



HAL
open science

Supplement to "Dimension reduction for regression over a general class of functions"

Quentin Paris

► **To cite this version:**

Quentin Paris. Supplement to "Dimension reduction for regression over a general class of functions". 2012. hal-00785657v1

HAL Id: hal-00785657

<https://hal.science/hal-00785657v1>

Preprint submitted on 6 Feb 2013 (v1), last revised 20 Mar 2013 (v2)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Supplement to “Dimension reduction for regression over a general class of functions”

Quentin PARIS

IRMAR, ENS Cachan Bretagne, CNRS, UEB
Campus de Ker Lann
Avenue Robert Schuman, 35170 Bruz, France
quentin.paris@bretagne.ens-cachan.fr

1 Proof of Lemma 8.3

Lemma 1.1. *Suppose assumption (A2) holds and $d \geq 3$. Then, for all $\varphi \in \mathcal{F}_d$, we have*

$$\mathbb{E} \|X_{(1)}^\varphi(X^\varphi) - X^\varphi\|^2 \leq \frac{16R^2}{n^{2/d}} \left(\frac{d}{2} - 1\right)^{4/d}.$$

Proof – Fix $\varphi \in \mathcal{F}_d$ and denote μ_φ the distribution of X^φ . Since $\dim S(\varphi) \leq d$, one may consider that X^φ is \mathbb{R}^d -valued and that μ_φ is of support in \mathbb{R}^d . Hence, for all $\varepsilon > 0$, we have

$$\begin{aligned} \mathbb{P} \left(\|X_{(1)}^\varphi(X^\varphi) - X^\varphi\| > \varepsilon \right) &= \mathbb{E} \left[\mathbb{P} \left(\|X_{(1)}^\varphi(X^\varphi) - X^\varphi\| > \varepsilon \mid X \right) \right] \\ &= \mathbb{E} \left[\mathbb{P} \left(\|X_1^\varphi - X^\varphi\| > \varepsilon \mid X \right)^n \right] \\ &= \mathbb{E} \left[\left(1 - \mu_\varphi(B_d(X^\varphi, \varepsilon)) \right)^n \right] \\ &= \int_{\|u\|_d \leq R} \left(1 - \mu_\varphi(B_d(u, \varepsilon)) \right)^n \mu_\varphi(du), \end{aligned}$$

since, under assumption (A3), the variable X^φ is $B_d(0, R)$ -valued. Now, for all $\varepsilon > 0$, the ε -covering number of $B_d(0, R)$ satisfies

$$N(\varepsilon) := N(B_d(0, R), \varepsilon) \leq \left(\frac{4R}{\varepsilon} \right)^d$$

(see *e.g.* Proposition 5 in Cucker and Smale, 2001). Thus, given $\varepsilon > 0$, one may find a finite collection of Euclidian balls $B_1, \dots, B_{N(\varepsilon)}$ of radius at most ε in \mathbb{R}^d such that

$$B_d(0, R) \subset \bigcup_{i=1}^{N(\varepsilon)} B_i.$$

We can notice that for all $i \in \{1, \dots, N(\varepsilon)\}$, we have $u \in B_i \Rightarrow B_i \subset B_d(u, \varepsilon)$. Then

$$\begin{aligned} \mathbb{P}\left(\|X_{(1)}^\varphi(X^\varphi) - X^\varphi\| > \varepsilon\right) &= \int_{\|u\|_d \leq R} (1 - \mu_\varphi(B_d(u, \varepsilon)))^n \mu_\varphi(du) \\ &\leq \sum_{i=1}^{N(\varepsilon)} \int_{B_i} (1 - \mu_\varphi(B_i))^n \mu_\varphi(du) \\ &= \sum_{i=1}^{N(\varepsilon)} \mu_\varphi(B_i) (1 - \mu_\varphi(B_i))^n. \end{aligned}$$

Then, since for all $t \in [0, 1]$ we have $t(1-t)^n \leq \frac{1}{n}$, it follows that for all $\varepsilon > 0$

$$\mathbb{P}\left(\|X_{(1)}^\varphi(X^\varphi) - X^\varphi\| > \varepsilon\right) \leq \frac{N(\varepsilon)}{n} \leq \frac{1}{n} \left(\frac{4R}{\varepsilon}\right)^d.$$

Now write $C_n := \frac{(4R)^d}{n}$. Using the fact that $d \geq 3$, we have

$$\begin{aligned} \mathbb{E}\|X_{(1)}^\varphi(X^\varphi) - X^\varphi\|^2 &= \int_0^{+\infty} \mathbb{P}\left(\|X_{(1)}^\varphi(X^\varphi) - X^\varphi\| > \sqrt{\varepsilon}\right) d\varepsilon \\ &\leq \int_0^{+\infty} \min\left(1, \frac{C_n}{\varepsilon^{d/2}}\right) d\varepsilon \\ &\leq \inf_{\delta > 0} \left(\int_0^\delta d\varepsilon + C_n \int_\delta^{+\infty} \frac{d\varepsilon}{\varepsilon^{d/2}} \right) \\ &= \inf_{\delta > 0} \left(\delta + C_n \left(\frac{d}{2} - 1\right) \delta^{1-d/2} \right) \\ &= C_n^{2/d} \left(\frac{d}{2} - 1\right)^{4/d} \\ &= \frac{16R^2}{n^{2/d}} \left(\frac{d}{2} - 1\right)^{4/d}, \end{aligned}$$

which concludes the proof. \square

Proof of Lemma 8.3 – Fix $k \in \{1, \dots, n\}$, $\rho > 0$ and $\varphi \in \mathbf{F}_d(\rho)$ such that $\|\varphi - \varphi^*\|_\infty \leq \rho$. To ease notations we set

$$\hat{r}_\varphi := \hat{r}_\varphi[k] \quad \text{and} \quad W_i(\varphi, \cdot) := W_i[k](\varphi, \cdot), \quad i = n+1, \dots, 2n.$$

First we have

$$\begin{aligned} \mathbb{E} \left[(r(X) - \hat{r}_\varphi(\varphi(X)))^2 \right] &\leq 2\mathbb{E} \left[(r(X) - r_\varphi(\varphi(X)))^2 \right] \\ &\quad + 2\mathbb{E} \left[(r_\varphi(\varphi(X)) - \hat{r}_\varphi(\varphi(X)))^2 \right]. \end{aligned}$$

Since $r = r_{\varphi^*} \circ \varphi^*$ and $r_{\varphi^*} \in \mathcal{L}$, we have under assumption **(A4)**

$$\begin{aligned} \mathbb{E} \left[(r(X) - r_\varphi(\varphi(X)))^2 \right] &\leq 2\mathbb{E} \left[(r_{\varphi^*}(\varphi^*(X)) - r_{\varphi^*}(\varphi(X)))^2 \right] \\ &\quad + 2\mathbb{E} \left[(r_{\varphi^*}(\varphi(X)) - r_\varphi(\varphi(X)))^2 \right] \\ &\leq 8L^2 \|\varphi - \varphi^*\|_\infty^2 \\ &\leq 8L^2 \rho^2. \end{aligned}$$

Therefore, we deduce that

$$\mathbb{E} \left[(r(X) - \hat{r}_\varphi(\varphi(X)))^2 \right] \leq 8L^2 \rho^2 + 2\mathbb{E} \left[(r_\varphi(\varphi(X)) - \hat{r}_\varphi(\varphi(X)))^2 \right].$$

Next, denoting $\mathcal{S} = \{X_{n+1}, \dots, X_{2n}\}$, we have

$$\begin{aligned} \mathbb{E} \left[(r_\varphi(\varphi(X)) - \hat{r}_\varphi(\varphi(X)))^2 \mid X \right] &= \mathbb{E} \left[\left(r_\varphi(\varphi(X)) - \mathbb{E} \left[\hat{r}_\varphi(\varphi(X)) \mid X, \mathcal{S} \right] \right)^2 \mid X \right] \\ &\quad + \mathbb{E} \left[\left(\hat{r}_\varphi(\varphi(X)) - \mathbb{E} \left[\hat{r}_\varphi(\varphi(X)) \mid X, \mathcal{S} \right] \right)^2 \mid X \right] \\ &=: E_1 + E_2. \end{aligned}$$

Here, we have used the fact that

$$\begin{aligned} &\mathbb{E} \left[\left(r_\varphi(\varphi(X)) - \mathbb{E} \left[\hat{r}_\varphi(\varphi(X)) \mid X, \mathcal{S} \right] \right) \left(\hat{r}_\varphi(\varphi(X)) - \mathbb{E} \left[\hat{r}_\varphi(\varphi(X)) \mid X, \mathcal{S} \right] \right) \mid X \right] \\ &= \mathbb{E} \left[\mathbb{E} \left[\left(r_\varphi(\varphi(X)) - \mathbb{E} \left[\hat{r}_\varphi(\varphi(X)) \mid X, \mathcal{S} \right] \right) \left(\hat{r}_\varphi(\varphi(X)) - \mathbb{E} \left[\hat{r}_\varphi(\varphi(X)) \mid X, \mathcal{S} \right] \right) \mid X, \mathcal{S} \right] \mid X \right] \\ &= \mathbb{E} \left[\left(r_\varphi(\varphi(X)) - \mathbb{E} \left[\hat{r}_\varphi(\varphi(X)) \mid X, \mathcal{S} \right] \right) \mathbb{E} \left[\hat{r}_\varphi(\varphi(X)) - \mathbb{E} \left[\hat{r}_\varphi(\varphi(X)) \mid X, \mathcal{S} \right] \mid X, \mathcal{S} \right] \mid X \right] \\ &= 0. \end{aligned}$$

First, we bound the term E_1 . We have

$$\begin{aligned}
\mathbb{E} \left[\hat{r}_\varphi(\varphi(X)) \middle| X, \mathcal{S} \right] &= \mathbb{E} \left[\sum_{i=n+1}^{2n} W_i(\varphi, \varphi(X)) Y_i \middle| X, \mathcal{S} \right] \\
&= \sum_{i=n+1}^{2n} W_i(\varphi, \varphi(X)) \mathbb{E} [Y_i | X_i] \\
&= \sum_{i=n+1}^{2n} W_i(\varphi, \varphi(X)) r(X_i). \tag{1.1}
\end{aligned}$$

Therefore

$$\begin{aligned}
E_1 &= \mathbb{E} \left[\left(r_\varphi(\varphi(X)) - \mathbb{E} \left[\hat{r}_\varphi(\varphi(X)) \middle| X, \mathcal{S} \right] \right)^2 \middle| X \right] \\
&= \mathbb{E} \left[\left(\sum_{i=n+1}^{2n} W_i(\varphi, \varphi(X)) (r_\varphi(\varphi(X)) - r_{\varphi^*}(\varphi^*(X_i))) \right)^2 \middle| X \right] \\
&\leq 3\mathbb{E} \left[\left(\sum_{i=n+1}^{2n} W_i(\varphi, \varphi(X)) (r_\varphi(\varphi(X)) - r_{\varphi^*}(\varphi(X))) \right)^2 \middle| X \right] \\
&\quad + 3\mathbb{E} \left[\left(\sum_{i=n+1}^{2n} W_i(\varphi, \varphi(X)) (r_{\varphi^*}(\varphi(X)) - r_{\varphi^*}(\varphi^*(X_i))) \right)^2 \middle| X \right] \\
&\quad + 3\mathbb{E} \left[\left(\sum_{i=n+1}^{2n} W_i(\varphi, \varphi(X)) (r_{\varphi^*}(\varphi(X_i)) - r_{\varphi^*}(\varphi^*(X_i))) \right)^2 \middle| X \right] \\
&=: J_1 + J_2 + J_3.
\end{aligned}$$

Since $r = r_{\varphi^*} \circ \varphi^*$ and $r_{\varphi^*} \in \mathcal{L}$, assumption **(A4)** leads to

$$J_i \leq 3L^2 \rho^2,$$

for $i = 1$ and $i = 3$. The fact that r_{φ^*} is L -Lipschitz leads to

$$J_2 \leq 3L^2 \mathbb{E} \left[\left(\frac{1}{k} \sum_{i=1}^k \|X^\varphi - X_{(i)}^\varphi(X^\varphi)\| \right)^2 \middle| X \right].$$

Now let $\tilde{n} = \lfloor n/k \rfloor$. We split the sample $\{X_{n+1}^\varphi, \dots, X_{n+k\tilde{n}}^\varphi\}$ into k subsamples of size \tilde{n} :

$$Z_j = \left\{ X_{n+(j-1)\tilde{n}+1}^\varphi, \dots, X_{n+j\tilde{n}}^\varphi \right\}, \quad j = 1, \dots, k,$$

and denote by $Z_j^{(1)}$ the closest element of Z_j from X^φ . Then

$$\sum_{j=1}^k \|X_{(j)}^\varphi(X^\varphi) - X^\varphi\| \leq \sum_{j=1}^k \|Z_j^{(1)} - X^\varphi\|.$$

Therefore, by Jensen's inequality and lemma 1.1, we have

$$\begin{aligned} \mathbb{E}[E_1] &\leq \frac{3L^2}{k} \sum_{j=1}^k \mathbb{E} \|Z_j^{(1)} - X^\varphi\|^2 + 6L^2 \rho^2 \\ &\leq \frac{48L^2 R^2}{\tilde{n}^{2/d}} \left(\frac{d-2}{2} \right)^{4/d} + 6L^2 \rho^2 \\ &\leq 48L^2 R^2 4^{1/d} \left(\frac{d-2}{2} \right)^{4/d} \left(\frac{k}{n} \right)^{2/d} + 6L^2 \rho^2, \end{aligned}$$

where the last inequality holds provided $\frac{k}{n} \lfloor n/k \rfloor \geq \frac{1}{2}$.

Now we turn to bounding the term E_2 . According to (1.1), we have

$$\begin{aligned} E_2 &= \mathbb{E} \left[\left(\sum_{i=n+1}^{2n} W_i(\varphi, \varphi(X)) (Y_i - r(X_i)) \right)^2 \middle| X \right] \\ &= \mathbb{E} \left[\sum_{i=n+1}^{2n} W_i(\varphi, \varphi(X))^2 (Y_i - r(X_i))^2 \middle| X \right], \end{aligned}$$

where we have used the fact that if $i \neq j \in \{n+1, \dots, 2n\}$

$$\mathbb{E} \left[W_i(\varphi, \varphi(X)) (Y_i - r(X_i)) W_j(\varphi, \varphi(X)) (Y_j - r(X_j)) \middle| X \right] = 0.$$

Using the properties $\sum_{i=n+1}^{2n} W_i(\varphi, \varphi(X)) = 1$, $W_i(\varphi, \varphi(X)) \leq \frac{1}{k}$ and $|Y_i - r(X_i)| \leq B$, for all $i \in \{n+1, \dots, 2n\}$, we conclude that

$$\mathbb{E}[E_2] \leq \frac{B^2}{k}.$$

To complete the proof, write

$$\begin{aligned}\mathbb{E}\left[\left(r(X) - \hat{r}_\varphi(\varphi(X))\right)^2\right] &\leq 2\mathbb{E}[E_1] + 2\mathbb{E}[E_2] + 8L^2\rho^2 \\ &\leq \frac{2B^2}{k} + 96L^2R^24^{1/d} \left(\frac{d-2}{2}\right)^{4/d} \left(\frac{k}{n}\right)^{2/d} + 20L^2\rho^2 \\ &\leq C \left\{ \frac{1}{k} + \left(\frac{k}{n}\right)^{2/d} \right\} + C\rho^2,\end{aligned}$$

where $C := \max\left\{2B^2; 20L^2; 96L^2R^24^{1/d} \left(\frac{d-2}{2}\right)^{4/d}\right\}$. \square

References

Cucker, F. and Smale, S. (2001). On the mathematical foundations of learning theory. *Bulletin of the American Mathematical Society*. Vol. 39, pp. 1-49.