



**HAL**  
open science

## Supplement to "Minimax adaptive dimension reduction for regression"

Quentin Paris

► **To cite this version:**

Quentin Paris. Supplement to "Minimax adaptive dimension reduction for regression". 2012. hal-00768913v1

**HAL Id: hal-00768913**

**<https://hal.science/hal-00768913v1>**

Preprint submitted on 29 Dec 2012 (v1), last revised 28 Jun 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  
**“Minimax adaptive dimension reduction for  
regression”**

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 Presentation

Let  $(X, Y)$  be an  $\mathcal{X} \times \mathbb{R}$ -valued random variable of distribution  $P$  where  $\mathcal{X}$  denotes a metric space. We fix  $T > 0$  and assume that  $|Y| \leq T$ . Let  $\mu$  be the distribution of  $X$ . For  $x \in \mathcal{X}$  we set

$$f_*(x) := \mathbb{E}(Y|X = x). \quad (1.1)$$

Let  $L > 0$  be fixed and  $\mathcal{F}$  be a given class of functions  $f : \mathcal{X} \rightarrow \mathbb{R}$  such that

$$\sup_{x \in \mathcal{X}} |f(x)| \leq L. \quad (1.2)$$

Let  $(X_1, Y_1), \dots, (X_n, Y_n)$  be a sample of  $n$  independent and identically distributed random variables with distribution  $P$ . The least-squares estimate  $f_n$  of  $f_*$  is defined as any random element in  $\mathcal{F}$  satisfying

$$f_n \in \arg \min_{f \in \mathcal{F}} \frac{1}{n} \sum_{i=1}^n (Y_i - f(X_i))^2.$$

Implicitly, it will be assumed that such an element exists. For any Borel probability measure  $\nu$  on  $\mathcal{X}$ , and all  $\rho > 0$ , we denote  $H(\rho, \mathcal{F}, \nu)$  the  $\rho$ -metric entropy of  $\mathcal{F}$  in  $\mathbb{L}^2(\nu)$  (i.e. the logarithm of the minimal number of metric balls with radius  $\rho$  that are needed to cover  $\mathcal{F}$  in  $\mathbb{L}^2(\nu)$ ). We define

$$H(\rho, \mathcal{F}) := \sup_{\nu} H(\rho, \mathcal{F}, \nu), \quad (1.3)$$

where the supremum is taken over all probability measures  $\nu$  with finite support in  $\mathcal{X}$ .

**Theorem 1.1.** *Let  $0 < s < 2$  and assume there exists  $A > 0$  such that for all  $\rho > 0$  we have  $H(\rho, \mathcal{F}) \leq A\rho^{-s}$ . Then, there exists a constant  $C$  depending only on  $s, A$  and  $L$  such that for all  $\varepsilon \in (0, 1]$  we have*

$$\mathbb{E} \|f_n - f_*\|_{\mu}^2 \leq (1 + \varepsilon) \inf_{f \in \mathcal{F}} \|f - f_*\|_{\mu}^2 + C\varepsilon^{-\frac{2-s}{2+s}} \left(\frac{b}{n}\right)^{\frac{2}{2+s}} + \frac{C}{\varepsilon} \left(\frac{b}{n}\right),$$

where  $b := (T + L)^2$  and where  $\|\cdot\|_{\mu}^2 := \int |\cdot|^2 d\mu$ .

## 2 Technical results

The outline of the proof follows Koltchinskii's approach for bounding the excess risk of empirical risk minimizers (for a detailed exposition of that topic, see e.g. Koltchinskii, 2006, or Chapter 4 in Koltchinskii, 2011 and the references cited therein).

Let  $\ell$  be the loss function defined by

$$\ell(u, z) := (z - u)^2.$$

For all  $f \in \mathbb{L}^2(\mu)$ , we denote  $\ell \bullet f$  the function defined on  $\mathcal{X} \times [-T, T]$  by

$$\ell \bullet f(x, z) := \ell(f(x), z).$$

It may be easily verified that

$$\begin{aligned} \|f_n - f_*\|_\mu^2 &= \int (\ell \bullet f_n - \ell \bullet f_*) dP \\ &= \left( \int \ell \bullet f_n dP - \inf_{f \in \mathcal{F}} \int \ell \bullet f dP \right) + \inf_{f \in \mathcal{F}} \|f - f_*\|_\mu^2. \end{aligned} \quad (2.1)$$

The first term in parenthesis on the right side of (2.1) is usually referred to as the estimation error and measures how well the estimate  $f_n$  behaves compared to the best possible approximation of  $f_*$  in  $\mathcal{F}$ . The second term, referred to as the approximation error, measures how far  $f_*$  is from its best possible approximation in  $\mathcal{F}$ . Hence, if for all  $g \in \mathcal{F}$  we denote its excess risk by

$$\mathcal{R}(g) := \int \ell \bullet g dP - \inf_{f \in \mathcal{F}} \int \ell \bullet f dP,$$

we obtain the so called bias-variance decomposition

$$\mathbb{E} \|f_n - f_*\|_\mu^2 = \mathbb{E} \mathcal{R}(f_n) + \inf_{f \in \mathcal{F}} \|f - f_*\|_\mu^2. \quad (2.2)$$

Since  $\mathcal{R}(f_n)$  is a positive random variable, we have

$$\mathbb{E} \mathcal{R}(f_n) = \int_0^{+\infty} \mathbb{P}(\mathcal{R}(f_n) > u) du,$$

so that finally, one may reduce the problem of finding a bound on the mean squared error of  $f_n$  to that of deriving bounds for the probabilities

$$\mathbb{P}(\mathcal{R}(f_n) > u), \quad u > 0.$$

To achieve this task, we follow the lines devised in Koltchinskii (2006). For all  $\delta > 0$ , let  $\mathcal{F}(\delta)$  be the  $\delta$ -minimal set of the excess risk defined by

$$\mathcal{F}(\delta) := \left\{ f \in \mathcal{F} : \mathcal{R}(f) \leq \delta \right\}.$$

For any class  $\mathcal{T}$  of function defined on  $\mathcal{X} \times \mathbb{R}$ , denote

$$\|P_n - P\|_{\mathcal{T}} := \sup_{\mathbf{t} \in \mathcal{T}} \left| \int \mathbf{t} d(P_n - P) \right|,$$

where  $P_n := n^{-1} \sum_{i=1}^n \delta_{(X_i, Y_i)}$  stands for the empirical distribution of the sample  $(X_1, Y_1), \dots, (X_n, Y_n)$ . For all  $\delta > 0$ , set finally

$$\mathcal{L}(\delta) := \left\{ (\ell \bullet f - \ell \bullet g) : f, g \in \mathcal{F}(\delta) \right\}.$$

Then, it may be easily verified that  $\hat{\delta} := \mathcal{R}(f_n)$  satisfies

$$\hat{\delta} \leq \|P_n - P\|_{\mathcal{L}(\hat{\delta})}.$$

As a result, given an upper bound  $U_n(\delta)$  for  $\|P_n - P\|_{\mathcal{L}(\delta)}$  which holds with high probability uniformly over all  $\delta > 0$ , we can bound the excess risk  $\mathcal{R}(f_n)$  with high probability by the largest  $\delta > 0$  satisfying  $\delta \leq U_n(\delta)$ . A powerful means to derive such bounds for  $\|P_n - P\|_{\mathcal{L}(\delta)}$  is Talagrand's concentration inequality (see Talagrand, 1996). Bousquet's inequality is an important improvement of Talagrand's concentration inequality where constants are explicit (see Bousquet, 2002). Bousquet's concentration inequality states that for all  $b > 0$  and for any class  $\mathcal{T}$  of  $[-b, b]$ -valued functions, we have for all  $t > 0$

$$\mathbb{P} \left( \|P_n - P\|_{\mathcal{T}} \geq \mathbb{E} \|P_n - P\|_{\mathcal{T}} + \sqrt{\frac{2t}{n} (\sigma^2(\mathcal{T}) + 4b \mathbb{E} \|P_n - P\|_{\mathcal{T}})} + \frac{2bt}{3n} \right) \leq e^{-t},$$

where we have denoted

$$\sigma^2(\mathcal{T}) := \sup_{\mathbf{t} \in \mathcal{T}} \left( \int \mathbf{t}^2 dP - \left( \int \mathbf{t} dP \right)^2 \right).$$

An application of Bousquet's inequality in our context allows to derive the following result.

**Lemma 2.1.** For all  $\delta > 0$  and for all  $t > 0$  we have

$$\mathbb{P} \left( \|P_n - P\|_{\mathcal{L}(\delta)} \geq 2\mathbb{E}\|P_n - P\|_{\mathcal{L}(\delta)} + 4\sqrt{b(\delta + \Delta)\frac{2t}{n} + \frac{8bt}{3n}} \right) \leq e^{-t},$$

where  $b := (T + L)^2$  and  $\Delta := \inf_{f \in \mathcal{F}} \|f - f_*\|_{\mu}^2$ .

**Proof** – For all  $f \in \mathcal{F}$  we have  $0 \leq \ell \bullet f \leq b$ . Therefore,  $\mathcal{L}(\delta)$  is a class of  $[-b, b]$ -valued functions. Denoting  $V := \|P_n - P\|_{\mathcal{L}(\delta)}$  and using the inequalities  $\sqrt{a+b} \leq \sqrt{a} + \sqrt{b}$  and  $2\sqrt{uv} \leq u + v$ , for  $u, v \geq 0$ , we obtain that

$$\mathbb{E}V + \sqrt{\frac{2t}{n} (\sigma^2(\mathcal{L}(\delta)) + 4b\mathbb{E}V)} + \frac{2bt}{3n} \leq 2\mathbb{E}V + \sigma(\mathcal{L}(\delta))\sqrt{\frac{2t}{n} + \frac{8bt}{3n}}.$$

As a result, we deduce from Bousquet's inequality that

$$\mathbb{P} \left( V \geq 2\mathbb{E}V + \sigma(\mathcal{L}(\delta))\sqrt{\frac{2t}{n} + \frac{8bt}{3n}} \right) \leq e^{-t}.$$

It remains only to prove that  $\sigma(\mathcal{L}(\delta)) \leq 4\sqrt{b(\delta + \Delta)}$ . To that aim, observe that for any  $f \in \mathcal{F}$  we have

$$\begin{aligned} \int (\ell \bullet f - \ell \bullet f_*)^2 dP &= \mathbb{E} \left( (Y - f(X))^2 - (Y - f_*(X))^2 \right)^2 \\ &= \mathbb{E} (2Y - f(X) - f_*(X))^2 (f(X) - f_*(X))^2 \\ &\leq 4b \mathbb{E} (f(X) - f_*(X))^2 \\ &= 4b \int (\ell \bullet f - \ell \bullet f_*) dP \\ &= 4b(\mathcal{R}(f) + \Delta). \end{aligned}$$

Then, for all  $f, g \in \mathcal{F}(\delta)$ , we have

$$\begin{aligned} \sqrt{\int (\ell \bullet f - \ell \bullet g)^2 dP} &\leq \sqrt{\int (\ell \bullet f - \ell \bullet f_*)^2 dP} + \sqrt{\int (\ell \bullet g - \ell \bullet f_*)^2 dP} \\ &\leq 4\sqrt{b(\delta + \Delta)}, \end{aligned}$$

Finally, we have proved that

$$\sigma(\mathcal{L}(\delta)) \leq \sup_{t \in \mathcal{L}(\delta)} \sqrt{\int t^2 dP} \leq 4\sqrt{b(\delta + \Delta)}.$$

The proof is complete.  $\square$

**Lemma 2.2.** *Let  $0 < s < 2$  and assume there exists  $A > 0$  such that for all  $\rho > 0$  we have  $H(\rho, \mathcal{F}) \leq A\rho^{-s}$ . Then, there exists a constant  $C$  depending only on  $s, A$  and  $L$  such that for all  $\delta > 0$  we have*

$$\mathbb{E}\|P_n - P\|_{\mathcal{L}(\delta)} \leq C \left\{ \frac{\sqrt{b}(\delta + \Delta)^{(2-s)/4}}{\sqrt{n}} \vee \frac{(\delta + \Delta)^{-s/2}}{n} \right\},$$

where  $b := (T + L)^2$  and  $\Delta := \inf_{f \in \mathcal{F}} \|f - f_*\|_{\mu}^2$ .

**Proof** – For simplicity, we assume that there exists  $\bar{f} \in \mathcal{F}$  such that

$$\int \ell \bullet \bar{f} \, dP = \inf_{f \in \mathcal{F}} \int \ell \bullet f \, dP.$$

From the proof of Lemma 2.1, it results that for all  $\delta > 0$  and for all functions  $f, g \in \mathcal{F}(\delta)$  we have

$$\int (\ell \bullet f - \ell \bullet g)^2 \, dP \leq 16b(\delta + \Delta).$$

Therefore, for all  $\delta > 0$ , we have

$$\begin{aligned} \mathbb{E}\|P_n - P\|_{\mathcal{L}(\delta)} &\leq 2\mathbb{E} \sup \left\{ \left| \int (\ell \bullet f - \ell \bullet \bar{f}) \, d(P_n - P) \right| : f \in \mathcal{F}(\delta) \right\} \\ &\leq 2\theta_n(16b(\delta + \Delta)), \end{aligned} \quad (2.3)$$

where we have denoted

$$\theta_n(\delta) := \mathbb{E} \sup \left\{ \left| \int (\ell \bullet f - \ell \bullet \bar{f}) \, d(P_n - P) \right| : f \in \mathcal{F}, \int (\ell \bullet f - \ell \bullet \bar{f})^2 \, dP \leq \delta \right\}.$$

Let  $\varepsilon_1, \dots, \varepsilon_n$  be a sequence of independent Rademacher random variables (i.e.  $\mathbb{P}(\varepsilon_i = 1) = \mathbb{P}(\varepsilon_i = -1) = 1/2$ ) independent from our sample. Then, according to the symmetrization inequality (see e.g. Theorem 2.1 in Koltchinskii, 2011) we have

$$\theta_n(\delta) \leq 2\mathbb{E} \sup_{\mathbf{t} \in \mathcal{L}(\delta)} \left| \int \mathbf{t} \, dP_n^\circ \right|, \quad (2.4)$$

with

$$\mathcal{L}(\delta) := \left\{ (\ell \bullet f - \ell \bullet \bar{f}) : f \in \mathcal{F}, \int (\ell \bullet f - \ell \bullet \bar{f})^2 \, dP \leq \delta \right\},$$

and

$$\int \mathbf{t} \, dP_n^\circ := \frac{1}{n} \sum_{i=1}^n \varepsilon_i \mathbf{t}(X_i, Z_i).$$

Since for all functions  $f, g \in \mathcal{F}$  we have

$$\int (\ell \bullet f - \ell \bullet g)^2 \, dP \leq 4b \|f - g\|_\mu^2,$$

we deduce that for all  $u > 0$  and for all  $\delta > 0$  we have

$$\begin{aligned} H(u, \mathcal{L}(\delta)) &= H\left(u, \left\{ \ell \bullet f : \int (\ell \bullet f - \ell \bullet \bar{f})^2 \, dP \leq \delta \right\}\right) \\ &\leq H\left(u, \left\{ \ell \bullet f : f \in \mathcal{F} \right\}\right) \\ &\leq H\left(\frac{u}{2\sqrt{b}}, \mathcal{F}\right) \\ &\leq A \left(\frac{2\sqrt{b}}{u}\right)^s. \end{aligned}$$

Hence, Theorem 3.12 in Koltchinskii (2011) yields that we have

$$\mathbb{E} \sup_{\mathbf{t} \in \mathcal{L}(\delta)} \left| \int \mathbf{t} \, dP_n^\circ \right| \leq K \left\{ \frac{b^{s/4} \delta^{(2-s)/4}}{\sqrt{n}} \vee \frac{b^{s/2} \delta^{-s/2}}{n} \right\}, \quad (2.5)$$

where  $K > 0$  denotes a constant depending only on  $s, A$  and  $L$ . Combining (2.3), (2.4) and (2.5) we deduce that for all  $\delta > 0$  we have

$$\mathbb{E} \|P_n - P\|_{\mathcal{L}(\delta)} \leq C \left\{ \frac{\sqrt{b}(\delta + \Delta)^{(2-s)/4}}{\sqrt{n}} \vee \frac{(\delta + \Delta)^{-s/2}}{n} \right\},$$

where  $C > 0$  denotes a constant depending only on  $s, A$  and  $L$ . This concludes the proof.  $\square$

**Lemma 2.3.** *Let  $V(u), u \geq 0$  be non-negative random variables such that  $V(u) \leq V(u')$  if  $u \leq u'$  and  $B(u, t), u \geq 0, t \geq 0$  be real numbers such that*

$$\mathbb{P}(V(u) \geq B(u, t)) \leq e^{-t}.$$

*Let  $\hat{u}$  be a non-negative random variable a priori upper bounded by a constant  $\bar{u} > 0$  and such that  $\hat{u} \leq V(\hat{u})$ . Then for all  $t \geq 0$  we have*

$$\mathbb{P}\left(\hat{u} \geq \inf \left\{ \sigma > 0 : \sup_{u \geq \sigma} \frac{B(u, \frac{tu}{\sigma})}{u} \leq 1 \right\}\right) \leq 3e^{-t}.$$



**Proof** – The bound is obvious if  $t \leq 1$ . Therefore, we will assume  $t \geq 1$ . The proof will be divided in two steps.

*Step 1.* Let  $u_j, j \geq 0$  be a decreasing sequence of positive numbers with  $u_0 = \bar{u}$  and let  $t_j, j \geq 0$  be a sequence of positive numbers. For all  $u \geq 0$ , denote

$$\bar{B}(u) := \sum_{j=0}^{+\infty} B(u_j, t_j) \mathbf{1}\{u_{j+1} < u \leq u_j\},$$

and set

$$u^* := \sup\{u \geq 0 : u \leq \bar{B}(u)\}.$$

The goal of this first step is to prove that

$$\forall u \geq u^* : \mathbb{P}(\hat{u} \geq u) \leq \sum_{u_j \geq u} e^{-t_j}.$$

Fix  $u > u^*$ . If for all  $j \geq 0$ , we denote  $E_j := \{V(u_j) \leq \bar{B}(u_j)\}$  and

$$E := \bigcap_{u_j \geq u} E_j,$$

it may be easily verified that

$$\mathbb{P}(E) \geq 1 - \sum_{u_j \geq u} e^{-t_j}.$$

On the event  $E$ , for all  $u' \geq u$  we have  $V(u') \leq \bar{B}(u')$  by monotonicity of  $V$  and by definition of  $\bar{B}$ . Thus, on the event  $\{\hat{u} \geq u\} \cap E$  we obtain

$$\hat{u} \leq V(\hat{u}) \leq \bar{B}(\hat{u}),$$

which implies that  $u \leq \hat{u} \leq u^*$ . Since this contradicts  $u > u^*$ , we deduce that  $\{\hat{u} \geq u\} \subset E^c$  which implies that

$$\mathbb{P}(\hat{u} \geq u) \leq \sum_{u_j \geq u} e^{-t_j}.$$

By continuity, this also holds for  $u = u^*$ .

Step 2. Denote

$$\sigma_t := \inf \left\{ \sigma > 0 : \sup_{u \geq \sigma} \frac{B(u, \frac{tu}{\sigma})}{u} \leq 1 \right\},$$

and fix  $\sigma > \sigma_t$ . Then, for all  $u \geq 0$ , let

$$\bar{B}_\sigma(u, t) := \sum_{j=0}^{+\infty} B\left(\frac{\bar{u}}{2^j}, \frac{t\bar{u}}{\sigma 2^j}\right) \mathbf{1}_{\left\{ \frac{\bar{u}}{2^{j+1}} < u \leq \frac{\bar{u}}{2^j} \right\}}.$$

It may be easily verified that

$$\frac{\bar{B}_\sigma(\sigma, t)}{\sigma} \leq \sup_{u \geq \sigma} \frac{B(u, \frac{tu}{\sigma})}{u} \leq 1,$$

which implies that

$$\sigma \geq u_t^* := \sup \{u \geq 0 : u \leq \bar{B}_\sigma(u, t)\}.$$

Then, according to step 1, we deduce that

$$\mathbb{P}(\hat{u} \geq \sigma) \leq \sum_{\frac{\bar{u}}{2^j} \geq \sigma} e^{-\frac{t\bar{u}}{\sigma 2^j}}.$$

The sum on the right hand side may be bounded as follows. Let

$$j_* := \max \left\{ j \geq 0 : \frac{\bar{u}}{2^j} \geq \sigma \right\}.$$

Then

$$\sum_{\frac{\bar{u}}{2^j} \geq \sigma} e^{-\frac{t\bar{u}}{\sigma 2^j}} = \sum_{j=0}^{j_*} e^{-\frac{t\bar{u}}{\sigma 2^j}} \leq \sum_{j=0}^{+\infty} e^{-t2^j}$$

and

$$\sum_{j=0}^{+\infty} e^{-t2^j} \leq e^{-t} + \sum_{j=1}^{+\infty} (2^j - 2^{j-1}) e^{-t2^j} \leq e^{-t} + \int_1^{+\infty} e^{-tu} du \leq 2e^{-t}.$$

Finally, we have proved that for all  $t \geq 0$  and for all  $\sigma > \sigma_t$  we have

$$\mathbb{P}(\hat{u} \geq \sigma) \leq 3e^{-t}.$$

The result follows by continuity.  $\square$

**Lemma 2.4.** *Let  $0 < s < 2$  and assume there exists  $A > 0$  such that for all  $\rho > 0$  we have  $H(\rho, \mathcal{F}) \leq A\rho^{-s}$ . Then, there exists a constant  $C > 0$  depending only on  $s, A$  and  $L$  such that for all  $t \geq 0$  and for all  $\varepsilon \in (0, 1]$  we have*

$$\mathbb{P} \left( \mathcal{R}(f_n) \geq \varepsilon \inf_{f \in \mathcal{F}} \|f - f_*\|_{\mu}^2 + \left\{ C\varepsilon^{-\frac{2-s}{2+s}} \left( \frac{b}{n} \right)^{\frac{2}{2+s}} \vee \frac{Ct}{\varepsilon} \left( \frac{b}{n} \right) \right\} \right) \leq 3e^{-t}, \quad (2.6)$$

where  $b := (T + L)^2$ .

**Proof** – Let  $t > 0$  and  $\varepsilon \in (0, 1]$  be fixed. In the proof,  $C > 0$  will denote a constant depending only on  $s, A$  and  $L$  which value may change from line to line. According to Lemmas 2.1, 2.2 and 2.3 we have that

$$\mathbb{P}(\mathcal{R}(f_n) \geq \sigma_{n,t}) \leq 3e^{-t},$$

where,

$$\sigma_{n,t} := \inf \left\{ \sigma > 0 : \sup_{\delta \geq \sigma} \frac{B_n \left( \delta, \frac{t\delta}{\sigma} \right)}{\delta} \leq 1 \right\},$$

with

$$B_n \left( \delta, \frac{t\delta}{\sigma} \right) := C \left\{ \frac{\sqrt{b}(\delta + \Delta)^{(2-s)/4}}{\sqrt{n}} \vee \frac{(\delta + \Delta)^{-s/2}}{n} \right\} + C \left\{ \sqrt{b(\delta + \Delta) \frac{t\delta}{n\sigma}} + \frac{bt\delta}{n\sigma} \right\}.$$

If  $\sigma_{n,t} \leq \varepsilon\Delta$ , then the bound is obvious. Therefore, we need only to prove the result for  $\sigma_{n,t} > \varepsilon\Delta$ . Fix  $\sigma > \sigma_{n,t} > \varepsilon\Delta$ . Then for all  $\delta \geq \sigma$  we have

$$\begin{aligned} & \frac{\sqrt{b}(\delta + \Delta)^{(2-s)/4}}{\sqrt{n}} \vee \frac{(\delta + \Delta)^{-s/2}}{n} \\ &= \frac{\sqrt{b}\delta^{(2-s)/4} \left(1 + \frac{\Delta}{\delta}\right)^{(2-s)/4}}{\sqrt{n}} \vee \frac{\delta^{-s/2} \left(1 + \frac{\Delta}{\delta}\right)^{-s/2}}{n} \\ &\leq \frac{\sqrt{b}\delta^{(2-s)/4} \left(1 + \frac{1}{\varepsilon}\right)^{(2-s)/4}}{\sqrt{n}} \vee \frac{\delta^{-s/2}}{n} \\ &\leq 2\sqrt{\frac{b}{n}} \left(\frac{\delta}{\varepsilon}\right)^{(2-s)/4} \vee \frac{\delta^{-s/2}}{n} \end{aligned}$$

Also, for all  $\delta \geq \sigma$ , we have

$$\sqrt{b(\delta + \Delta) \frac{t\delta}{n\sigma}} \leq \sqrt{\frac{2bt\delta^2}{\varepsilon n\sigma}}.$$

Hence, we obtain that

$$\begin{aligned} \sup_{\delta \geq \sigma} \frac{B_n\left(\delta, \frac{t\delta}{\sigma}\right)}{\delta} &\leq C \left\{ \frac{\sqrt{b}}{\varepsilon^{(2-s)/4} \sigma^{(2+s)/4} \sqrt{n}} \vee \frac{1}{n\sigma^{(2+s)/2}} \right\} + C \left\{ \sqrt{\frac{bt}{\varepsilon n \sigma}} + \frac{bt}{n\sigma} \right\} \\ &=: h_1(\sigma) + h_2(\sigma). \end{aligned}$$

We deduce by monotonicity that

$$\sigma_{n,t} \leq \inf \{ \sigma > 0 : h_1(\sigma) + h_2(\sigma) \leq 1 \}.$$

For any two non-increasing functions  $h_1, h_2 : \mathbb{R}_+ \rightarrow \mathbb{R}_+$  it may be easily verified that

$$\inf \{ \sigma : h_1(\sigma) + h_2(\sigma) \leq 1 \} \leq \inf \{ \sigma : h_1(\sigma) \leq \frac{1}{2} \} \vee \inf \{ \sigma : h_2(\sigma) \leq \frac{1}{2} \}.$$

In our case we have

$$\inf \{ \sigma : h_1(\sigma) \leq \frac{1}{2} \} \leq \frac{C}{\varepsilon^{(2-s)/(2+s)}} \left( \frac{b}{n} \right)^{2/(2+s)},$$

and

$$\inf \{ \sigma : h_2(\sigma) \leq \frac{1}{2} \} \leq \frac{Cbt}{\varepsilon n}.$$

Finally, for some constant  $C$  depending only on  $s, A$  and  $L$  we obtain

$$\sigma_{n,t} \leq C \left\{ \frac{1}{\varepsilon^{(2-s)/(2+s)}} \left( \frac{b}{n} \right)^{2/(2+s)} \vee \frac{bt}{\varepsilon n} \right\}.$$

The proof is complete.  $\square$

### Proof of Theorem 1.1

Let  $\varepsilon \in (0, 1]$ . According to the bias-variance decomposition (2.2), we need only to prove that for a constant  $C > 0$  depending only on  $s, A$  and  $L$  we have

$$\mathbb{E} \mathcal{R}(f_n) \leq \varepsilon \inf_{f \in \mathcal{F}} \|f - f_*\|_{\mu}^2 + C\varepsilon^{-\frac{2-s}{2+s}} \left( \frac{b}{n} \right)^{\frac{2}{2+s}} + \frac{C}{\varepsilon} \left( \frac{b}{n} \right).$$

To that aim, note that (2.6) implies that there exists a constant  $C > 0$  depending only on  $s, A$  and  $L$  such that for all  $t > 0$  we have

$$\mathbb{P} \left( \mathcal{R}(f_n) \geq \varepsilon \inf_{f \in \mathcal{F}} \|f - f_*\|_{\mu}^2 + C\varepsilon^{-\frac{2-s}{2+s}} \left(\frac{b}{n}\right)^{\frac{2}{2+s}} + \frac{Ct}{\varepsilon} \left(\frac{b}{n}\right) \right) \leq 3e^{-t}.$$

Therefore, for all  $u > 0$ , we obtain that

$$\mathbb{P}(\mathcal{R}(f_n) \geq u) \leq 3 \exp \left( -\frac{n\varepsilon}{Cb} \left\{ u - \varepsilon \inf_{f \in \mathcal{F}} \|f - f_*\|_{\mu}^2 - C\varepsilon^{-\frac{2-s}{2+s}} \left(\frac{b}{n}\right)^{\frac{2}{2+s}} \right\} \right).$$

Using the last inequality and bounding by 1 the probability  $\mathbb{P}(\mathcal{R}(f_n) \geq u)$  for all  $u > 0$  satisfying

$$u \leq A(\varepsilon, n) := \varepsilon \inf_{f \in \mathcal{F}} \|f - f_*\|_{\mu}^2 + C\varepsilon^{-\frac{2-s}{2+s}} \left(\frac{b}{n}\right)^{\frac{2}{2+s}},$$

we deduce that

$$\begin{aligned} \mathbb{E}\mathcal{R}(f_n) &= \int_0^{+\infty} \mathbb{P}(\mathcal{R}(f_n) \geq u) \, du \\ &= \int_0^{A(\varepsilon, n)} \mathbb{P}(\mathcal{R}(f_n) \geq u) \, du + \int_{A(\varepsilon, n)}^{+\infty} \mathbb{P}(\mathcal{R}(f_n) \geq u) \, du \\ &\leq A(\varepsilon, n) + \int_{A(\varepsilon, n)}^{+\infty} \mathbb{P}(\mathcal{R}(f_n) \geq u) \, du \\ &\leq A(\varepsilon, n) + \int_{A(\varepsilon, n)}^{+\infty} 3 \exp \left( -\frac{n\varepsilon}{Cb} \{u - A(\varepsilon, n)\} \right) \, du \\ &= A(\varepsilon, n) + \int_0^{+\infty} 3 \exp \left( -\frac{n\varepsilon u}{Cb} \right) \, du \\ &= \varepsilon \inf_{f \in \mathcal{F}} \|f - f_*\|_{\mu}^2 + C\varepsilon^{-\frac{2-s}{2+s}} \left(\frac{b}{n}\right)^{\frac{2}{2+s}} + \frac{3Cb}{n\varepsilon}. \end{aligned}$$

This concludes the proof.  $\square$

## References

- Bousquet, O. (2002). A Bennett concentration inequality and its application to suprema of empirical processes. *Comptes Rendus de l'Académie des Sciences de Paris*. Vol. 334, pp. 495-500.
- Koltchinskii, V. (2006). Local Rademacher complexities and oracle inequalities in risk minimization. *The Annals of Statistics*. Vol. 34, pp. 2593-2656.
- Koltchinskii, V. (2011). *Oracle Inequalities in Empirical Risk Minimization and Sparse Recovery Problems*. Ecole d'été de Probabilités de Saint-Flour 2008. Lectures Notes in Mathematics 2033.
- Talagrand, M. (1996). New concentration inequalities in product spaces. *Inventiones Mathematicae*. Vol. 126, pp. 505-563.