AISM (2007) 59: 435–464
DOI 10.1007/s10463-006-0063-7

D. Fourdrinier · S. Pergamenshchikov

Improved model selection method for a regression function with dependent noise

Received: 9 July 2004 / Revised: 21 October 2005 /
Published online: 17 June 2006

Abstract This paper is devoted to nonparametric estimation, through the $\mathcal{L}_2$-risk, of a regression function based on observations with spherically symmetric errors, which are dependent random variables (except in the normal case). We apply a model selection approach using improved estimates. In a nonasymptotic setting, an upper bound for the risk is obtained (oracle inequality). Moreover asymptotic properties are given, such as upper and lower bounds for the risk, which provide optimal rate of convergence for penalized estimators.

Keywords Model selection · Nonparametric estimation · Spherically symmetric distribution · Spherically symmetric regression model

1 Introduction

The paper deals with nonparametric estimation of a regression function under observations with dependent errors. More specifically, we consider the model

$$Y = S + \xi \tag{1}$$

with

$$Y = (Y_1, \ldots, Y_n)', \quad S = (S(x_1), \ldots, S(x_n))' \quad \text{and} \quad \xi = (\xi_1, \ldots, \xi_n)',$$

where $S : [0, 1] \to \mathbb{R}$ is some unknown function, $\{x_1, \ldots, x_n\}$ is the partition of the interval $[0, 1]$, such that $x_i = i/n$, and ξ is a vector of dependent errors whose

The second author is supported by the RFFI grant 04-01-00855

D. Fourdrinier (✉) · S. Pergamenshchikov
Laboratoire de Mathématiques Raphaël Salem
Université de Rouen, UMR CNRS 6085
Avenue de l'Université BP. 12 76801 Saint-Étienne-du-Rouvray, France
E-mail: Dominique.Fourdrinier@univ-rouen.fr
E-mail: Serge.Pergamenchtchikov@univ-rouen.fr

distribution will be specified below (here the notation $'$ holds for transposition). The aim is to estimate the function S based on the vector of observations Y.

Numerous papers are devoted to this problem with i.i.d. errors $\xi_1, \ldots, \xi_n$. See, for example, Donoho et al. (1995), Efroimovich (1999) and Nemirovski (2000).In this context, various properties have been studied such as, asymptotic minimaxity, adaptive estimation etc. In fact, in practical regression problems, the observations are more often dependent. A broad statistical review of long-range dependence has been given by Beran (1992) with discussions and references. In particular, Dahlhaus (1995) studies this problem for long-range dependent linear models. In a nonparametric setting, Hall and Hart (1990) and Csörgo and Mielniczuk (1995) have investigated the asymptotic properties of the kernel estimate.

Fourdrinier and Wells (1994) considered the model (1) for linear function with dependent observations, more specifically under spherically symmetric distributions. They address the selection problem of significant predictors of the linear function through a decision theory approach based on quadratic loss estimation.

In this paper, we consider the spherically symmetric model (1) proposed in Fourdrinier and Strawderman (1994) and Fourdrinier and Wells (1994) in a parametric estimation context. More specifically, we assume that ξ has a density of the form $g(\|\cdot\|^2)$ for some nonegative function g. This class of densities gives rise to a natural extension of normal distributions which lead to dependent components ξ_i. Actually the only case where the (ξ_i) are independent is when $g(t) \sim \mathrm{e}^{-t/2\sigma^2}$, that is, when the distribution of ξ is normal $\mathcal{N}(0, \sigma^2 I)$ where I is the identity matrix.

General distributions of current use belong to this class. This is the case with the multivariate Student's distribution with ν degrees of freedom for which $g(t) \sim (1 + t/\nu)^{-\frac{\nu+n}{2}}$. It is worth noting that this distribution can be viewed as a variance mixture of normal distributions, that is,

$$g(t) = \mathbf{E}\, \frac{1}{(2\pi\, \vartheta)^{\frac{n}{2}}}\, \mathrm{e}^{-t/2\vartheta}\,, \tag{2}$$

where the mixing random variable ϑ has the inverse gamma distribution $\mathcal{IG}(\frac{\nu}{2}\,,\ \frac{\nu}{2})$ (see Kariya and Sinha (1993)). Another example (which is not a variance mixture of normal distributions) is the Kotz distribution with

$$g(t) = \frac{\Gamma(n/2)}{\pi^{n/2}(2\sigma^2)^{n/2+q}\Gamma(n/2+q)}\, t^q\, \mathrm{e}^{-t/2\sigma^2}\,, \tag{3}$$

for some real number q such as $q > -n/2$ (see Fang et al. 1989). Introduced by Kotz (1975), this distribution was found to be useful in constructing models in which the usual normality assumption is not applicable. Note that the case $q = 0$ corresponds to the normal case mentioned above.

As a last example, quoted by Berger (1975) (who gives an extensive class of spherically symmetric distributions), the function

$$g(t) \sim \frac{\mathrm{e}^{-(\alpha t+\beta)}}{(1+\mathrm{e}^{-(\alpha t+\beta)})^2} \quad \alpha > 0\,, \quad \beta \in \mathbb{R} \tag{4}$$

leads to a logistic type distribution.

Under this spherical distributional context, our main goal is to tackle the nonparametric estimation problem of S through the $\mathcal{L}_2$-norm in a non-asymptotic setting. We make use of *the model selection method* (see, for example, Gunst and Mason 1980) developed by Barron et al. (1999) in the non-parametric case. In this last paper, the authors construct the model selection procedure in the particular case where the model (1) is gaussian. They give a non-asymptotic upper bound for the quadratic risk (oracle inequality), i.e. the main term in this bound is the infimum, under all the models, of the sum of the approximate term and the penalization term.

Note that the model selection procedure proposed by Barron, Birge and Massart (1999) is based on the least squares estimator. It is well known, since the founding of Stein (1956), that the least squares estimator is not optimal in the entire class of the estimators. Indeed, in the gaussian case, James and Stein (1961) constructed improved estimators with quadratic risk less than the least squares estimator. Fourdrinier and Strawderman (1996) compared these various improved estimators for the general linear spherically symmetric regression model.

In this paper, we construct a general model selection procedure based on any general parametric estimator (non necessarily the least square estimator) under spherically symmetric distributions in the model (1), for which we obtain an oracle inequality (Theorem 1). Then we adapt the improved estimates from Fourdrinier and Strawderman (1996) and Fourdrinier and Wells (1994) to this model (Theorems 2, 3) and we construct the model selection procedure applying the improved estimates. For this procedure, we specify an oracle inequality as well (Theorem 4).

Moreover, we also study asymptotic properties of the model selection procedures. We find asymptotic upper and lower bounds for the minimax usual risk and the adaptive risk which provide the optimal convergence rate of the proposed procedures.

The paper is organised as follows. In Sect. 2, we describe the model selection method for which we obtain a non-asymptotical upper bound for the $\mathcal{L}_2$-risk. In Sect. 3, we illustrate the model in using improved estimation method. Sect. 4 yields the upper bound for the asymptotical minimax risk of the penalized estimator in case of known smoothness and Sect. 5 is devoted to a lower bound for the $\mathcal{L}_2$-risk. In Sect. 6, we consider the adaptive estimation problem for the model (1). In Sect. 7, we give some general conclusions. Finally, Sect. 8 is an appendix which contains some technical results.

2 Penalization method

We consider the non-parametric regression estimation problem for the model (1). We suppose that the vector ξ has a spherically symmetric distribution with density $g(\|\cdot\|^2)$ in $\mathbf{R}^n$, where $\|\cdot\|$ is the euclidean norm and g is some positive function. Futhermore we assume that this distribution satisfies the following condition

$\mathcal{H}_1$) *there exist two known constants* $\mu_0 > 0$ *and* $\mu_1 > 0$ *and there exists* $M^* > 0$ *such that*

$$\sup_{1\le k\le n}\left(\ln \mathbf{P}\left(\sum_{j=1}^{k}\xi_j^2 > b\right) - \mu_0\, k\right) \le \ln M^* - \mu_1\, b \tag{5}$$

for any $b > 0$.

In Appendices 8.1, 8.2, we show that, for any variance mixture of normal distributions as (2) such that the mixing random variable ς is bounded from above by a known positive constant σ^*, we have $\mu_0 = -0.5\,\ln(1-2\mu_1\sigma^*)$, $0 < \mu_1 < 1/2\sigma^*$ and $M^* = 1$. We also show that, for any Kotz's distribution as (3) with variance parameter σ^2 bounded by a known constant σ^*, we have $\mu_0 = -0.5\,\ln(1-2\mu_1\sigma^*)$, $0 < \mu_1 < 1/2\sigma^*$ and $M^* = \mathrm{e}^{2q\,\mu_0}$. Moreover, in Remark 1, we give a method for choosing μ_1 (and hence μ_0) in an optimal way.

Let $\mathcal{L}_2[\mathcal{X}\,,\,1/n]$ be the Hilbert space of functions on $\mathcal{X} = \{x_1, \ldots, x_n\}$ with the scalar product $(t,u)_n = n^{-1}\sum_{i=1}^n t_i\,u_i$, where t and u are functions on $\mathcal{X}$ defined by $t_i = t(x_i)$ and $u_i = u(x_i)$ for $1 \le i \le n$. In the following, we fix a system of orthonormalized functions $\phi_1, \ldots, \phi_n$ in $\mathcal{L}_2[\mathcal{X}\,,\,1/n]$, i.e. $(\phi_i, \phi_j)_n = 0$, if $i \neq j$ and $\|\phi_i\|_n^2 = (\phi_i, \phi_i)_n = 1$.

Given a class $\mathcal{M}_n$ of subsets of $\{1, \ldots, n\}$, we consider a family $(\mathcal{D}_m\,,\,m \in \mathcal{M}_n)$ of linear subspaces of R^n. For any $m \in \mathcal{M}_n$, the space $\mathcal{D}_m$ can be written as $\mathcal{D}_m = \{t \in R^n \,:\, t = \sum_{j\in m}\alpha_j\phi_j\,,\alpha_j \in R\}$ and we denote by d_m its dimension ($d_m = \dim \mathcal{D}_m$). In this setting, Barron et al. (1999) construct a model selection procedure based on the least squares estimators, that is, on the estimators $\hat{S}_m$, which are minimizers with respect to $t \in \mathcal{D}_m$ of the distance $\|Y - t\|_n$ or, equivalently, of the empirical contrast

$$\gamma_n(t) = \|t\|_n^2 - 2(Y,t)_n\,. \tag{6}$$

In contrast to the Barron–Birgé–Massart procedure, we construct a model selection procedure based on a general family of estimators $\tilde{S}_m$ of S, i.e. the $\tilde{S}_m$'s are any mesurable functions of the observations Y taking their values in $\mathcal{D}_m$. Through a family of prior weights $\{l_m\,,\,m \in \mathcal{M}_n : l_m \ge 1\}$ such that

$$l_\infty = \sup_{n\ge 1}\sum_{m\in\mathcal{M}_n}\mathrm{e}^{-l_m\,d_m} < \infty\,, \tag{7}$$

we choose the penality term $P_n(m)$ of the form

$$P_n(m) = 4\frac{(\mu_0 + l_m)d_m}{\mu_1\,n}\,, \tag{8}$$

and denote

$$\tilde{m} = \mathrm{argmin}_{m\in\mathcal{M}_n}\{\gamma_n(\tilde{S}_m) + P_n(m)\}\,, \tag{9}$$

where $\gamma_n(\tilde{S}_m) + P_n(m)$ is the penalized empirical contrast. For the least squares estimators, we use

$$\hat{m} = \mathrm{argmin}_{m\in\mathcal{M}_n}\{\gamma_n(\hat{S}_m) + P_n(m)\}\,. \tag{10}$$

Our aim, in this section, is to prove the following oracle inequality.

Theorem 1 *Under the condition $\mathcal{H}_1$) the estimator $\tilde{S}_{\tilde{m}}$ of S satisfies the inequality*

$$\mathbf{E}_S\,\|\tilde{S}_{\tilde{m}} - S\|_n^2 \le \inf_{m\in\mathcal{M}_n}\left\{3\,\mathbf{E}_S\,\|\tilde{S}_m - S\|_n^2 + l_m^*\,P_n(m)\right\} + \frac{\mu^*}{n}\,, \tag{11}$$

where

$$l_m^* = 2\left(1 + \frac{\mu_0}{\mu_0 + l_m}\right) \quad \text{and} \quad \mu^* = 8\,M^*\,l_\infty\frac{1}{\mu_1}\,. \tag{12}$$

As an immediate consequence of Theorem 1, we give in the following corollary an upper bound for the left hand side of (11).

Corollary 1 *Under the conditions of Theorem 1, we have*

$$\mathbf{E}_S \,\|\tilde{S}_{\tilde{m}} - S\|_n^2 \le \inf_{m\in\mathcal{M}_n} \left\{ 3\,\mathbf{E}_S\,\|\tilde{S}_m - S\|_n^2 + 8\kappa\,\frac{l_m\,d_m}{n} \right\} + \frac{\mu^*}{n} \tag{13}$$

with $\kappa = (2\mu_0 + 1)/\mu_1$.

Remark 1 Notice that, for the Kotz distribution (given in (3)) with variance parameter $\sigma^2 \le \sigma^*$ and for variance mixtures of normal distributions (given in (2)) with mixing random variable ϑ bounded by σ^*, $\kappa = \kappa(\mu_1) = (1 - \ln(1 - 2\mu_1\sigma^*))\,/\mu_1$ with $0 < \mu_1 < 1/2\sigma^*$. By choosing μ_1 to minimize $\kappa(\mu_1)$, we obtain that

$$\mu_0 = \frac{x_* - 2}{2}, \quad \mu_1 = \frac{x_* - 1}{2\sigma^*\,x_*}, \ \ \kappa \;=\; 2\sigma^*\,x_*\,, \tag{14}$$

where x_* is the maximal root of the equation $\ln x = x - 2$.

Proof of Theorem 1 From (1) and (6) we have, for any $t \in \mathbb{R}^n$,

$$\|t - S\|_n^2 = \gamma_n(t) + 2(\xi, t)_n + \|S\|_n^2\,.$$

As by definition of $\tilde{m}$ we have, for any $m \in \mathcal{M}_n$,

$$\gamma_n(\tilde{S}_{\tilde{m}}) + P_n(\tilde{m}) \le \gamma_n(\tilde{S}_m) + P_n(m)$$

this identity leads to

$$\begin{aligned}\|\tilde{S}_{\tilde{m}} - S\|_n^2 &= \gamma_n(\tilde{S}_{\tilde{m}}) \;+\; 2(\xi, \tilde{S}_{\tilde{m}})_n \;+\|S\|_n^2 \\ &\le \gamma_n(\tilde{S}_m) \;+\; P_n(m) \;-\; P_n(\tilde{m}) \;+\; 2(\xi, \tilde{S}_{\tilde{m}})_n \;+\|S\|_n^2 \\ &= \|\tilde{S}_m - S\|_n^2 + P_n(m) - P_n(\tilde{m}) + 2(\xi, \tilde{t})_n \end{aligned} \tag{15}$$

with $\tilde{t} = \tilde{S}_{\tilde{m}} - \tilde{S}_m$.

Let m be fixed. For any $\iota \in \mathcal{M}_n$, we introduce the random variable

$$Z_\iota(t) = \frac{2(\xi, t)_n}{\|t\|_n^2 + \varrho_n^2(\iota)}, \quad t \in \mathcal{D}_\iota + \mathcal{D}_m\,,$$

where $\varrho_n(\iota)$ will be choosen later. Let the functions $\phi_{i_1}, \ldots, \phi_{i_N}$ be a basis in $\mathcal{D}_\iota + \mathcal{D}_m$ where $N = \dim(\mathcal{D}_\iota + \mathcal{D}_m)$ (note that $N \le d_\iota + d_m$). Thus one can write any normalized vector $\bar{t} = t/\|t\|_n \in \mathcal{D}_\iota + \mathcal{D}_m$ as $\bar{t} = \sum_{j=1}^N a_j \phi_{i_j}$ with $\sum_{j=1}^N a_j^2 = 1$. This leads to the representation of $Z_\iota(t)$ as

$$Z_\iota(t) = \frac{2\,n^{-1/2}\,\|t\|_n}{\|t\|_n^2 + \varrho_n^2(\iota)}\eta_n(\iota), \quad \eta_n(\iota) = \sum_{j=1}^N a_j\,\zeta_{i_j}\,,$$

where $\zeta_i = (\xi, \phi_i)_n \sqrt{n}$. Moreover

$$\sup_{t\in\mathcal{D}_\iota+\mathcal{D}_m} |Z_\iota(t)| \le \frac{n^{-1/2}}{\varrho_n(\iota)}\zeta^*\,,$$

where $\zeta^* = \left(\sum_{j=1}^N \zeta_{i_j}^2\right)^{1/2}$.

Notice now that the vector $\zeta = (\zeta_1, \dots, \zeta_n)'$ is an orthonormal transformation of the vector $\xi = (\xi_1, \dots, \xi_n)'$ (i.e. $\zeta = Q\xi$ where Q is a matrix such that $QQ' = I$). By making use of the fact that a spherically symmetric distribution is invariant with respect to any orthonormal transformation, we see that the vector ζ has the same distribution that the vector ξ. Thus the condition $\mathcal{H}_1$) implies that, for any $b > 0$,

$$\begin{aligned}\mathbf{P}_S(\zeta^* > b) &= \mathbf{P}_S\left(\sum_{j=1}^N \zeta_{i_j}^2 > b^2\right) = \mathbf{P}_S\left(\sum_{j=1}^N \xi_{i_j}^2 > b^2\right)\\ &= \mathbf{P}_S\left(\sum_{k=1}^N \xi_k^2 > b^2\right) \le M^* \exp\{\mu_0 N - \mu_1 b^2\}\,.\end{aligned}$$

Choosing $b = b_*(\iota, x) = \sqrt{(\mu_0 N + d_\iota l_\iota + x)/\mu_1}$ with $x > 0$ yields

$$\mathbf{P}_S(\zeta^* > b_*(\iota, x)) \le M^* \mathrm{e}^{-x-d_\iota l_\iota}\,.$$

Therefore, setting $\Gamma_n(x) = \{\sup_{\iota\in\mathcal{M}_n} \zeta^*/b_*(\iota, x) \le 1\}$, we get, for $x > 0$,

$$\mathbf{P}_S(\Gamma_n^c(x)) \le \sum_{\iota\in\mathcal{M}_n} \mathbf{P}_S(\zeta^* > b_*(\iota, x)) \le M^* l_\infty \mathrm{e}^{-x} \tag{16}$$

according to (7).

We now set $\varrho_n(\iota) = n^{-1/2} b_*(\iota, x)/\tau$ where τ is some positive constant which will be choosen below. On the set $\Gamma_n(x)$ we have

$$\sup_{\iota\in\mathcal{M}_n} \sup_{t\in\mathcal{D}_\iota+\mathcal{D}_m} |Z_\iota(t)| \le \tau\,.$$

Thus, on $\Gamma_n(x)$,

$$\begin{aligned}2(\xi, \tilde{t})_n = \;& Z_{\tilde{m}}(\tilde{t})(\|\tilde{t}\|_n^2 + \varrho_n^2(\tilde{m}))\\ \le\;& \tau\,\|\tilde{t}\|_n^2 + b_*^2(\tilde{m}, x)/n\tau \le 2\tau\|\tilde{S}_{\tilde{m}} - S\|^2 + 2\tau\|\tilde{S}_m - S\|^2\\ &+\frac{1}{n\mu_1\tau}(\mu_0\, d_m + \mu_0\, d_{\tilde{m}} + d_{\tilde{m}} l_{\tilde{m}}) + \frac{1}{\tau}\,\frac{x}{n\mu_1}\,.\end{aligned}$$

Applying this inequality to (15), we get on the set $\Gamma_n(x)$

$$\begin{aligned}\|\tilde{S}_{\tilde{m}} - S\|_n^2 \le\;& \frac{(1+2\tau)}{1-2\tau}\|\tilde{S}_m - S\|_n^2 + \frac{P_n(m) - P_n(\tilde{m})}{1-2\tau}\\ &+\frac{1}{n\mu_1\tau(1-2\tau)}\left(\mu_0\, d_m + \left(\mu_0 + l_{\tilde{m}}\right)\, d_{\tilde{m}}\right) + \frac{1}{\tau(1-2\tau)}\,\frac{x}{n\mu_1}\,.\end{aligned}$$

In choosing τ to maximize the function $(1-2\tau)\tau$ (i.e. $\tau = 1/4$) we minimize the upper bound of the estimation accuracy and we obtain on $\Gamma_n(x)$

$$\|\tilde{S}_{\tilde{m}} - S\|_n^2 \le 3\,\|\tilde{S}_m - S\|_n^2 + l_m^*\, P_n(m) \,+\, \tfrac{8}{n\mu_1}\, x\,, \tag{17}$$

where l_m^* is defined in (12). Furthermore, setting

$$\eta = \|\tilde{S}_{\tilde{m}} - S\|_n^2 - 3\,\|\tilde{S}_m - S\|_n^2 - l_m^*\, P_n(m)\,,$$

we have

$$\|\tilde{S}_{\tilde{m}} - S\|_n^2 \le 3\,\|\tilde{S}_m - S\|_n^2 + l_m^*\, P_n(m) + \eta_+\,,$$

where $\eta_+ = \max(\eta, 0)$. Inequality (17) implies that $\eta \le 8\,x/(\mu_1 n)$ on the set $\Gamma_n(x)$. Therefore, by definition of η and by (16), we obtain that for $x > 0$,

$$\mathbf{P}_S\,(\eta_+ > 8\,x/(\mu_1 n)) = \mathbf{P}_S(\eta > 8\,x/(\mu_1 n),\quad \Gamma_n^c(x)) \le M^*\, l_\infty \mathrm{e}^{-x}\,.$$

Then it follows that

$$\mathbf{E}_S \eta_+ = \int_0^\infty \mathbf{P}_S(\eta_+ > z)\mathrm{d}z = \frac{8}{\mu_1 n} \int_0^\infty \mathbf{P}_S(\eta_+ > 8\,x/(\mu_1 n))\mathrm{d}x \le \mu^*\, n^{-1}\,,$$

which implies, by definition of η, that

$$\mathbf{E}_S\, \|\tilde{S}_{\tilde{m}} - S\|_n^2 \le 3\,\mathbf{E}_S\, \|\tilde{S}_m - S\|_n^2 + l_m^*\, P_n(m) \,+\, \mu^*/n\,.$$

As m is arbitary choosen in $\mathcal{M}_n$, we obtain Inequality (11) and, finally, Theorem 1 is proved. □

Remark 2 We can calculate the right hand side of the oracle inequality (13) for the least squares model selection procedure $\hat{S}_{\hat{m}}$ defined by (10). Indeed, denoting by S_m the orthogonal projection of S on $\mathcal{D}_m$, it is easy to show that, for the least squares estimate, $\mathbf{E}_S\, \|\hat{S}_m - S_m\|_n^2 = \frac{d_m\, \varsigma_n}{n}$, where $\varsigma_n = \mathbf{E}\, \xi_1^2$. Therefore, since $\|\hat{S}_m - S\|_n^2 = \|S_m - S\|_n^2 + \|\hat{S}_m - S_m\|_n^2$, we can rewrite Inequality (13) as

$$\mathbf{E}_S\, \|\hat{S}_{\hat{m}} - S\|_n^2 \le \inf_{m\in\mathcal{M}_n}\, \hat{a}_m(S) + \frac{\mu^*}{n} = \hat{a}(S)\,, \tag{18}$$

where

$$\begin{aligned} \hat{a}_m(S) &= 3\,\mathbf{E}_S\, \|\hat{S}_m - S\|_n^2 \,+\, 8\kappa\, \frac{l_m\, d_m}{n} \\ &\le\, 3\,\|S_m - S\|_n^2 \,+\, \left(3\varsigma_n + 8\frac{2\mu_0 + 1}{\mu_1}\right) \frac{d_m\, l_m}{n}\,. \end{aligned} \tag{19}$$

Then it is particularly interesting to consider the Kotz distribution (3) with known scale parameter (i.e. $\sigma^* = \sigma^2$). In that case, $\varsigma_n = \sigma^2(1 + 2q/n)$ so that, for the model selection procedure $\hat{S}_{\hat{m}}$ with parameters μ_0 and μ_1 defined by (14), we obtain that

$$\hat{a}_m(S) \le\, 3\,\|S_m - S\|_n^2 + \sigma^2\,(3(1 + 2q/n) + 16\,x_*)\tfrac{l_m\, d_m}{n}$$

and hence

$$\hat{a}(S) \leq \inf_{m \in \mathcal{M}_n} \hat{a}_m(S) + 16\, l_\infty \frac{e^{2q\mu_0}\, x_* \sigma^2}{(x_* - 1)n} .$$

Note that, for the gaussian model (i.e. the Kotz distribution with parameter $q = 0$) with $\sigma^2 = 1$, we have

$$\hat{a}(S) \leq \inf_{m \in \mathcal{M}_n} \left\{ 3\, \|S_m - S\|_n^2 + (3 + 16\, x_*) \frac{l_m\, d_m}{n} \right\} + \frac{16\, x_*}{(x_* - 1)}\, l_\infty \frac{1}{n} .$$

Taking into account that $x_* \approx 3.1462$, we obtain that $3 + 16\, x_* \approx 53.34$ and $16\, x_*/(x_* - 1) \approx 23.46$. This bound is sharper than the corresponding one in Barron et al. formula (2.3), for which the upper bound in (18) is equal to

$$\inf_{m \in \mathcal{M}_n} \left\{ 3\, \|S_m - S\|_n^2 + 72\, \frac{l_m\, d_m}{n} \right\} + 32\, l_\infty \frac{1}{n} .$$

3 Improved estimators under projections

In this section, we consider competitive estimators

$$S_m^* \, : \, Y \in \mathbf{R}^n \longmapsto S_m^*(Y) \in \mathcal{D}_m$$

of S which improve on $\hat{S}_m$ in the sense that, for any $S \in \mathbf{R}^n$,

$$\mathbf{E}_S\, \|S_m^* - S\|_n^2 \leq \mathbf{E}_S\, \|\hat{S}_m - S\|_n^2 \tag{20}$$

with strict inequality for some S.

Note that $\hat{S}_m$ is also the orthogonal projector from $\mathbf{R}^n$ onto $\mathcal{D}_m$ with respect to the usual inner product $(.,.)$ given by $(t, u) = \sum_{i=1}^n t_i u_i$ and that the inequality in (20) is equivalent to

$$\mathbf{E}_S\, \|S_m^* - S\|^2 \leq \mathbf{E}_S\, \|\hat{S}_m - S\|^2 \tag{21}$$

where $\|\cdot\|$ is usual norm ($\|u\|^2 = \sum_{i=1}^n u_i^2$). Note also that, if S_m denotes the orthogonal projection of S on $\mathcal{D}_m$, then the inequality in (21) reduces to

$$\mathbf{E}_S\, \|S_m^* - S_m\|^2 \leq \mathbf{E}_S\, \|\hat{S}_m - S_m\|^2 \tag{22}$$

since, for any $t \in \mathcal{D}_m$, $\|t - S\|^2 = \|t - S_m\|^2 + \|S_m - S\|^2$. The expectations in (22) represent the quadratic risks of the considered estimators. It is clear that the risk of $\hat{S}_m$ is finite as soon as the distribution of Y has a finite second moment.

Then, as any estimator S_m^* can be written as $S_m^* = \hat{S}_m + \psi_m$ (with $\psi_m = \hat{S}_m - S_m^*$), S_m^* has a finite risk if and only if $\mathbf{E}_S\, \|\psi_m\|^2 < \infty$ (which is assumed in the following). This can be seen through the risk difference between S_m^* and $\hat{S}_m$ expressed, for any S, as

$$\begin{aligned} \Delta_m(S) &:= \mathbf{E}_S\, \|S_m^* - S_m\|^2 - \mathbf{E}_S\, \|\hat{S}_m - S_m\|^2 \\ &= 2\, \mathbf{E}_S\, (\psi_m, \hat{S}_m - S_m) + \mathbf{E}_S\, \|\psi_m\|^2 \end{aligned} \tag{23}$$

and through Schwarz's inequality.

In the following, we yield another expression of $\Delta_m(S)$. To this end, we need to specify, for $m = \{j_1, \ldots, j_{d_m}\} \subseteq \{1, \ldots, n\}$ fixed, the expressions of $\hat{S}_m$ and S^*_m. First note that $\mathcal{D}_m = \{t \in \mathbf{R}^n \ : \ t = \sum_{k=1}^{d_m} \alpha_k \varphi_{j_k}, \ \alpha_k \in \mathbf{R}\}$, where $\varphi_j = \phi_j/\sqrt{n}$. Therefore the least squares estimator is equal to $\hat{S}_m = \sum_{k=1}^{d_m} \hat{s}_{j_k} \varphi_{j_k}$ with $\hat{s}_j = (Y, \varphi_j)$. Moreover, if we denote by $i \ : \ \mathcal{D}_m \to \mathbf{R}^{d_m}$ the natural isomorphism (i.e. $i(t) = (\alpha_1, \ldots, \alpha_{d_m})'$ for $t = \sum_{k=1}^{d_m} \alpha_k \varphi_{j_k}$), the random vector $\hat{J}_m = i(\hat{S}_m) = (\hat{s}_{j_1}, \ldots, \hat{s}_{j_{d_m}})'$ has a spherically symmetric distribution in $\mathbf{R}^{d_m}$ with density $g_m(\|u - J_m\|^2)$, where $J_m = i(S_m)$ and where $g_m(w) = g_{d_m}(w) = \int_{\mathbf{R}^{n-d_m}} g(w + \|z\|^2)\,\mathrm{d}z$.

Now, the alternative estimators S^*_m that we consider take their values in $\mathcal{D}_m$ and, more specifically, are functions of $\hat{S}_m$, that is

$$S^*_m = \hat{S}_m + \psi_m(\hat{S}_m) \quad \text{with} \quad \psi_m(\hat{S}_m) = \Psi_m(\hat{J}_m)\,, \tag{24}$$

where $\Psi_m(\cdot)$ is a function from $\mathbf{R}^{d_m}$ into $\mathcal{D}_m$.

Finally, for $u = (u_1, \ldots, u_{d_m})$, we define the divergence of $v_m(\cdot) = i(\Psi_m(\cdot))$ as $\operatorname{div} v_m(u) = \sum_{i=1}^{d_m} \partial < v_m(u) >_i /\partial u_i$, where $< v_m(u) >_i$ is the ith component of the vector $v_m(\cdot)$.

Theorem 2 *Let S^*_m be an estimator as in* (24) *such that $\mathbf{E}_S \|\psi_m\|^2 < \infty$. Then the risk difference between S^*_m and $\hat{S}_m$ equals, for any S,*

$$\Delta_m(S) = \mathbf{E}_S \left(2\, G_m(\|\hat{S}_m - S_m\|^2) \operatorname{div} v_m(\hat{J}_m) + \|\psi_m\|^2 \right),$$

where $G_m(w) = \int_w^\infty g_m(a)\mathrm{d}a/2g_m(w)$.

Remark 3 The statement of Theorem 2 assumes that $\operatorname{div}\psi_m$ exists. In order to include basic examples of functions ψ_m proportional to the function $u \longrightarrow \frac{u}{\|u\|^2}$ which blows up at zero, it is assumed that ψ_m is a weakly differentiable function. This assumption is well adapted to the Stokes theorem which is the basis of the proof of Theorem 2 (see, for example, Ziemer 1989).

Proof of Theorem 2 According to (23) the proof reduces to show that

$$\mathbf{E}_S\,(\psi_m, \hat{S}_m - S_m) = \mathbf{E}_S\, G_m(\|\hat{S}_m - S_m\|^2) \operatorname{div} v_m(\hat{J}_m)\,.$$

Now, taking into account that

$$(\psi_m, \hat{S}_m - S_m) = (i(\psi_m), i(\hat{S}_m) - i(S_m)) = (v_m(\hat{J}_m), \hat{J}_m - J_m)\,,$$

we obtain

$$\begin{aligned}
\mathbf{E}_S\,(\psi_m, \hat{S}_m - S_m) &= \int_{\mathbf{R}^{d_m}} (v_m(u), u - J_m)\, g_m\left(\|u - J_m\|^2\right) \mathrm{d}u \\
&= \int_0^\infty \left(\int_{\mathcal{S}_{r,m}} (v_m(u), u - J_m)\, \varrho_{r,m}(\mathrm{d}u) \right) g_m(r^2)\,\mathrm{d}r
\end{aligned}$$

where $\varrho_{r,m}$ is the superficial measure on the sphere $\mathcal{S}_{r,m} = \{u \in \mathbf{R}^{d_m} \, : \, \|u - J_m\| = r\}$ of radius r and centered at J_m. Introducing the unit outward normal vector $n(u) = (u - J_m)/\|u - J_m\|$ at $u \in \mathcal{S}_{r,m}$, it follows that

$$\mathbf{E}_S\,(\psi_m, \hat{S}_m - S_m) = \int_0^\infty \left(\int_{\mathcal{S}_{r,m}} (v_m(u), n(u))\, \varrho_{r,m}(\mathrm{d}u) \right) r\, g_m(r^2)\mathrm{d}r$$
$$= \int_0^\infty \left(\int_{B_{r,m}} \mathrm{div}\, v_m(u)\mathrm{d}u \right) r\, g_m(r^2)\,\mathrm{d}r$$

by applying the Stokes theorem where $B_{r,m} = \{u \in \mathbf{R}^{d_m} \, : \, \|u - J_m\| \leq r\}$. Now, through Fubini's theorem, we have

$$\mathbf{E}_S\,(\psi_m, \hat{S}_m - S_m) = \int_{\mathbf{R}^{d_m}} \left(\int_{\|u - S_m\|}^\infty r\, g_m(r^2)\mathrm{d}r \right) \mathrm{div}\, v_m(u) du$$
$$= \int_{\mathbf{R}^{d_m}} \left(\frac{1}{2} \int_{\|u - S_m\|^2}^\infty g_m(a)\,\mathrm{d}a \right) \mathrm{div}\, v_m(u)\mathrm{d}u$$

with the change of variable $a = r^2$. Finally, dividing and multiplying trough by $g_m(\|u - J_m\|^2)$, we obtain

$$\mathbf{E}_S\,(\psi_m, \hat{S}_m - S_m) = \int_{\mathbf{R}^{d_m}} G_m\left(\|u - J_m\|^2\right)\, g_m\left(\|u - J_m\|^2\right)\, \mathrm{div}\, v_m(u)\mathrm{d}u$$
$$= \mathbf{E}_S\, G_m(\|\hat{J}_m - J_m\|^2)\, \mathrm{div}\, v_m(\hat{J}_m)$$
$$= \mathbf{E}_S\, G_m(\|\hat{S}_m - S_m\|^2)\, \mathrm{div}\, v_m(\hat{J}_m)\,,$$

which is the desired result. □

Remark 4 It is easy to check that, when the sampling distribution is normal $\mathcal{N}(0, \sigma^2)$, then the function G_m is constant and equal to σ^2. In that case, the risk difference is

$$\Delta_m(S) = 2\,\sigma^2\, \mathbf{E}_S\, \mathrm{div}\, v_m(\hat{J}_m) + \mathbf{E}_S\, \|\psi_m\|^2\,,$$

which is the expression given by Stein (1981).

We now give a sufficient condition, denoted by $\mathcal{H}_2$), for which S_m^* improves on $\hat{S}_m$, that is, for which $\Delta_m(S) \leq 0$. It is well known that such an improvement can only happen if the dimension d_m is greater than or equal to 3 (see Stein (1956) in the normal case and Brown (1966) in the general case). This fact means that the least squares estimator $\hat{S}_m$ is admissible when $d_m \leq 2$.

$\mathcal{H}_2$) *There exists a constant $c > 0$ such that, for any $w \geq 0$, $G_m(w) \geq c$.*

Theorem 3 *Assume that the condition $\mathcal{H}_2$) holds. Assume also that $d_m \geq 3$ and let S_m^* an estimator as in* (24) *such that* $\mathbf{E}_S \, \|\psi_m\|^2 < \infty$.

Then a sufficient condition for which S_m^* improves on $\hat{S}_m$ is that, for any $u \in \mathbf{R}^{d_m}$,

$$L_m(u) = 2\, c \operatorname{div} v_m(u) + \| \, v_m(u)\|^2 \leq 0 \tag{25}$$

(with strict inequality on a set of positive Lebesgue measure). Moreover,

$$\Delta_m(S) \leq \mathbf{E}_S \, L_m(\hat{J}_m) \,. \tag{26}$$

Proof The result follows immediately from Theorem 2 since condition (25) implies that div $\psi_m(u) \leq 0$. Note that, the fact that $d_m \geq 3$ is implicit in (25). Indeed, Blanchard and Fourdrinier (1999) showed that Inequation (25) has non trivial solutions only in the case where $d_m \geq 3$. □

Remark 5 While condition (25) is a condition on the estimator S_m^*, the assumption $G_m(w) \geq c$ is a condition on the sampling distribution.

Example 1 Consider the Kotz distribution in $\mathbf{R}^n$ with density (48), where we assume $q \geq 0$. Thus the generating function $g(w)$ is proportional to $w^q \mathrm{e}^{-w/2\sigma^2}$ and we have

$$G(w) = \frac{1}{2} \frac{\int_w^\infty g(a)\,\mathrm{d}a}{g(w)} = \sigma^2 + \sigma^2\, q\, \frac{\int_w^\infty a^{q-1}\, \mathrm{e}^{-a/2\sigma^2}\,\mathrm{d}a}{w^q \mathrm{e}^{-w/2\sigma^2}} \geq \sigma^2 \;.$$

Then, according to Lemma 4 in the Sect 8.3, the generating function g_{d_m} satisfies

$$G_{d_m}(w) = \frac{1}{2} \frac{\int_w^\infty g_{d_m}(a)\,\mathrm{d}a}{g_{d_m}(w)} \geq \sigma^2 \;.$$

Thus the distributional assumption $\mathcal{H}_2$) is satisfied with the constant $c = \sigma^2$. Now Inequation (25) has the classical James–Stein type solutions

$$v_m(u) = -\alpha \frac{u}{\|u\|^2} \tag{27}$$

with $0 < \alpha < 2(d_m - 2)\sigma^2$. It is shown in the Appendix (see Sect. 8.4) that

$$\mathbf{E}_S \, \|\psi_m(\hat{S}_m)\|^2 = \mathbf{E}_S \, \|v_m(\hat{S}_m)\|^2 < \infty \tag{28}$$

for $d_m \geq 3$. Moreover, with $\alpha = (d_m - 2)\sigma^2$, it is easy to check that

$$L_m(u) = -\frac{(d_m - 2)^2 \sigma^4}{\|u\|^2} \,. \tag{29}$$

Now we can apply the improved estimator to the selection model procedure. We pose

$$m^* = \operatorname{argmin}_{m\in\mathcal{M}_n}\{\gamma_n(S^*_m) + P_n(m)\}\,, \tag{30}$$

where $\gamma_n(\cdot)$ and $P_n(m)$ are defined by (6) and (8), respectively.

Theorem 4 *Under the conditions $\mathcal{H}_1$)–$\mathcal{H}_2$) the improved model selection procedure $S^*_{m^*}$ of S satisfies the inequality*

$$\mathbf{E}_S\,\|S^*_{m^*} - S\|^2_n \le \inf_{m\in\mathcal{M}_n}\left\{3\,\mathbf{E}_S\,\|S^*_m - S\|^2_n + l^*_m\,P_n(m)\right\} + \frac{\mu^*}{n}\,, \tag{31}$$

where l^*_m and μ^* are defined in (12). Moreover, for any $m\in\mathcal{M}_n$,

$$\mathbf{E}_S\,\|S^*_m - S\|^2_n \le \mathbf{E}_S\,\|\hat{S}_m - S\|^2_n\,.$$

Proof This theorem follows immediately from Theorems 1 and 3. □

Corollary 2 *Under the conditions $\mathcal{H}_1$)–$\mathcal{H}_2$) the improved selection model estimator $S^*_{m^*}$ of S satisfies the inequality*

$$\mathbf{E}_S\,\|S^*_{m^*} - S\|^2_n \le \inf_{m\in\mathcal{M}_n}\ a^*_m(S) + \frac{\mu^*}{n}\,, \tag{32}$$

with $a^*_m(S) = \hat{a}_m(S) + 3\Delta_m(S)/n$, where $\hat{a}_m(S)$ is defined in (19) *and where $\Delta_m(S)\le 0$, for any $m\in\mathcal{M}_n$.*

Proof The corollary follows immediately from Inequality (18) and Theorem 3. □

Now we compare the improved estimator S^*_m in (24), corresponding to the function (27) for $\alpha = (d_m-2)\sigma^2$, with the least squares estimator $\hat{S}_m$ in terms of relative efficiency (of S^*_m with respect to $\hat{S}_m$) defined by

$$\epsilon_m(S) = \frac{\mathbf{E}_S\,\|S^*_m - S\|^2_n}{\mathbf{E}_S\,\|\hat{S}_m - S\|^2_n}\,.$$

An upper asymptotic bound for $\epsilon_m(S)$ is given in the following theorem.

Theorem 5 *For the Kotz distribution given by* (3)*, the improved estimator S^*_m of S satisfies*

$$\limsup_{d_m\to\infty}\,\epsilon_m(S) \le \epsilon^*(S) = \frac{\gamma(S) + d^*_1(S)}{\sigma^2 + d^*_1(S)} \tag{33}$$

where

$$\gamma(S) = \frac{\sigma^2 d^*_0(S)}{\sigma^2 + d^*_0(S)}\,,\quad d^*_0(S) = \limsup_{d_m\to\infty}\,\frac{\|S_m\|^2}{d_m}$$

$$\text{and}\ \ d^*_1(S) = \limsup_{d_m\to\infty}\,\frac{\|S_m - S\|^2}{d_m}\,.$$

Proof By (26) and (29) we have, for $d_m \geq 3$,

$$\begin{aligned}\mathbf{E}_S \|S^*_m - S_m\|^2 &\leq \mathbf{E}_S \|\hat{S}_m - S_m\|^2 + \mathbf{E}_S\, L_m(\hat{S}_m)\\ &= \mathbf{E}_S \|\hat{S}_m - S_m\|^2 - (d_m-2)^2\,\sigma^4 \mathbf{E}_S \,\frac{1}{\|\hat{S}_m\|^2}\\ &\leq \mathbf{E}_S \|\hat{S}_m - S_m\|^2 - (d_m-2)^2\,\sigma^4 \,\frac{1}{\mathbf{E}_S \|\hat{S}_m\|^2}\,.\end{aligned}$$

Thus

$$\begin{aligned}\frac{\mathbf{E}_S \|S^*_m - S_m\|^2_n}{\mathbf{E}_S \|\hat{S}_m - S_m\|^2_n} &\leq 1 - \frac{(d_m-2)^2\sigma^4}{d_m\,\varsigma_n}\,\frac{1}{\mathbf{E}_S \|\hat{S}_m\|^2}\\ &\leq \frac{d_m^2\,((\varsigma_n)^2 - \sigma^4) + 4(d_m-1)\sigma^4 + \varsigma_n \|S_m\|^2\, d_m}{d_m\varsigma_n(\|S_m\|^2 + d_m\,\varsigma_n)}\,.\end{aligned}$$

As, for (3), $\varsigma_n = \sigma^2(1 + 2q/n)$ we have the desired result. □

Remark 6 Note that, for $S = 0$, $\epsilon^*(S) = 0$. Now, for a continuous function S, if the dimension d_m goes to infinity with n (more specifically, if $d_m/n \to \delta > 0$) and if $\|S_m - S\|^2 = \mathrm{o}(n)$ (i.e. $d^*_1(S) = 0$) then

$$\epsilon^*(S) = \frac{\int_0^1 S^2(x)\,\mathrm{d}x}{\sigma^2\delta + \int_0^1 S^2(x)\,\mathrm{d}x}\,.$$

4 Estimation with known smoothness properties

Using the following system of trigonometric functions in $\mathcal{L}_2[0,1]$

$$\begin{aligned}&e_1(x) = 1,\quad e_2(x) = \sqrt{2}\cos(2\pi\, x),\quad e_3(x) = \sqrt{2}\sin(2\pi\, x),\ldots,\\ &e_{2i}(x) = \sqrt{2}\cos(2i\pi\, x),\quad e_{2i+1}(x) = \sqrt{2}\sin(2i\pi\, x),\ldots.\end{aligned}$$

we assume that the function S to estimate belongs to

$$\Theta_{\beta,r} = \{S\,:\,S(x) = \sum_{j=1}^{\infty} s_j\, e_j(x)\,,\ \max_{l\geq 1}\, l^{2\beta}\, Q_l(S) \leq r^2\}\,, \tag{34}$$

where $Q_l(S) = \sum_{j=l}^{+\infty} s_j^2$ and $\beta > 1$ is a known constant, and is 1-periodic (i.e. $S(0) = S(1)$). Any estimator $\tilde{S}_n$ (i.e. a measurable function of the observations $(y_1,\ldots,y_n)$ in (5)) is evaluated by the risk defined by

$$\mathcal{R}_n(\tilde{S}_n) = \sup_{S\in\Theta_{\beta,r}} \mathbf{E}_S\|\varphi_n(\tilde{S}_n - S)\|^2_n\,, \tag{35}$$

where $\varphi_n = n^{\beta/2\beta+1}$. For this problem, we apply the ordered variables selection (see Barron et al. (1999) p. 315), where the set $\mathcal{M}_n$ is defined as $\mathcal{M}_n = \{m_1, \ldots, m_n\}$, with $m_i = \{1, \ldots, i\}$ (therefore $d_{m_i} = i$). Notice that, in this case,

$$\mathcal{D}_{m_i} = \{t = (t_1, \ldots, t_n) : \, t_l = \sum_{j=1}^{i} \alpha_j \phi_j(x_l)\,, \alpha_j \in R\,,\; x_l = l/n\}$$

where $\phi_j(x) = \nu_j e_j(x)$, $\nu_j = 1$ for $1 \le j \le n-1$, $\nu_n = 1/\sqrt{2}$ if n is even and $\nu_n = 1$ if n is odd. Note also that the system (ϕ_j) is an orthonormal basis in $\mathcal{L}_2[\mathcal{X}\,, 1/n]$, i.e. $(\phi_j, \phi_l)_n = 0$ for $j \neq l$ and $1 \le j, l \le n$ and $\|\phi_j\|_n^2 = 1$ for $1 \le j \le n$. In this context, one takes $l_m = 1$ for any $m \ge 1$, so that, according to (7), $l_\infty = \sum_{i=1}^{\infty} \exp(-\mathrm{d}_{\mathrm{m_i}}) = 1/(\mathrm{e}-1)$.

Theorem 6 *Under condition $\mathcal{H}_1)$ the upper bound of the risk $\mathcal{R}_n(\hat{S}_{\hat{m}})$ for the least squares model selection procedure $\hat{S}_{\hat{m}}$ in (10) is finite, i.e.*

$$\limsup_{n\to\infty} \mathcal{R}_n(\hat{S}_{\hat{m}}) < \infty\,. \tag{36}$$

Proof First of all, we show that condition $\mathcal{H}_1)$ implies that

$$\limsup_{n\to\infty} \varsigma_n \;\le\; \mu_0/\mu_1 \tag{37}$$

where μ_0 and μ_1 are the two known constants involved in $\mathcal{H}_1)$. Notice that

$$\varsigma_n \;=\; \mathbf{E}\,\xi_1^2 \;=\; \frac{1}{n}\,\mathbf{E}\sum_{j=1}^{n} \xi_j^2 \;=\; \frac{1}{n}\int_0^\infty \mathbf{P}\left(\sum_{j=1}^{n} \xi_j^2 \;\ge\; x\right)\,\mathrm{d}x$$

and hence, for any $x_0 > 0$,

$$\varsigma_n \;\le\; \frac{x_0}{n} \;+\; \frac{1}{n}\int_{x_0}^\infty \mathbf{P}\left(\sum_{j=1}^{n} \xi_j^2 \;\ge\; x\right)\,\mathrm{d}x$$

which implies, by (5), that $\varsigma_n \;\le\; n^{-1} + (\mu_1\, n)^{-1}\, M^*\, \mathrm{e}^{\mu_0 n - \mu_1 x_0}$. Thus, setting $x_0 = \mu_0 n/\mu_1$, we obtain (37). Now, (18) and (19) imply that

$$\begin{aligned}\mathbf{E}_S\|\hat{S}_{\hat{m}} - S\|_n^2 &\le \inf_{m\in\mathcal{M}_n}\left\{3\|S - S_m\|_n^2 + \left(3\varsigma_n \;+\; \frac{8(2\mu_0+1)}{\mu_1}\right)\frac{d_m}{n}\right\} + \mu^*\,\frac{1}{n}\\ &\le c_n^* \inf_{1\le k\le n}\left\{\|S - S_{m_k}\|_n^2 + \frac{k}{n}\right\} + \mu^*\,\frac{1}{n}\,,\end{aligned}$$

where $c_n^* = 3 + 3\varsigma_n \;+\; 8(2\mu_0+1)/\mu_1$ and $S_m(x) = \sum_{j\in m} \alpha_j^* \phi_j(x)$ with $\alpha_j^* = (S, \phi_j)_n$. Taking into account that $\|S_m - S\|_n^2 = \inf_{\alpha_j}\; \|S - \sum_{j\in m} \alpha_j \phi_j\|_n^2$, it is easy to derive that $\|S_{m_k} - S\|_n^2 \le \|\Delta_k\|_n^2$ where $\Delta_k = \sum_{j=k+1}^{+\infty} s_j\, e_j$. Notice now that Inequality (34) allows

$$\int_0^1 \dot{\Delta}_k^2(x)\,\mathrm{d}x \le \pi^2 \sum_{l\ge 1} s_l^2\, l^2 = Q_1(S) + 2\sum_{l\ge 1} l\, Q_{l+1}(S) + \sum_{l\ge 1} Q_{l+1}(S)$$
$$\le r^2(1 + 2\sum_{l\ge 1} l^{-2\beta+1} + \sum_{l\ge 1} l^{-2\beta}) = r^* < \infty$$

for any $S \in \Theta_{\beta,r}$ with $\beta > 1$. In order to estimate $\|\Delta_k\|_n^2$, we can write

$$\begin{aligned}
\|\Delta_k\|_n^2 &= \frac{1}{n}\sum_{l=1}^n \Delta_k^2(x_l)\\
&= \sum_{l=1}^n \int_{x_{l-1}}^{x_l} [\Delta_k(x) + (\Delta_k(x_l) - \Delta_k(x))]^2\,\mathrm{d}x\\
&\le 2\int_0^1 \Delta_k^2(x)\mathrm{d}x + 2\sum_{l=1}^n \int_{x_{l-1}}^{x_l} (\Delta_k(x_l) - \Delta_k(x))^2\,\mathrm{d}x\,.
\end{aligned}$$

Now, by Bunyakovski–Cauchy–Schwarz inequality, we have

$$\begin{aligned}
\sum_{l=1}^n \int_{x_{l-1}}^{x_l} (\Delta_k(x_l) - \Delta_k(x))^2\,\mathrm{d}x &= \sum_{l=1}^n \int_{x_{l-1}}^{x_l} \left(\int_x^{x_l} \dot{\Delta}_k(t)\,\mathrm{d}t\right)^2 \mathrm{d}x\\
&\le \sum_{l=1}^n \int_{x_{l-1}}^{x_l} \int_x^{x_l} (\dot{\Delta}_k(t))^2\,\mathrm{d}t\,(x_l - x)\,\mathrm{d}x\\
&\le n^{-2}\sum_{l=1}^n \int_{x_{l-1}}^{x_l} (\dot{\Delta}_k(t))^2\,\mathrm{d}t\\
&= n^{-2}\int_0^1 (\dot{\Delta}_k(x))^2\,\mathrm{d}x
\end{aligned}$$

and hence $\|\Delta_k\|_n^2 \le 2Q_{k+1}(S) + 2\,r^*\,n^{-2}$. Therefore, by Inequality (34),

$$\begin{aligned}
\mathbf{E}_S\|\hat{S}_{\hat{m}} - S\|_n^2 &\le c^* \inf_{1\le k\le n}\left\{2Q_{k+1}(S) + \frac{k}{n}\right\} + \mu^*\,\frac{1}{n} + 2\,r^*\,\frac{1}{n^2}\\
&\le c^*\,(2\,r^2\,(k+1)^{-2\beta} + k/n) + \mu^*\,\frac{1}{n} + 2\,r^*\,\frac{1}{n^2}
\end{aligned}$$

for any $1 \le k \le n$. Choosing $k = k_n = [n^{\frac{1}{2\beta+1}}]$ yields

$$\mathbf{E}_S \| \hat{S}_{\hat{m}} - S \|_n^2 \le c_n^* \, (2\, r^2 + 1)\, n^{-2\beta/(2\beta+1)} + \mu^* \, \frac{1}{n} + 2\, r^* \, \frac{1}{n^2} \,. \tag{38}$$

Finally Condition (36) follows from (38) and (37). □

Theorem 6 and Corollary 2 immediately imply the next result.

Theorem 7 *Under conditions $\mathcal{H}_1$)–$\mathcal{H}_2$) the upper bound of the risk $\mathcal{R}_n$ for the improved model selection estimator $S^*_{m^*}$ defined by (24) and (30) is finite, i.e.*

$$\limsup_{n\to\infty} \mathcal{R}_n(S^*_{m^*}) \, < \, \infty \,. \tag{39}$$

Remark 7 Notice that Theorems 6 and 7 give the classical nonparametric convergence rate for the problem with known regularity, i.e. as for the case of independent observations.

5 Lower bound

In this section, we give a simple condition for the lower bound of the risk (35) to be positive. More precisely, we suppose that the density g in model (5) satisfies the following finitness condition:

$$F_g \, = \limsup_{n\to\infty} \; F_n(g) \, < \, \infty \,, \tag{40}$$

where $F_n(g) \, = \, 4 \int_{\mathbf{R}^n} \, u_1^2 \, \frac{\dot{g}^2(\|u\|^2)}{g(\|u\|^2)} \, \mathrm{d}u$. Notice that, for the Kotz distribution (48), we have

$$F_n(g) = \frac{2\Gamma(n/2+q+1) - 4q\Gamma(n/2+q) + 2q^2\Gamma(n/2+q-1)}{\sigma^2\, n\, \Gamma(n/2+q)}$$

and, therefore, (40) is satisfied with

$$F_g \, = \, \lim_{n\to\infty} \; F_n(g) = 1/\sigma^2 \,.$$

Theorem 8 *Under the condition* (40)*, for $\beta > 1$ in* (34)*, the lower bound of the risk $\mathcal{R}_n$ over all estimates is strictly positive, i.e.*

$$\liminf_{n\to\infty} \inf_{\tilde{S}_n} \mathcal{R}_n(\tilde{S}_n) > 0 \,. \tag{41}$$

Proof First, it will be convenient to write β as $\beta = k + \alpha$ where $k \ge 1$ is an integer and $0 \le \alpha < 1$. Now, for $z = (z_1, \ldots, z_m) \in \Pi_\delta = (-\delta, \delta)^m$ with $m = m_n = [n^{\frac{1}{2\beta+1}}]$ and $\delta = \nu/\varphi_n$ for some $\nu > 0$, define a function S_z by

$$S_z(x) = \sum_{l=1}^{m} z_l \psi_l(x) \quad \text{with} \quad \psi_l(x) = V\left(\frac{x - a_l}{h}\right) \,. \tag{42}$$

Here $V(\cdot)$ is a positive function $k+1$ times continuously differentiable such that $V(x)=0$ for $|x|\ge 1$; moreover $a_l = l/(m_n+1)$ and $h=1/2(m_n+1)$. Notice that, for all $1\le j\le k$, the jth derivative of S_z equals

$$S_z^{(j)}(x) = \sum_{l=1}^{m} z_l\, \frac{1}{h^j} V^{(j)}\left(\frac{x-a_l}{h}\right)$$

and hence $S_z^{(j)}(0) = S_z^{(j)}(1) = 0$. Note also that, for all $z\in\Pi_\delta$ and for all $x, y\in[0,1]$,

$$\max_{1\le j\le k-1} |S_z^{(j)}(x)| \le \nu\, \frac{1}{h^{k-1}\varphi_n}\, V^* \le 4^{k-1}\nu\, V^*$$

with $V^* = \max_{1\le j\le k+1}\sup_{|a|\le 1} |\,V^{(j)}(a)|$ and furthermore

$$|S_z^{(k)}(x) - S_z^{(k)}(y)| \le 2\nu\, \frac{1}{h^\beta\varphi_n}\, V^*\,|x-y|^\alpha \le 2\,\nu\, 4^\beta\, V^*\,|x-y|^\alpha\,.$$

Therefore, by Lemma 6, there exists $\nu>0$ such that $S_z\in\Theta_{\beta,r}$ for all $z\in\Pi_\delta$ and $n\ge 1$.

Now a lower bound for $\mathcal{R}_n(\tilde{S}_n)$ will be obtained through introducing the prior distribution on Π_δ with density

$$\pi_\delta(z) = \pi_\delta(z_1,\ldots,z_m) = \prod_{j=1}^{m} \lambda_\delta(z_j) \quad \text{with} \quad \lambda_\delta(z) = \frac{1}{\delta} G\left(\frac{z}{\delta}\right)$$

and $G(u) = G_* \mathrm{e}^{-\frac{1}{1-u^2}} \mathbf{1}_{\{|u|\le 1\}}$ where $G_* = \left(\int_{-1}^{1} \mathrm{e}^{-\frac{1}{1-v^2}}\,\mathrm{d}v\right)^{-1}$.

For $\tilde{S}_n(x)$, an estimate of $S(x)$ based on observations $y_1,\ldots,y_n$ in (1), we have

$$\sup_{S\in\Theta_{r,\beta}} \mathbf{E}_S\|\tilde{S}_n - S\|_n^2 \ge \int_{\Pi_\delta} \mathbf{E}_z\,\|\tilde{S}_n - S_z\|_n^2\, \pi_\delta(z)\,\mathrm{d}z$$

where $\mathbf{E}_z$ is the expectation under the distribution of $y=(y_1,\ldots,y_n)$ in (1) with $S=S_z$. By definition of S_z, $\|\tilde{S}_n - S_z\|_n^2 \ge \sum_{l=1}^m (\tilde{z}_l - z_l)^2\,\|\psi_l\|_n^2$, where $\tilde{z}_l = \tilde{z}_l(y) = (\tilde{S}_n\,,\psi_l)_n/\|\psi_l\|_n^2$. Therefore

$$\mathcal{R}_n(\tilde{S}_n) = \sup_{S\in\Theta_{r,\beta}} \mathbf{E}_S\|\varphi_n(\tilde{S}_n - S)\|_n^2 \ge \varphi_n^2 \sum_{l=1}^{m} \Lambda_l\,\|\psi_l\|_n^2 \tag{43}$$

where

$$\Lambda_l = \int_{\Pi_\delta} \mathbf{E}_z(\tilde{z}_l - z_l)^2\,\pi_\delta(z)\,\mathrm{d}z = \int_{(-\delta,\delta)^{m-1}} \prod_{j\ne l}^{m} \lambda_\delta(z_j)\; I_l(z) \prod_{j\ne l}^{m} \mathrm{d}z_j$$

and

$$I_l(z) = \int_{-\delta}^{\delta} \mathbf{E}_z (\tilde{z}_l - z_l)^2 \lambda_\delta(z_l) \mathrm{d}z_l = \int_{-\delta}^{\delta} \int_{\mathbf{R}^n} (\tilde{z}_l(y) - z_l)^2\, f(y,z) \mathrm{d}y \lambda_\delta(z_l) \mathrm{d}z_l$$

with $f(y,z) = g(\|y - S_z\|^2)$. In this case, it is easy to see that $\partial f(y,z)/\partial z_l = \zeta_l(y,z) f(y,z)$ where

$$\zeta_l(y,z) = -2 \frac{\dot{g}(\|y - S_z\|^2)}{g(\|y - S_z\|^2)} \sum_{i=1}^{n} (y_i - S_z(x_i))\, V\left(\frac{x_i - a_l}{h}\right).$$

By definition of λ_δ, we have

$$\int_{-\delta}^{\delta} \frac{\partial}{\partial z_l} \left(f(y,z)\lambda_\delta(z_l)\right) \mathrm{d}z_l = \lambda_\delta(\delta) f(y,z)\Big|_{z_l=\delta} - \lambda_\delta(-\delta) f(y,z)\Big|_{z_l=-\delta} = 0$$

and

$$\int_{\mathbf{R}^n} \int_{-\delta}^{\delta} z_l \frac{\partial}{\partial z_l} \left(f(y,z)\lambda_\delta(z_l)\right) \mathrm{d}z_l\, \mathrm{d}y = -\int_{-\delta}^{\delta} \int_{\mathbf{R}^n} f(y,z)\, \mathrm{d}y \lambda_\delta(z_l)) \mathrm{d}z_l = -1\,.$$

Therefore

$$\begin{aligned}
&\int_{\mathbf{R}^n} \int_{-\delta}^{\delta} (\tilde{z}_l(y) - z_l) \frac{\partial}{\partial z_l} \left(f(y,z)\lambda_\delta(z_l)\right) \mathrm{d}z_l\, \mathrm{d}y \\
&\quad = \int_{-\delta}^{\delta} \int_{\mathbf{R}^n} (\tilde{z}_l(y) - z_l) \gamma_l(y,z)\, f(y,z)\, \mathrm{d}y\, \lambda_\delta(z_l)\, \mathrm{d}z_l \\
&\quad = \int_{-\delta}^{\delta} \mathbf{E}_z (\tilde{z}_l - z_l) \gamma_l(z)\, \lambda_\delta(z_l)\, \mathrm{d}z_l = 1\,,
\end{aligned}$$

where

$$\gamma_l(z) = \gamma_l(y,z) = \frac{\frac{\partial}{\partial z_l} \left(f(y,z)\lambda_\delta(z_l)\right)}{f(y,z)\lambda_\delta(z_l)} = \zeta_l(y,z) + i_\delta(z_l)$$

with

$$i_\delta(u) = \frac{\dot{\lambda}_\delta(u)}{\lambda_\delta(u)} = \delta^{-1} \frac{\dot{G}\left(\frac{u}{\delta}\right)}{G\left(\frac{u}{\delta}\right)}\,.$$

Thus, by Cauchy-Bunyakovski-Schwarz inequality, $\Lambda_l \geq 1/J_l(\delta)$, where

$$J_l(\delta) = \int_{\Pi_\delta} \mathbf{E}_z\, \gamma_l^2(z) \pi_\delta(z)\, \mathrm{d}z = \int_{\Pi_\delta} \mathbf{E}_z\, (\zeta_l(z) + i_\delta(z_l))^2 \pi_\delta(z)\, \mathrm{d}z$$

and $\zeta_l(z) = \zeta_l(\cdot, z)$. Taking into acount that

$$\mathbf{E}_z\, \zeta_l(z) = \int_{\mathbf{R}^n} \zeta_l(y, z)\, f(y, z)\, \mathrm{d}y = \int_{\mathbf{R}^n} \frac{\partial}{\partial z_l} f(y, z)\, \mathrm{d}y = 0\,,$$

we obtain that

$$\mathbf{E}_z\, (\zeta_l(z) + i_\delta(z_l))^2 = \mathbf{E}_z\, \zeta_l^2(z) + i_\delta^2(z_l)\,.$$

As for the first term of the right hand side of this last equality, we have

$$\begin{aligned} \mathbf{E}_z\, \zeta_l^2(z) &= 4 \int_{\mathbf{R}^n} \frac{\dot{g}^2(\|u\|^2)}{g(\|u\|^2)} \left(\sum_{i=1}^n u_i\, V\left(\frac{x_i - a_l}{h} \right) \right)^2 \mathrm{d}u \\ &= F_n(g) \sum_{i=1}^n V^2 \left(\frac{x_i - a_l}{h} \right) \\ &= F_n(g)\, n\, \|\psi_l\|_n^2\,, \end{aligned}$$

where the function $F_n(g)$ is defined in Condition (40). Then we obtain that

$$\begin{aligned} J_l(\delta) &= F_n(g)\, n\, \|\psi_l\|_n^2 + \int_{\Pi_\delta} i_\delta^2(z_l) \pi_\delta(z)\, \mathrm{d}z \\ &= F_n(g)\, n\, \|\psi_l\|_n^2 + \int_{-\delta}^{\delta} i_\delta^2(u) \lambda_\delta(u)\, \mathrm{d}u \\ &= F_n(g)\, n\, \|\psi_l\|_n^2 + \varphi_n^2 \nu^{-2} I_G\,, \end{aligned}$$

where $I_G = 8 \int_0^1 u^2 (1 - u^2)^{-4} G(u)\, \mathrm{d}u$.

Finally, according to (43), the risk of $\tilde{S}_n$ satisfies

$$\mathcal{R}_n(\tilde{S}_n) \geq \sum_{l=1}^m \frac{\varphi_n^2\, \|\psi_l\|_n^2}{F_n(g)\, n\, \|\psi_l\|_n^2 + \varphi_n^2 \nu^{-2} I_G}$$

and, furthermore, taking into account that $n\, h/\varphi_n^2 \to 1/2$ as $n \to \infty$ and that

$$\lim_{n\to\infty} \frac{1}{h} \|\psi_l\|_n^2 = \lim_{n\to\infty} \tfrac{1}{n\, h} \textstyle\sum_{k=1}^n V^2 \left(\frac{x_k - a_l}{h} \right) = \displaystyle\int_{-1}^{1} V^2(u)\, \mathrm{d}u = \|V\|^2\,,$$

we obtain that Condition (40) implies that

$$\liminf_{n\to\infty} \inf_{\tilde{S}} \mathcal{R}_n(\tilde{S}_n) \geq \frac{\|V\|^2}{F(g)\, \|V\|^2 + 2\, \nu^{-2} I_G} > 0\,.$$

$\square$

Remark 8 Theorem 8 means that the estimators $\hat{S}_{\hat{m}}$ and $S^*_{m^*}$ are optimal in the sense of the risk (35).

6 Estimation of functions with unknown smoothness properties

We consider here the estimation problem of the function $S \in \cup_{\beta>1}^{\infty} \Theta_{\beta,r}$, where the set $\Theta_{\beta,r}$ is defined in (34).

Let us introduce the adaptive risk related to an estimator $\tilde{S}_n$ as follows

$$\mathcal{R}_n^a(\tilde{S}_n) = \sup_{\beta>1} \sup_{S\in\Theta_{\beta,r}} \mathbf{E}_S \|\varphi_n(\tilde{S}_n - S)\|^2 \quad \text{with} \quad \varphi_n = n^{\frac{\beta}{2\beta+1}} . \tag{44}$$

Note that, in this case, the smoothness parameter $1 < \beta < \infty$ is unknown. Theorem 8 yields immediately that

$$\liminf_{n\to\infty} \inf_{\tilde{S}} \mathcal{R}_n^a(\tilde{S}_n) > 0$$

and Inequality (38) implies that, under the condition $\mathcal{H}_1$),

$$\limsup_{n\to\infty} \mathcal{R}_n^a(\hat{S}_{\hat{m}}) < \infty \tag{45}$$

and, under the conditions $\mathcal{H}_1) - \mathcal{H}_2$),

$$\limsup_{n\to\infty} \mathcal{R}_n^a(S^*_{m^*}) < \infty \tag{46}$$

respectively.

Notice that $\hat{S}_{\hat{m}}$ and $S^*_{m^*}$ are independent of β and (45)–(46) express that they attain the optimal rate for every class $\Theta_{\beta,r}$ uniformly over $\beta > 1$. Such estimators are called optimal rate adaptive (see e.g. Tsybakov 1998).

7 Conclusions

In this article, we provide an extension of the classical model $Y = S + \xi$, where $\xi = (\xi_1, \ldots, \xi_n)'$ is a vector of independent errors ξ_i, to the case where these errors are dependent, but non correlated. Such an extension is given in the framework of the spherically symmetric distributions. Our basic examples are variance mixture of normal distributions and the Kotz distributions.

For this model, we develop the selection model method proposed by Barron et al. (1999) for a regression of the form (1). Our contribution to this problem consists in the construction of a model selection procedure based on general estimators for which we can derive an Oracle inequality. Note that the procedure proposed by Barron et al. (1999) is founded on the least squares estimator alone.

An important feature of our method is that we can improve their procedure replacing the least squares estimator by improved estimators with respect to the quadratic risk.

8 Appendix

8.1 A property of variance mixtures of normal distributions

In this section, we consider a generating function g corresponding to a variance mixture of normal distributions, that is,

$$g(t) = \mathbf{E}\,\frac{1}{(2\,\vartheta\,\pi)^{n/2}}\,\mathrm{e}^{-\frac{t}{2\vartheta}}\,, \tag{47}$$

where ϑ is a non negative random variable.

Lemma 1 *Let $\xi = (\xi_1, \ldots, \xi_n)$ a random vector having a generating function as in (47). If the random variable ϑ is bounded by a constant $\sigma^* > 0$ then, for any $0 < \mu_1 < 1/2\sigma^*$,*

$$\sup_{1\le k\le n}\left(\ln\,\mathbf{E}\,e^{\mu_1\sum_{j=1}^k \xi_j^2} - \mu_0\,k\right) \le\, 0\,,$$

where $\mu_0 = -\frac{1}{2}\ln(1 - 2\sigma^*\mu_1)$.

Proof If $\tilde{\xi}_1, \ldots, \tilde{\xi}_n$ are i.i.d. $\mathcal{N}(0, 1)$ and F_ϑ is the distribution of ϑ

$$\mathbf{E}\,\mathrm{e}^{\mu_1\sum_{j=1}^k \xi_j^2} = \int_0^{\sigma^*} \mathbf{E}\,\mathrm{e}^{v\mu_1\sum_{j=1}^k \tilde{\xi}_j^2}\,F_\vartheta(\mathrm{d}v)$$

Therefore the result follows from the fact that, if $\mu < 1/2\sigma^2$, then

$$\mathbf{E}\,\mathrm{e}^{\mu_1\sum_{j=1}^k \xi_j^2} \le\, \mathbf{E}\,\mathrm{e}^{\sigma^*\mu_1\sum_{j=1}^k \tilde{\xi}_j^2} = \mathrm{e}^{\mu_0 k}\,.$$

$\square$

8.2 A property of the Kotz distribution

Lemma 2 *Let $\xi = (\xi_1, \ldots, \xi_n)$ a random vector in $\mathbf{R}^n$ with spherically symmetric density of the form $g(\|\cdot\|^2)$ with*

$$g(t) = \frac{\Gamma(n/2)}{(\pi)^{n/2}(2\sigma^2)^{n/2+q}\,\Gamma(n/2+q)}\,t^q\,e^{-t/2\sigma^2}\,, \tag{48}$$

where q is a real number such that $n/2 + q > 0$. Then for any $0 < \mu_1 < 1/2\sigma^2$

$$\sup_{1\le k\le n}\,(\ln\,\mathbf{E}\,\mathrm{e}^{\mu_1\sum_{j=1}^k \xi_j^2} - \mu_0\,k) \le\, \ln\,M^*\,,$$

where $\mu_0 = -\frac{1}{2}\ln(1 - 2\sigma^2\mu_1)$ and $M^* = \mathrm{e}^{2q\mu_0}$.

Proof First note that, for $k = n$,

$$\begin{aligned}
\mathbf{E}_S\, e^{\mu_1 \sum_{j=1}^{n} \xi_j^2} &= \int_{\mathbf{R}^n} e^{\mu_1 \sum_{j=1}^{n} x_j^2}\, g\left(\sum_{j=1}^{n} x_j^2\right) \mathrm{d}x_1, \ldots, \mathrm{d}x_n \\
&= \int_0^\infty e^{\mu_1 r}\, \frac{2\pi^{n/2}}{\Gamma(n/2)}\, r^{n-1}\, g(r^2)\, \mathrm{d}r \\
&= \frac{1}{2^{n/2+q}\,\Gamma(n/2+q)} \int_0^\infty t^{n/2+q-1}\, e^{-\frac{1}{2\sigma^2}(1-2\sigma^2\mu_1)t}\, \mathrm{d}t \\
&= \frac{1}{(1-2\sigma^2\mu_1)^{n/2+q}} = e^{\mu_0\, n+2q\mu_0}\,.
\end{aligned}$$

Now, for $1 \le k \le n$, we can write, through conditioning on $\|\xi\| = r$,

$$\begin{aligned}
\mathbf{E}_S\, e^{\mu_1 \sum_{j=1}^{k} \xi_j^2} &= \int_{\mathbf{R}^n} e^{\mu_1 \sum_{j=1}^{k} x_j^2}\, g\left(\sum_{j=1}^{n} x_j^2\right) \mathrm{d}x_1, \ldots, \mathrm{d}x_n \\
&= \frac{2\pi^{n/2}}{\Gamma(n/2)} \int_0^\infty I_r\, r^{n-1}\, g(r^2)\, \mathrm{d}r
\end{aligned}$$

with

$$I_r = \int_{\mathcal{S}_r} e^{\mu_1 \sum_{j=1}^{k} x_j^2}\, \overline{\varrho_r}\, \mathrm{d}x \tag{49}$$

and where $\overline{\varrho_r}(\cdot)$ is the uniform distribution on the sphere

$$\mathcal{S}_r = \{x \in \mathbf{R}^n \,:\, \|x\| = r\}\,,$$

i.e. $\overline{\varrho_r}(\cdot) = \varrho_r(\cdot)/\varrho_r(\mathcal{S}_r)$ where $\varrho_r(\cdot)$ is the superficial measure on $\mathcal{S}_r$.

The integrand term in (49) depends of $\xi = (\xi_1, \ldots, \xi_n)$ through the orthogonal projection $\tilde{\xi}_k = (\xi_1, \ldots, \xi_k)$. Now, under $\overline{\varrho_r}$, this projection has the density

$$\frac{\Gamma(n/2)\, r^{2-n}}{\Gamma((n-k)/2)\pi^{k/2}} \left(r^2 - \sum_{j=1}^{k} x_j^2\right)^{(n-k)/2-1} \mathbf{1}_{B_r}\,(x_1, \ldots, x_k)\,,$$

where $\mathbf{1}_{B_r}$ denotes indicator function of the ball $B_r = \{x \in \mathbf{R}^k \,:\, \|x\| \le r\}$ of radius r and centered at 0 in $\mathbf{R}^n$ (see Fourdrinier and Strawderman (1996) for more references). Hence the integral in (49) can be written as

$$I_r = \frac{\Gamma(n/2)\, r^{2-n}}{\Gamma((n-k)/2)\pi^{k/2}} \int_{B_r} \mathrm{e}^{\mu_1 \sum_{j=1}^{k} x_j^2} \left(r^2 - \sum_{j=1}^{k} x_j^2 \right)^{(n-k)/2-1} \mathrm{d}x_1, \ldots, \mathrm{d}x_k$$

$$= \frac{2\Gamma(n/2)\, r^{2-n}}{\Gamma((n-k)/2)\Gamma(k/2)} \int_0^r \mathrm{e}^{\mu_1 u^2} (r^2 - u^2)^{(n-k)/2-1} u^{k-1}\, \mathrm{d}u$$

$$= \frac{\Gamma(n/2)}{\Gamma((n-k)/2)\Gamma(k/2)} \int_0^1 \mathrm{e}^{\mu_1 r^2 t} (1-t)^{(n-k)/2-1} t^{k/2-1}\, \mathrm{d}t$$

$$= \int_0^1 \mathrm{e}^{\mu_1 r^2 t}\, \nu_{k,n}(\mathrm{d}t)\,, \tag{50}$$

where $\nu_{k,n}(\cdot)$ is the Beta distribution with parameters $k/2$ and $(n-k)/2$.

Therefore it follows from (50) that

$$\mathbf{E}_S\, \mathrm{e}^{\mu_1 \sum_{j=1}^{k} \xi_j^2} = \int_0^1 J(t)\, \nu_{k,n} \mathrm{d}t \tag{51}$$

where

$$J(t) = \frac{2\pi^{n/2}}{\Gamma(n/2)} \int_0^\infty \mathrm{e}^{\mu_1 r^2 t}\, r^{n-1}\, g(r^2)\, \mathrm{d}r$$

$$= \frac{2}{(2\sigma^2)^{n/2+q}\, \Gamma(n/2+q)} \int_0^\infty r^{2q+n-1}\, \mathrm{e}^{-(1-2\mu_1\sigma^2 t)r^2/2\sigma^2}\, \mathrm{d}r$$

$$= \frac{1}{(2\sigma^2)^{n/2+q}\, \Gamma(n/2+q)} \int_0^\infty v^{n/2+q-1}\, \mathrm{e}^{-(1-2\mu_1\sigma^2 t)v/2\sigma^2}\, \mathrm{d}v$$

$$= \frac{1}{(1-2\mu_1\sigma^2 t)^{n/2+q}}\,.$$

Finally (51) gives that

$$\mathbf{E}_S\, \mathrm{e}^{\mu_1 \sum_{j=1}^{k} \xi_j^2} = \int_0^1 \frac{1}{(1-2\mu_1\sigma^2 t)^{n/2+q}}\, \nu_{k,n} \mathrm{d}t$$

$$\leq \frac{1}{(1-2\mu_1\sigma^2)^q} \int_0^1 \frac{1}{(1-2\mu_1\sigma^2 t)^{n/2}}\, \mathrm{d}\nu_{k,n}(t)$$

$$= \frac{1}{(1-2\mu_1\sigma^2)^{q+k/2}} = \mathrm{e}^{\mu_0\, k + 2q\mu_0}$$

since the last integral corresponds to the case $q = 0$ (that is the normal case) and equals $(1 - 2\mu_1\sigma^2)^{-k/2}$. Hence we obtain the desired result. $\square$

Notice that this lemma implies the condition $\mathcal{H}_1)$ by the Markov inequality.

8.3 Some properties of spherically symmetric distributions

Lemma 3 *Consider a spherically symmetric distribution on $\mathbf{R}^n$ with generating function g. Let $\mathcal{D}$ be a linear subspace of $\mathbf{R}^n$ with dimension d such that $0 < d < n$ and let $\hat{S}$ be the orthogonal projector from $\mathbf{R}^n$ onto $\mathcal{D}$. Then the distribution of $\hat{S}$ is spherically symmetric on $\mathcal{D}$ with generating function g_d defined, for any $w \in \mathbf{R}_+$, by*

$$g_d(w) = \frac{\pi^{(n-d)/2}}{\Gamma((n-d)/2)} \int_0^\infty v^{(n-d)/2-1} g(w+v)\, \mathrm{d}v\,.$$

Proof The density of the spherically symmetric distribution on $\mathbf{R}^n$ is the form $x \longmapsto g(\|x-\theta\|^2)$ for some $\theta \in \mathbf{R}^n$. It is clear that the proof reduces to the case where $\theta = 0$ and to consider that $\mathcal{D}$ is isomorphic to $\mathbf{R}^d$ so that the density of $\hat{S}$ can be written as

$$y \longmapsto \int_{\mathbf{R}^{n-d}} g(\|y\|^2 + \|z\|^2)\mathrm{d}z\ .$$

Thus the distribution of $\hat{S}$ is spherically symmetric (around 0) and has the generating function g_d defined, for any $w \geq 0$, by

$$\begin{aligned} g_d(w) &= \int_{\mathbf{R}^{n-d}} g(w + \|z\|^2)\mathrm{d}z \\ &= \frac{2\pi^{(n-d)/2}}{\Gamma((n-d)/2)} \int_0^\infty g(w+r^2)\, r^{n-d-1}\mathrm{d}z \end{aligned}$$

through polar coordinates. With the change of variable $v = r^2$, we obtain the stated result. $\square$

Lemma 4 *In the context of Lemma 3, consider the functions G and G_d defined, for any $w \geq 0$, by*

$$G(w) = \frac{1}{2}\, \frac{\int_w^\infty g(v)dv}{g(w)} \quad \textit{and} \quad G_d(w) = \frac{1}{2}\, \frac{\int_w^\infty g_d(v)dv}{g_d(w)}\,.$$

If there exists a constant $c > 0$ such that, for any $w \geq 0$, $G(w) \geq c$ then $G_d(w) \geq c$.

Proof Let $w \geq 0$. According to Lemma 20, we have

$$2\, G_d(w) \int_0^\infty v^{(n-d)/2-1} g(w+v) \mathrm{d}v = \int_w^\infty \int_0^\infty v^{(n-d)/2-1} g(u+v) \mathrm{d}v \mathrm{d}u$$

$$= \int_0^\infty \int_w^\infty g(u+v) \mathrm{d}u \;\; v^{(n-d)/2-1} \mathrm{d}v$$

$$= \int_0^\infty \int_{w+v}^\infty g(t) \mathrm{d}t \;\; v^{(n-d)/2-1} \mathrm{d}v \, .$$

Therefore

$$G_d(w) = \int_0^\infty \frac{1}{2} \, \frac{\int_{w+v}^\infty g(t) \mathrm{d}t}{g(w+v)} \, \frac{v^{(n-d)/2-1} g(w+v)}{\int_0^\infty z^{(n-d)/2-1} g(w+z) dz} \, \mathrm{d}v$$

which can be written as

$$G_d(w) \;=\; \mathbf{E}_w \, G(V+w) \, ,$$

where $\mathbf{E}_w$ denotes the expectation with respect to the density

$$v \longmapsto \frac{v^{(n-d)/2-1} g(v+w)}{\int_0^\infty z^{(n-d)/2-1} g(z+w) \mathrm{d}z} \; .$$

Thus, if $G(w) \geq c$ for any $w \geq 0$, then $G_d(w) \geq c$ as well. □

8.4 Proof of condition (28)

Notice that, to check (28), it suffices to show that, for any $d_m \geq 3$,

$$\mathbf{E}_S \, \|\hat{S}_m\|^{-2} \; < \infty \, .$$

Indeed, we have that

$$\begin{aligned}
\mathbf{E}_S \, \|\hat{S}_m\|^{-2} \; &= \int_{\mathbf{R}^{d_m}} \frac{1}{\|u+S_m\|^2} \, g_{d_m}(u) \, \mathrm{d}u \\
&= \int_{\|u+S_m\| \leq 1} \frac{g_{d_m}(u)}{\|u+S_m\|^2} \mathrm{d}u + \int_{\|u+S_m\| \geq 1} \frac{g_{d_m}(u)}{\|u+S_m\|^2} \mathrm{d}u \\
&\leq \sup_{\|u+S_m\| \leq 1} g_{d_m}(u) \int_{\|u+S_m\| \leq 1} \frac{1}{\|u+S_m\|^2} \, \mathrm{d}u + \int_{\mathbf{R}^{d_m}} g_{d_m}(u) \mathrm{d}u \\
&= \sup_{\|u+S_m\| \leq 1} g_{d_m}(u) \int_{\|u\| \leq 1} \frac{1}{\|u\|^2} \, \mathrm{d}u + 1 \, .
\end{aligned}$$

By using polar coordinates in the last integral, we obtain the desired result. □

8.5 Some properties of the Fourier coefficients

The following lemma is inspired from Timan (1963) (p. 254). We give a proof which underlines the link between the upper bound in (53) and the regularity condition expressed in (52).

Lemma 5 *Let f be a function in $\mathcal{C}^1[0,1]$ such that $f(0)=f(1)$, $f'(0)=f'(1)$ and, for some constants $L_0>0$, $L>0$ and $0\le\alpha<1$,*

$$\sup_{0\le x\le 1}|f(x)|\le L_0 \quad \textit{and} \quad |\dot{f}(x)-\dot{f}(y)|\le L\,|x-y|^{\alpha} \tag{52}$$

for all $x\,,y\in[0,1]$. Then the Fourier coefficients $(a_k)_{k\ge0}$ and $(b_k)_{k\ge1}$ of the function f, defined through

$$f(x)=\frac{a_0}{2}+\sum_{k=1}^{\infty}(a_k\cos(2\pi\,kx)+b_k\sin(2\pi\,kx))\,,$$

satisfy the following inequality

$$\sup_{n\ge0}(n+1)^{\beta}\left(\sum_{k=n}^{\infty}(a_k)^2+(b_k)^2\right)^{1/2}\le c^*\,(L+L_0) \tag{53}$$

where $\beta=1+\alpha$ and

$$c^*=1+2^{\beta}+\pi^4\,9^{\beta}\,\frac{\int_0^{\infty}u^{\alpha-3}\,\sin^4(\pi\,u)\,\mathrm{d}u}{8\,\int_0^{1/2}u^{-4}\,\sin^4(\pi\,u)\,\mathrm{d}u}\,.$$

Proof Let $F(x)$ be a 1-periodic function on $\mathbf{R}$ such that $F(x)=f(x)$ for $x\in[0,1]$. Note that we can also represent this function as

$$F(x)=\frac{a_0}{2}+\sum_{k=1}^{\infty}\left(\tilde{a}_k\cos(\pi\,kx)+\tilde{b}_k\sin(\pi\,kx)\right)\,,$$

where $\tilde{a}_k=a_m$ and $\tilde{b}_k=b_m$ if $k=2m$ and $\tilde{a}_k=\tilde{b}_k=0$ for $k=2m+1$, $m\ge0$. Then, by denoting $p_n(x)=\frac{a_0}{2}+\sum_{k=1}^{n}(a_k\cos(2\pi\,kx)+b_k\sin(2\pi\,kx))$, we obtain that, for $n\ge2$,

$$\begin{aligned}\sum_{k=n}^{\infty}(a_k^2+b_k^2)&=\int_{-1/2}^{1/2}(F(x)-p_{n-1}(x))^2\,\mathrm{d}x\\&=\frac{1}{2}\inf_{\alpha_j,\beta_j}\int_{-1}^{1}\left(F(x)-\frac{\alpha_0}{2}-\sum_{k=1}^{2(n-1)}(\alpha_k\cos(\pi\,kx)+\beta_k\sin(\pi\,kx))\right)^2\,\mathrm{d}x\\&\le\frac{1}{2}\int_{-1}^{1}(F(x)-t_m(x))^2\,\mathrm{d}x\,,\end{aligned}$$

where

$$t_m(x) = 2\int_{-1/2}^{1/2} F(u)\rho_m(u-x)\mathrm{d}u - \frac{1}{2}\int_{-1}^{1} F(u)\rho_m\left(\frac{u-x}{2}\right)\mathrm{d}u\,,$$

$$\rho_m(u) = \frac{1}{\gamma_m}\left(\frac{1}{2}+\sum_{j=1}^{m}\cos(2\pi\, ju)\right)^4 = \frac{1}{\gamma_m}\left(\frac{\sin((2m+1)\pi\, u)}{2\sin(\pi\, u)}\right)^4\,,$$

$$\gamma_m = \int_{-1/2}^{1/2}\left(\frac{\sin((2m+1)\pi\, u)}{2\sin(\pi\, u)}\right)^4\mathrm{d}u$$

and $m = [n/8]$. Here $[a]$ denote a integer part of any number a. We rewrite the term $t_m(x)$ as

$$t_m(x) = \int_{-1/2}^{1/2}(2F(u+x) - F(x+2u))\rho_m(u)\,\mathrm{d}u\,.$$

Note that the condition $f'(0) = f'(1)$ implies that the function F is continuously differentiable and satisfies the following inequality

$$|F(x+2u) - 2F(x+u) + F(x)| \le 2\,L\,|u|^{\beta}$$

for all $u\,, x\;\in\;\mathbf{R}$. Hence, taking into account that $2\,u\;\le\;\sin(\pi u)\;\le\;\pi\,u$ for $0 \le u \le 1/2$, we obtain that

$$\begin{aligned}|t_m(x) - F(x)| &\le \int_{-1/2}^{1/2}|F(x+2u) - 2F(x+u) + F(x)|\rho_m(u)\mathrm{d}u\\ &\le \frac{L}{4\gamma_m}\int_{0}^{1/2}u^{\beta}\left(\frac{\sin((2m+1)\pi\, u)}{\sin(\pi\, u)}\right)^4\mathrm{d}u\\ &\le \frac{L}{64\gamma_m}\int_{0}^{1/2}u^{\beta}\left(\frac{\sin((2m+1)\pi\, u)}{u}\right)^4\mathrm{d}u\,.\end{aligned}$$

By a change of variable in the last integral and setting

$$\theta_1 = \int_{0}^{\infty}u^{\beta}\left(\frac{\sin(\pi\, u)}{u}\right)^4\mathrm{d}u\,,$$

we obtain

$$|t_m(x) - F(x)| \le \frac{L(2m+1)^3\,\theta_1}{64\,\gamma_m}\left(\frac{1}{2m+1}\right)^{\beta}\,.$$

Furthermore, for $m \geq 0$,

$$\gamma_m = \int_{-1/2}^{1/2} \left(\frac{\sin((2m+1)\pi\, u)}{2\sin(\pi\, u)} \right)^4 \mathrm{d}u$$
$$\geq \frac{1}{8\pi^4} \int_0^{1/2} \left(\frac{\sin((2m+1)\pi\, u)}{u} \right)^4 \mathrm{d}u$$
$$\geq \frac{(2m+1)^3}{8\pi^4}\, \theta_2$$

where

$$\theta_2 = \int_0^{1/2} \left(\frac{\sin(\pi\, u)}{u} \right)^4 \mathrm{d}u\,.$$

The last inequality implies that

$$\sup_{x\in\mathbf{R}} |t_m(x) - F(x)| \leq \rho^*\, L \left(\frac{1}{2m+1} \right)^{\beta} \quad \text{with} \quad \rho^* = \frac{\pi^4 \theta_1}{8\theta_2}\,.$$

Taking into account that

$$\sup_{n\geq 2} \frac{n+1}{2m+1} = \sup_{n\geq 2} \frac{n+1}{2[n/8]+1} \leq 9\,,$$

we obtain that

$$\sup_{n\geq 2} (n+1)^{\beta} \left(\sum_{k=n}^{\infty} (a_k)^2 + (b_k)^2 \right)^{\frac{1}{2}} \leq 9^{\beta}\, \rho^*\, L\,.$$

Moreover, by the Parseval equality and the condition of this lemma, we have that

$$\left(\sum_{k=0}^{\infty} (a_k)^2 + (b_k)^2 \right)^{\frac{1}{2}} = \left(\int_0^1 f^2(x)\,\mathrm{d}x \right)^{\frac{1}{2}} \leq L_0\,.$$

Finally we obtain (53).

Now we extend this lemma to a k time differential function with $k \geq 1$.

Lemma 6 *Let S be a function in $\mathcal{C}^k[0,1]$ such that $S^{(j)}(0) = S^{(j)}(1)$ for all $0 \leq j \leq k$ and, such that, for some contants $L_0 > 0$, $L > 0$ and $0 \leq \alpha < 1$*

$$\max_{0\leq j\leq k-1} \max_{0\leq x\leq 1} |S^{(j)}(x)| \leq L_0 \;\; \textit{and} \;\; |S^{(k)}(x) - S^{(k)}(y)| \leq L\,|x-y|^{\alpha} \qquad (54)$$

for all $x, y \in [0,1]$. *Then the Fourier coefficients* $(a_k)_{k\geq 0}$ *and* $(b_k)_{k\geq 1}$ *of the function* S *satisfies the inequality*

$$\sup_{n\geq 0}(n+1)^{\beta}\left(\sum_{j=n}^{\infty}(a_j)^2+(b_j)^2\right)^{\frac{1}{2}} \leq c^*\,(L+L_0)\,, \tag{55}$$

where $\beta = k+\alpha$ *(*k *being an integer and* $0\leq\alpha<1$*) and* c^* *is defined in (53).*

Proof For $k=1$, Inequality (55) is shown in Lemma 3. For $k\geq 2$, first, note that, by Lemma 3, the Fourier coeficients (a'_j) and (b'_j) of the function $f = S^{(k-1)}$ satisfy Inequality (53). Therefore, taking into account that

$$(a'_j)^2+(b'_j)^2 = \left(4\pi^2 j^2\right)^{k-1}\left((a_j)^2+(b_j)^2\right)\,,$$

we obtain Inequality (55). □

References

1. Barron, A., Birgé, L., Massart, P. (1999). Risk bounds for model selection via penalization. *Probability Theory and Related Fields*, *113*, 301–413
2. Beran, J. (1992). Statistical methods for data with long-range depedence. *Statistical Science*, *7*, 404–427
3. Berger, J. O. (1975). Minimax estimation of location vectors for a wide class of densities. *The Annals of Statistics*, *3*(6), p. 1318–1328
4. Blanchard, D., Fourdrinier, D. (1999). Non trivial solutions of non-linear partial differential inequations and order cut-off. *Rendiconti di Matematica*, Serie VII, *19*, 137–154
5. Brown, L. D. (1966). On the admissibility of invariant estimators of one or more location parameters. *The Annals of Mathematical Statistics*, *37*, 1087–1136
6. Csörgó, S., Mielniczuk, J. (1995). Nonparametric regression under longe range dependent normal errors. *The Annals of Statistics*, *23*(3), 1000–1014
7. Dahlhaus, R. (1995). Efficient location and regression estimation for longrange dependent regression models. *The Annals of Statistics*, *23*(3), 1029–1047
8. Donoho, D.L., Johnstone, I.M., Kerkyacharian, G. Picard, D. (1995). Wavelet shrinkage: asymptotia? *Journal of Royal Statistical Society, Services B*, *57*, 301–369
9. Efroimovich, S. Yu. (1999). *Nonparametric curve estimation. Methods, theory and applications*. Berlin, Haidelberg New York: Springer
10. Fang, K.-T., Kotz, S., Ng, K.-W. (1989). *Symmetric Multivariate and Related Distributions*. New york: Chapman and Hall
11. Fourdrinier, D., Strawderman, W. E. (1996). A paradox concerning shrinkage estimators: should a known scale parameter be replaced by an estimated value in the shrinkage factor? *Journal of Multivariate Analysis*, *59*(2), 109–140
12. Fourdrinier, D., Wells, M. T. (1994). Comparaison de procédures de sélection d'un modèle de régression: une approche décisionnelle. *Comptes Rendus de l'Académie des Sciences Paris*, v. 319, série I, p. 865–870
13. Gunst, R. F., Mason, R.L. (1980). *Regression analysis and its applications: a data oriented approach*. Marcel Dekker: New York
14. James, W., Stein, C. (1961). Estimation with quadratic loss. *In Proceedings of the Fourth Berkeley Symposium Mathematics, Statistics and Probability*, University of California Press, Berkeley, *1*, 361–380
15. Hall, P., Hart, J. (1990). Nonparametric regression with long range dependence. *Stochastics Processes and their Applications 36*, 339–351
16. Kariya, T., Sinha, B.K. (1993). *The robustness of statistical tests*. New York: Academic
17. Kotz, S. (1975). Multivariate distributions at a cross-road. In Patil, G.J., Kotz, S., Ord, J.K. Statistical Distributions in Scientific Work. (Eds.) *1*, 245–270

18. Nemirovski, A. (2000). Topics in non-parametric statistics. *Lectures on probability theory and statistics, Saint-Flour 1998. In Lecture Notes in Mathematics*, *1738*. Berlin Heidelberg New York: Springer, pp 85–277
19. Stein, C.M. (1956). Inadmissibility of usual estimator of the mean of a multivariate distribution. *In Proceedings of the 3rd Berkely Symposium in Mathematical Statistics and Probability*, *1*, 197–206
20. Stein, C.M. (1981). Estimation of the mean of a multivariate normal distribution. *The Annals of Statistics*, *9*, 1135–1151
21. Timan, A. F. (1963). *Theory of approximation of functions of a real variable.* Oxford: Pergamon
22. Tsybakov, A. (1998). Pointwise and sup-norm sharp adaptive estimation of functions on the Sobolev classes. *The Annals of Statistics*, *26*(6), 221–245
23. Ziemer, W.P. (1989). *Weakly differentiable functions — sobolev spaces and functions of bounded variation*. Berlin Heidelberg New York: Springer