{k,\boldsymbol{j}}(r^*_{k,\boldsymbol{j}})\} \mathrm{d}(P - P_n) \right| \\ &\qquad = O_p\left(\frac{M^{D/(2\ell+D)}}{n^{\ell/(2\ell+D)}}\right).\end{aligned}$$

Note that Theorem A holds even for any fixed sample size n. Suppose that M_n is of the poly-logarithmic order such as $(\log(n))^\gamma$. Then, for sufficiently large $h (> D)$, the upper bound of $\|\widehat{r}_{k,\boldsymbol{j}} - r^*_{k,\boldsymbol{j}}\|^2_P$ is of the order $c_n/n^{1/(2+D/\ell)}$ where c_n is of the poly-logarithmic order of n.

References

F. Cucker and S. Smale. On the mathematical foundations of learning. *Bulletin of the American Mathematical Society*, 39:1–49, 2002.

D. Zhou. Derivative reproducing properties for kernel methods in learning theory. *Journal of Computational and Applied Mathematics*, 220(1-2):456–463, 2008.