Confidence balls in nonparametric Gaussian regression

Y. BARAUD

DMA - 01 - 28
Confidence balls in nonparametric Gaussian regression

Y. BARAUD*

DMA - 01 - 28

November 2001
CONFIDENCE BALLS IN NONPARAMETRIC GAUSSIAN REGRESSION

YANNICK BARAUD

Abstract. Starting from the observation of an $\mathbb{R}^n$-Gaussian vector of mean $f$ and covariance matrix $\sigma^2 I$ ($I$ is the identity matrix), we propose two methods for building an Euclidean confidence ball around $f$, with prescribed probability of coverage. The first method is based on prior information on the variance $\sigma^2$ and is free from any assumption on $f$. In contrast, the other method is based on prior information on the ratio signal-noise, $f/\sigma$. For each $n$, we describe the nonasymptotic properties of the so-defined confidence balls and show their optimality with respect to some criteria.

1. Introduction

We consider the following statistical model

$$Y_i = f_i + \sigma \varepsilon_i, \ i = 1, \ldots, n$$

(1)

where $f = (f_1, \ldots, f_n)'$ is an unknown vector, $\sigma$ a positive number and $\varepsilon_1, \ldots, \varepsilon_n$ a sequence of i.i.d. standard Gaussian random variables. The aim of this paper is to build a nonasymptotic Euclidean confidence ball for $f$ with prescribed probability of coverage from the observation of $Y = (Y_1, \ldots, Y_n)'$.

The ideas underlying our approach are due to Lepski and have been exposed by their initiator in a series of lectures at the Institute Henri Poincaré in Paris. We shall now give a brief account of these ideas and recommend that the reader have a look at Lepski (1999) for more details.

Let us consider a statistical experiment generating the observation $X^n$ of law $\mathbb{P}_F$ where $F$ is an unknown function belonging to some known functional space $\Sigma$. Under suitable assumptions on the statistical model, a minimax approach allows to obtain both an estimator of

\begin{footnotesize}
\begin{itemize}
\item 1991 Mathematics Subject Classification. Primary 62G15; Secondary 62G03,62G10.
\item Key words and phrases. Confidence ball, nonparametric regression, hypothesis testing, estimation.
\end{itemize}
\end{footnotesize}
and a control on the accuracy of the estimation. Yet, unless one has a strong guess on the particular features of $F$, $\Sigma$ is usually too large to obtain an accurate estimation of $F$. The idea of Lepski is to test one or several additional structures on $F$ in order to improve the accuracy of estimation. Unlike an adaptive approach, an attractive feature of Lepski’s lies in the fact that the accuracy of estimation is available to the statistician and consequently that a nonparametric confidence ball for $F$ can be derived. This is explained in the papers by Lepski (1999) and by Hoffmann and Lepski (2001).

However, the procedure described there for the purpose of building $\mathbb{L}^2$-confidence balls suffers from the following weaknesses. First, the point of view is purely asymptotic. The procedure does not lead to confidence balls with prescribed probability of coverage, say $1 - \beta$, for fixed values of $n$. Furthermore, a careful look at the proofs shows that for fixed $n$, the squared radius of the confidence ball is equal to a constant plus some term which essentially proportional to the number of hypotheses to test. Consequently, the number of these cannot be large if one wants to keep the confidence ball of a reasonable size. In addition, the squared radius of the confidence ball is proportional to $1/\beta$ and is thus very large for small values of $\beta$. Finally, the applications developed in Lepski (1999) and Hoffmann and Lepski (2001) mainly address the Gaussian white noise model and an adaptation of the procedure to the regression case would require an estimation of the unknown $\sigma$.

The results of the present paper are nonasymptotic and the procedures which are described here aim at obtaining confidence balls as sharp as possible. In particular, the dependency with respect to $\beta$ and the number of hypotheses to test is only logarithmic. This allows us to handle the variable selection problem described in Section 2.2 below. Besides, the problem of building a confidence ball when $\sigma$ is unknown is also considered. We show that without any prior information on the target vector $f$, the problem of building a non trivial confidence ball for $f$ is impossible unless $\sigma$ is known or nearly known. In the case of a known $\sigma$, our result recovers those previously established by Li (1989). Namely, Li showed that if asymptotically a confidence ball ensures a probability of coverage larger than $1 - \beta$ whatever $f \in \mathbb{R}^n$ then necessarily for each $f \in \mathbb{R}^n$, the radius of the confidence ball cannot converge towards 0 faster than $n^{1/4}$. 
When $\sigma$ is known or nearly known, we propose a procedure to build a nonasymptotic confidence ball which does not require any prior information on the target vector $f$. For a known $\sigma$, this problem was also considered by Li (1989), and an asymptotic Euclidean confidence ball based on Stein estimates can be found there.

Under the prior information that the ratio $f/\sigma$ belongs to some set $\Sigma \subset \mathbb{R}^n$, we consider the problem of building a confidence ball when $\sigma$ is totally unknown. An application to the cases where $\Sigma$ is an ellipsoid or is related to an Hölderian functional class is developed.

Let us now briefly describe our procedure. We start with a collection of linear spaces $\{S_m, m \in \mathcal{M}_n\}$ and associate to each of these, the least-squares estimator of $f$, denoted by $\hat{f}_m$, a test $\phi_m$ to test the null hypothesis “$f \in S_m” against the alternative “$f \in \mathbb{R}^n \setminus S_m” and some positive number $\rho_m$. The $\rho_m$’s are suitably calibrated to satisfy the property that if $\phi_m$ accepts the null then with probability close to one, the Euclidean distance between $f$ and its least-squares estimator $\hat{f}_m$ is not larger than $\rho_m$. We then select $\hat{m}$ as the minimizer of $\rho_m$ among those $m$ for which $\phi_m$ accepts the null and define the confidence ball as the Euclidean ball centered at $\hat{f}_{\hat{m}}$ of radius $\rho_{\hat{m}}$.

The optimality (in a suitable sense) of the so-defined procedure is established. The proof relies on lower bounds for the minimax estimation rates and for the minimax separation rates over linear spaces. These lower bounds hold true for fixed values of $n$ and involve explicit numerical constants.

The problem of giving a confidence “ball” in sup-norm has been considered by several authors, the more recent work being, to our knowledge, due to Picard and Tribouley (2000) and we refer to the references therein for an overview on the topic.

The paper is organized as follows. In Section 2 we consider the case of a known $\sigma$. In Section 3, we assume that $\sigma$ is unknown but belongs to some known interval. The case of a totally unknown $\sigma$ with some prior information on the ratio $f/\sigma$ is treated in Section 4. The proofs are postponed to Section 5.

Notations: Throughout this paper we use the following notations. We equip $\mathbb{R}^n$ with the Euclidean norm denoted by $\|\|$. For a triplet $(z, d, u) \in \mathbb{R}_+ \times \mathbb{N} \setminus \{0\} \times [0, 1]$ and a random variable $X$ distributed as a (non)central $\chi^2$ with noncentrality parameter $z$ and $d$ degrees of freedom, we denote by $\chi^2_{z, d}(u)$ the distribution function of $X$ and by
$q_{z,d}(u)$ its 1 − $u$ quantile. In particular

$$\mathbb{E}[X] = z + d, \quad \chi^2_{z,d}(u) = \mathbb{P}(X \leq u) \quad \text{and} \quad \mathbb{P}(X \geq q_{z,d}(u)) = u.$$ 

We use the convention that

$$q_{z,0}(u) = 0 \quad \text{and} \quad q_{z,d}(1) = -\infty.$$ 

We denote by $\Pi_S$ the orthogonal projector onto the linear space $S \subset \mathbb{R}^n$ and by $B(x,r)$, the Euclidean ball centered at $x \in \mathbb{R}^n$ of radius $r > 0$. Throughout the paper, $\alpha$ and $\beta$ denote two numbers in $]0, 1[$ satisfying $1 - \alpha > \beta$, and $\{S_m, \ m \in \mathcal{M}_n\}$ denotes a collection of linear subspaces of $\mathbb{R}^n$. For each $m$, $D_m$ is the dimension of $S_m$ and $N_m$ the quantity $n - D_m$. Finally, $C, C'$... denote constants that may vary from line to line.

2. Confidence balls when the variance is known

This section is devoted to the case of a known $\sigma$. The construction of the confidence ball for $f$ is given below. Note that it is assumption free on $f$.

2.1. Construction of the confidence ball. To each $m \in \mathcal{M}_n$, we associate to $S_m$ some positive number $\beta_m$ and assume that the following assumption is fulfilled.

**Assumption 2.1.** The subscript $n$ belongs to $\mathcal{M}_n$ and $S_n = \mathbb{R}^n$. We have $\sum_{m \in \mathcal{M}_n} \beta_m \leq \beta$ and for each $m \in \mathcal{M}_n \setminus \{n\}$, $D_m \leq n/2$.

We define the sequence of nonnegative numbers $\{\rho_m, \ m \in \mathcal{M}_n\}$ as follows. If $m = n$ then

$$\rho^2_m = q_{0,n}(\beta_n)\sigma^2,$$

if $m \in \mathcal{M}_n \setminus \{n\}$ and $D_m \neq 0$ then

$$\rho^2_m = \sup_{z \geq 0} \left[ z + q_{0,D_m} \left( \beta_m \left( \frac{\chi^2_{z,N_m}(q_{0,N_m}(\alpha))}{\chi^2_{z,N_m}(q_{0,N_m}(\alpha)) \wedge 1} \right) \right) \right] \sigma^2,$$ \hfill (2)

if $m \in \mathcal{M}_n \setminus \{n\}$ and $D_m = 0$ then

$$\rho^2_m = \inf \{ z \geq 0, \chi^2_{z,n}(q_{0,n}(\alpha)) \leq \beta_m \} \sigma.$$

For each $m \in \mathcal{M}_n$, we set $\hat{f}_m = \Pi_{S_m} Y$ and test the hypothesis "$f \in S_m$" against the alternative "$f \not\in S_m$" by rejecting the null when the
quantity $\|Y - \hat{f}_m\|^2 - q_{0, N_m}(\alpha)\sigma^2$ is positive. We denote by $A$ the set of indices $m \in M_n$ for which the hypothesis is accepted, namely

$$A = \left\{ m \in M_n, \; \|Y - \hat{f}_m\|^2 \leq q_{0, N_m}(\alpha)\sigma^2 \right\},$$

and define

$$\hat{m} = \arg\min_{m \in A} \rho_{m}, \quad \hat{\rho} = \rho_{\hat{m}}, \quad \hat{f} = \hat{f}_{\hat{m}}. \quad (3)$$

We have the following result.

**Theorem 2.1.** Let $(\hat{f}, \hat{\rho})$ be the pair of random variables defined by (3). If the mapping $m \mapsto \rho_m$ is one to one then $B(\hat{f}, \hat{\rho})$ is a confidence ball with probability of coverage $1 - \beta$, i.e.

$$P_{f, \sigma} \left[ f \in B(\hat{f}, \hat{\rho}) \right] \geq 1 - \beta, \; \forall f \in \mathbb{R}^n. \quad (4)$$

Moreover, for each $m \in M_n$

$$\inf_{f \in S_m} P_{f, \sigma} \left[ \hat{\rho} \leq \rho_m \right] \geq 1 - \alpha. \quad (5)$$

The assumption that $m \mapsto \rho_m$ is one to one is usually met. If not, a slight modification of the $\beta_m$’s allows to satisfy it.

By using the convention $q_{0,0}(\cdot) \equiv 0$ note that $n$ belongs to $A$ and therefore $A$ is nonvoid. Besides, the inequality $\hat{\rho} \leq \rho_n$ is always verified.

If $N_m \neq 0$, the set $Z_m$ collecting the $z$’s for which $\chi^2_{z, N_m}(q_{0, N_m}(\alpha)) > \beta_m$ is nonvoid. In fact $0 \in Z_m$ since

$$\chi^2_{0, N_m}(q_{0, N_m}(\alpha)) = 1 - \alpha > \beta \geq \beta_m.$$ 

Therefore, by using the convention that $q_{0, D_m}(1) = -\infty$, the supremum in (2) can be restricted to those $z$ in $Z_m$.

It follows from the proof that actually with probability larger than $1 - \beta$, $f$ belongs to the smaller set

$$\bigcap_{m \in A} B(\hat{f}_m, \rho_m),$$

which is not a ball in general.

For each $m \in M_n$, Inequality (5) shows that if $f$ belongs to $S_m$, the radius of the confidence ball is not larger than $\rho_m$ with probability larger than $1 - \alpha$. An upper bound for $\rho_m$ is given in the following proposition the proof of which provides explicit constants.
Proposition 2.1. There exists some constant $C$ depending on $\alpha$ only such that for all $m \in \mathcal{M}_n$

$$\rho_m^2 \leq C \max \{ D_m, \sqrt{n \log(1/\beta_m)}, \log(1/\beta_m) \} \sigma^2.$$ 

If $\mathcal{M}_n$ reduces to $\{n\}$ then $\hat{\rho} = \rho_n$ and the radius of the ball is of order $n\sigma^2$ by taking $\beta_n = \beta$. By considering several linear spaces $S_m$ we have the opportunity to capture some specific features of $f$ and consequently to reduce the order of magnitude of $\hat{\rho}$. The number of tests $|\mathcal{M}_n|$ to perform is taken into account via the quantity $\beta_m$. If one chooses $\beta_m = \beta/|\mathcal{M}_n|$ for all $m \in \mathcal{M}_n$, one gets that the radius of the confidence ball depends logarithmically on $|\mathcal{M}_n|$. However, a choice of $\beta_m$ depending on $m$ via the dimension of the linear space $S_m$ for example, is recommended. An example is given below.

2.2. Application to variable selection. In this section, we illustrate the procedure in the variable selection problem. Assume that $f$ is of the form $XU$ where $X$ is a known $p \times n$ full-rank matrix with $p \in \{1, \ldots, n\}$ and $U$ some unknown vector in $\mathbb{R}^p$. The problem of variable selection is to determine from the data the nonzero coordinates of $U$, that is

$$m^* = \{j \in \{1, \ldots, p\}, U_j \neq 0\}.$$

In this section we give a way to select those coefficients and provide simultaneously a confidence ball for $f$. We apply the procedure as follows:

Let $x_1, \ldots, x_p$ be the column vectors of the matrix $X$ and $\mathcal{P}_n$ the class of nonempty subsets $m$ of $\{1, \ldots, p\}$ with cardinality $|m|$ not larger than $n/2$. For all $m \in \mathcal{P}_n$, we define $S_m$ as the the linear span of the $x_j$’s for $j \in m$ and set

$$\beta_m = \beta \left[ n \binom{n}{D} \right]^{-1} \text{ with } D = |m|.$$ 

We define $\mathcal{M}_n = \mathcal{P}_n \cup \{n\}$ and set $\beta_n = \beta/2$. Note that Assumption 2.1 is fulfilled since

$$\sum_{m \in \mathcal{M}_n} \beta_m = \frac{\beta}{2} + \sum_{m \in \mathcal{P}_n} \beta_m$$

$$= \frac{\beta}{2} + \sum_{1 \leq D \leq n/2} \sum_{m \in \mathcal{P}_n, 1 \leq D \leq \beta} \beta_m \leq \beta.$$ 

By applying the procedure described in Section 2.1 we select a set of indices $\hat{m}$ for which the Euclidean distance between the least-squares
estimator \( \hat{f}_m \) and \( f \) is not larger than \( \rho_m \) with probability larger than \( 1 - \beta \). Since \( f \) belongs to the linear space \( S_{m^*} \), with probability larger than \( 1 - \alpha \) the set \( m^* \) belongs to \( \mathcal{A} \) and consequently \( \rho_m \) is not larger than \( \rho_{m^*} \). Therefore, either \( \hat{m} = m^* \) and then the procedure selects the target subset \( m^* \) or \( \hat{m} \neq m^* \) and then the resulting confidence ball is at least as accurate as if the target subset \( m^* \) were selected. In addition, thanks to the inequality

\[
(n \choose D) \leq \exp(D \log(e n / D))
\]

and Proposition 2.1, with probability larger than \( 1 - \alpha \), the following upper bound holds: there exists some constant \( C \) depending on \( \alpha \) and \( \beta \) only such that

\[
\hat{\rho}^2 \leq C \max \left\{ \sqrt{n|m^*| \log(e n / |m^*|)}, |m^*| \log(e n / |m^*|) \right\} \sigma^2.
\]

Let us denote by \( B \) this upper bound. By choosing \( \beta_m = \beta_n = \beta / |\mathcal{M}_n| \) for all \( m \in \mathcal{M}_n \), the bound on \( \hat{\rho}^2 \) would then be of order \( B' = \max \{ \sqrt{n n_p} \sigma^2 \} \) as \( |\mathcal{M}_n| \) is of order \( n^p \). Note that \( B \) and \( B' \) are of the same order when \( |m^*| \) is of order \( n \) but that \( B \) can be much smaller than \( B' \) when \( p \) is large (say of order \( n \)) and \( |m^*| \) small. This illustrates the advantage of taking \( \beta \) as a function of \( m \).

3. Confidence balls under some information on the variance

In this section, we no longer assume \( \sigma \) to be known but assume that there exists a pair of known numbers \( (\tau^2, \eta) \in \mathbb{R}_+ \times [0, 1] \) which satisfy

\[
(1 - \eta) \tau^2 \leq \sigma^2 \leq \tau^2.
\]

This section is only of theoretical interest since this assumption is seldom verified in practice. Its aim is to highlight the dependency of the radius of a confidence ball with respect to the parameter \( \eta \) which measures in some sense our reliability on \( \tau^2 \) as a “good” upper bound for \( \sigma^2 \). In the sequel we set \( I = [\sqrt{1 - \eta}, \tau] \).

3.1. How sharp can the confidence ball be? We have the following result.

**Theorem 3.1.** Let \( \alpha \) and \( \beta \) be numbers in \( [0, 1] \) satisfying \( 2 \beta + \alpha < 1 - \exp(-1/36) \). Let \((\hat{f}, \hat{r})\) be a pair of random variables with values in \( \mathbb{R}^n \times \mathbb{R}_+ \) verifying both
(i) for all \( f \in \mathbb{R}^n \) and \( \sigma \in I \),
\[
P_{f,\sigma} \left[ f \in \mathcal{B}(\hat{f}, \hat{r}) \right] \geq 1 - \beta \tag{7}
\]
(ii) for all \( m \in \mathcal{M}_n \) there exists some positive number \( r_m \) such that
\[
\inf_{f \in \mathcal{S}_m} P_{f,\sigma} [\hat{r} \leq r_m] \geq 1 - \alpha. \tag{8}
\]
Then there exists some constant \( C \) depending on \( \alpha \) and \( \beta \) only such that for all \( m \in \mathcal{M}_n \),
\[
\tau_m^2 \geq C \max \{ \eta N_m, D_m(1 - \eta), \sqrt{N_m(1 - \eta)} \} \tau^2. \tag{9}
\]

To keep our formula as legible as possible, the above theorem involves an inexplicit constant \( C \). However, lower bounds including explicit numerical constants are available from the proof in Section 5.2.

If \( D_m \leq n/2 \) then \( N_m \geq n/2 \) and we derive from the theorem that for some constant \( C \) depending on \( \alpha \) and \( \beta \) only
\[
\tau_m^2 \geq C \max \{ \eta n, D_m, \sqrt{n} \} \tau^2.
\]
Consequently, when \( \eta = 0 \) we obtain that from an asymptotic point of view (that is as \( n \) becomes large), this lower bound is of the same order as the upper bound on \( \rho_m^2 \) established in Proposition 2.1 provided that \( \beta_m \) is free from \( n \). This is the case if \( \beta_m = \beta / |\mathcal{M}_n| \) and if the cardinality of the collection, \( |\mathcal{M}_n| \), does not depend on \( n \). The procedure is optimal in the sense given by Lepski (1999).

Another consequence is that if \( f = 0 \), by applying the theorem with \( S_m = \{0\} \) we get that whatever the procedure satisfying (7) (these procedures are called honest by Li (1989)) the radius of the confidence ball is under \( P_0 \) at least of order \( \max \{ \eta n, \sqrt{n} \} \tau^2 \). Clearly, by translating the data \( Y \) of a vector \( f \in \mathbb{R}^n \), we deduce that the result is not only true for \( 0 \) but for all \( f \in \mathbb{R}^n \). This means that whatever \( f \in \mathbb{R}^n \) the accuracy of the confidence ball under \( P_f \) cannot be better (in order) than \( \max \{ \eta n, \sqrt{n} \} \tau^2 \) and consequently that the confidence ball is useless unless \( \eta \) is small. This result generalizes those obtained by Li (1989) for \( \eta = 0 \).

3.2. Construction of a confidence ball. In this section we build a confidence ball under the information that \( \sigma \) belongs to \( I \).

The following result holds.

**Theorem 3.2.** Let \( \sigma \in I \) and assume that Assumption 2.1 is fulfilled. Consider the construction of \((\hat{f}, \hat{r})\) described in Section 2.1 with the
following definitions for the \( \rho_m \)'s and \( A \): if \( m = n \) then

\[
\rho_n^2 = q_{0,n}(\beta_n) \tau^2,
\]

if \( m \in \mathcal{M}_n \setminus \{n\} \) and \( D_m \neq 0 \)

\[
\rho_m^2 = \sup_{z \geq 0, \sigma \in I} \left[ z \sigma^2 + q_{0,D_m} \left( \frac{\beta_m}{\chi_{2,N_m}(q_{0,N_m}(\alpha) \tau^2)} \wedge 1 \right) \sigma^2 \right]
\]

if \( m \in \mathcal{M}_n \setminus \{n\} \) and \( D_m = 0 \)

\[
\rho_m^2 = \inf \left\{ x \geq 0, \sup_{\sigma \in I} \chi_{2,\frac{m}{\sigma^2}}(q_{0,n}(\alpha) \tau^2 / \sigma^2) \leq \beta_m \right\}
\]

and

\[
A = \left\{ m \in \mathcal{M}_n, \| Y - \hat{f} \|^2 \leq q_{0,N_m}(\alpha) \tau^2 \right\}.
\]

If the mapping \( m \mapsto \rho_m \) is one to one then \( B(\hat{f}, \rho) \) is a confidence ball with probability of coverage \( 1 - \beta \) i.e. (4) is verified. Moreover, for each \( m \in \mathcal{M}_n \)

\[
\inf_{f \in S_m} \mathbb{P}_{f,\sigma} [ \rho \leq \rho_m ] \geq 1 - \alpha. \tag{10}
\]

An upper bound for \( \rho_m \) is given by the following proposition.

**Proposition 3.1.** There exists some constant \( C \) depending on \( \alpha \) only such that for all \( m \in \mathcal{M}_n \)

\[
\rho_m^2 \leq C \max \{ n, D_m, \sqrt{n \log(1/\beta_m)}, \log(1/\beta_m) \} \tau^2.
\]

From an asymptotic point of view, we derive from Theorem 3.1 the optimality of the procedure whenever the cardinality of the collection \( |\mathcal{M}_n| \) does not depend on \( n \) by taking \( \beta_m = \beta/|\mathcal{M}_n| \) for all \( m \in \mathcal{M}_n \). For more general collections, the procedure is also optimal for those \( m \in \mathcal{M}_n \) for which \( \beta_m \) does not decrease with \( n \).

4. **Confidence balls under some information on the ratio signal/noise**

In this section we consider that \( \sigma \) is unknown but we make an assumption on \( f/\sigma \). Let us set \( f_\sigma \) the vector \( f/\sigma \), we assume that there exists a linear space \( S \subset \mathbb{R}^n \) with \( \dim(S) = D \leq n/2 \) and a number \( b \) such that

\[
d(f_\sigma, S) = \inf_{s \in S} \| f_\sigma - s \| \leq b. \tag{11}
\]

The pair \((S, b)\) is assumed to be known. We denote by \( V_S(b) \) the set of vectors \( f \) in \( \mathbb{R}^n \) satisfying (11).
4.1. **Some examples.** Let us present here two examples where such information is available.

(Ell) We assume that \( f_{\sigma} \) belongs to the ellipsoid
\[
\mathcal{E}_{a, \epsilon} = \left\{ g \in \mathbb{R}^n, \sum_{i=1}^{n} \frac{g_i^2}{a_i^2} \leq \frac{1}{\epsilon^2} \right\},
\]
where \( a = (a_i)_{i=1, \ldots, n} \) is a nonincreasing sequence of positive numbers and \( \epsilon \) some positive number. We assume \( a \) and \( \epsilon \) to be known. For some \( D \in \{1, \ldots, [n/2]\} \) ([x] denotes the integer part of x), let 
\( S = S_D^* \) be the linear space generated by the \( D \) first elements of canonical basis of \( \mathbb{R}^n \). In this case \( d(f_{\sigma}, S_D^*) \leq a_D \epsilon^{-1} = b \).

(Höld) Let \( x_1, \ldots, x_n \) be deterministic points in \([0, 1]\). For some \( L > 0 \) and \( s \in [0, 1] \) let \( \mathcal{H}_s(L) \) be the Hölder space
\[
\mathcal{H}_s(L) = \{ G, |G(x) - G(y)| \leq L |x - y|^s \ \forall x, y \in [0, 1] \}.
\]
We assume that \( f_{\sigma} \) belongs to
\[
\Sigma_s(L) = \{ g = (G(x_1), \ldots, G(x_n))^t, G \in \mathcal{H}_s(L) \},
\]
s and \( L \) being known. For some \( D \in \{1, \ldots, [n/2]\} \), let \( S_D^* \) be the space generated by the piecewise constant functions on \([0, 1]\) based on the regular grid \( \{k/D, \ k = 0, \ldots, D\} \). Then, setting
\[
S = S_D^* = \{ (G(x_1), \ldots, G(x_n))^t, G \in S_D \},
\]
we have \( d(f_{\sigma}, S_D^*)/\sqrt{n} \leq L D^{-s} \). For a proof of this inequality, see Baraud, Huet and Laurent (2000). We shall assume that the \( x_i \)'s are such that \( \dim(S) = D \).

When the vector \( f \) is related to some continuous function \( F \) on a compact interval \( I \) by the relation \( f = (F(x_1), \ldots, F(x_n))^t \), it is more natural to consider a confidence ball with respect to the normalized Euclidean norm, \( \| \cdot \|_n = \| \cdot \|/\sqrt{n} \), which converges asymptotically, for an adequate repartition of the \( x_i \)'s, towards the \( \mathbb{L}^2([0, 1], dx) \)-norm. The radius \( \hat{\rho} \) of the confidence ball should then be renormalized by the factor \( \sqrt{n} \).

4.2. **Construction of the confidence ball.** Let \( \alpha, \beta, \gamma \) be numbers in \([0, 1/2]\), we set \( \tilde{D} = n - D \),
\[
\tau^2 = \frac{\| Y - \Pi_SY \|^2}{q_0, \tilde{D}(1 - \gamma)}, \tag{12}
\]
and

\[ \eta = \frac{q_{b^2,D}(\gamma) - q_{0,D}(1 - \gamma)}{q_{b^2,D}(\gamma)}. \]  

(13)

Note that \( \eta \) belongs to [0, 1]. Let \( v_1, ..., v_D \) be an orthonormal basis of \( \mathbb{R}^D \) and \( \mathcal{I} \) the isometry from \( S \) to \( \mathbb{R}^D \) defined by

\[ S \rightarrow \mathbb{R}^D \]

\[ s = \sum_{j=1}^{D} \bar{s}_j v_j \mapsto \mathcal{I}(s) = (\bar{s}_1, ..., \bar{s}_D)'. \]

We consider a collection of pairs \( \{(S_m, \beta_m), m \in \mathcal{M}_n\} \) such that the collection \( \{(\mathcal{I}(S_m), \beta_m), m \in \mathcal{M}_n\} \) satisfies Assumption 2.1 with \( D \) in place of \( n \). Given the data

\[ Z_j = < Y, v_j > \text{ for } j = 1, ..., D \]

we define \( \{\rho_m, m \in \mathcal{M}_n\}, \hat{f} \in \mathbb{R}^D \) and \( \hat{\rho} \) as in Section 3.2 with \( \tau^2 \) and \( \eta \) respectively defined by (12) and (13). For each \( m \in \mathcal{M}_n \), we set

\[ \tilde{\rho}_m^2 = \rho_m^2 + b^2\tau^2 \]

and define the pair \( (\hat{f}, \hat{\rho}) \) as

\[ \hat{f} = \mathcal{I}^{-1}(\tilde{f}) \quad \text{and} \quad \hat{\rho} = \tilde{\rho}^2 + b^2\tau^2. \]

(14)

We then have the following result.

**Theorem 4.1.** The pair of random variables \( (\hat{f}, \hat{\rho}) \) defined by (14) satisfies for all \( \sigma > 0 \) and \( f \in \mathcal{V}_S(b) \)

\[ \mathbb{P}_{f,\sigma} \left[ f \in \mathcal{B}(\hat{f}, \hat{\rho}) \right] \geq (1 - 2\gamma)(1 - \beta). \]

(15)

In addition, for each \( m \in \mathcal{M}_n \),

\[ \inf_{f \in S_m} \mathbb{P}_{f,\sigma} [\hat{\rho} \leq \tilde{\rho}_m] \geq (1 - 2\gamma)(1 - \alpha). \]

(16)

When \( v_1, ..., v_D \) are the \( D \) first elements of the canonical basis and \( b = 0 \) (that is \( f_{D+1} = ... = f_n = 0 \)), note that the procedure amounts to building the confidence ball as described in Section 3.2 by using the data \( Y_1, ..., Y_D \) and to taking advantage of the data \( Y_{D+1}, ..., Y_n \) which are i.i.d. distributed as \( \mathcal{N}(0, \sigma^2) \) to estimate the variance \( \sigma^2 \).

Theorem 4.1 is actually a corollary of Theorem 3.2 with the quantities \( \tau^2 \) and \( \eta \) built to satisfy (6) on a set \( \Omega_\gamma \) of probability at least \( 1 - 2\gamma \).

Let us now shed lights on the order of magnitude of the \( \tilde{\rho}_m \)'s. For the sake of the simplicity we assume that \( \mathcal{M}_n = \{m, n\} \) and \( \beta_m = \beta_n = \)
We then derive from Proposition 3.1 that there exists a constant $C = C(\alpha, \beta)$ such that

$$\tilde{\rho}_m^2 \leq C \max\{b^2, \eta D, D_m, \sqrt{D}\} \tau^2,$$

We restrict ourself to the case where $b^2$ is small compared to $n$ as otherwise $\tilde{\rho}^2$ is larger than $n\tau^2$ and the confidence ball useless. In this case, by using upper and lower bounds on quantiles of noncentral $\chi^2$ (we refer to Inequalities (19) and (20)) we have

$$\eta \leq C \left( \frac{b^2}{n} + \frac{1}{\sqrt{n}} \right)$$

for some constant $C$ depending on $\gamma$. We derive that if $b^2$ is small compared to $\sqrt{n}$ then

$$\tilde{\rho}_m^2 \leq C \max\{b^2, D_m, \sqrt{D}\} \tau^2$$

and if not, as $\sqrt{D} \leq \sqrt{n}$ we have

$$\tilde{\rho}_m^2 \leq C \max\{b^2, D_m\} \tau^2.$$

Note that whenever $\max\{D_m, \sqrt{D}\} \geq b^2$ we know from Theorem 3.1 that these bounds are optimal in order. On the other hand, under the constraint that the center of the confidence ball belongs to $S$, we clearly have that $\tilde{\rho}^2 \geq b^2 \sigma^2$ which leads to the optimality of these bounds.

4.3. How to choose $D$ in examples (Elli) and (Höld)? For these examples, different choices of spaces $S_D$ are possible, each corresponding to a choice of the parameter $D = \dim(S_D)$ among $\{1, \ldots, \lfloor n/2 \rfloor\}$. In the sequel, we emphasize the dependency of $b, \tilde{\rho}_{\gamma}$ with respect to the parameter $D$ through the notation $b(D), \tilde{\rho}(D)$... and raise the following question: how can we take advantage of the fact that different choices of $D$ are possible?

A way to take advantage of this fact is to choose $\gamma = \gamma(D)$ and $\beta = \beta(D)$ as functions of $D$. If these functions are suitably chosen to verify

$$\sum_{D=1}^{n-1} [1 - (1 - 2\gamma(D))(1 - \beta(D))] \leq \delta,$$

for some $\delta \in [0, 1]$, then with probability larger than $1 - \delta$ for all $D$, $f$ belongs to all the confidence balls $B(f(D), \tilde{\rho}(D))$ and consequently to their intersection.

Another way is to select some $D$ “at best”. For example one can choose $D$ to minimize the right-hand side of (17), that is to take $D =$
$D^*$ as the minimizer among $D \in \{1, ..., [n/2]\}$ of the quantity

$$\max\{b^2(D), \sqrt{D}\}.$$ 

As typical, the minimum is then achieved when $D$ realizes the best trade-off between $b^2(D)$, which is decreasing with $D$, and $\sqrt{D}$ which is increasing with $D$, and thus roughly speaking

$$b^2(D^*) \approx \sqrt{D^*} < \sqrt{n}.$$  \hfill (18)

In example (Höld) elementary computations show that for $s > 1/4$, $D^*$ is of order $n^{2/(1+4s)}$ (which belongs to $\{1, ..., [n/2]\}$ at least for $n$ large enough). In this case, by normalizing $\rho_m^2$ by $n$, we derive that

$$\frac{\rho_m^2}{n} \leq C \max \left\{ \frac{D_n}{n}, v^2(n) \right\} \tau^2,$$

where $v^2(n) = n^{-4s/(1+4s)}$. It is shown in Baraud (2000) that $v(n)$ is the minimax separation rate with respect to the norm $||| \cdot |||_{n} = ||| \cdot ||/\sqrt{n}$ for the problem of testing $f = 0$ against the alternative

$$f \in \{g \in \mathbb{R}^n, \forall D \in \{1, ..., [n/2]\}, \ d(g, S_{D}^\circ)/\sqrt{n} \leq LD^{-s}\}.$$ 

Let us now consider the case (Elli). We take $\epsilon$ as the scaling parameter, $\epsilon$ tends to 0 for asymptotical issues. By arguing similarly we deduce from the monotonicity of the sequences $b^2(D) = a_D^2 \epsilon^{-2}$ and $\sqrt{D}$ with respect to $D$, that $D^*$ which satisfies (18) also maximizes (up to a possible constant) the quantity

$$\min\{b^2(D), \sqrt{D}\}$$

among $D \in \{1, ..., [n/2]\}$. By setting

$$v^2(\epsilon) = \max_{D \in \{1, ..., [n/2]\}} \min\{a_D^2 \epsilon^{-2}, \sqrt{D}\},$$

we derive from (17) that for some constant $C = C(\alpha, \beta, \gamma)$,

$$\rho_m^2 \leq C \max\{D_n, v^2(\epsilon)\} \tau^2.$$ 

It is proved in Baraud (2000) that the quantity $v(\epsilon)$ is related to the minimax separation rate (with respect the Euclidean distance) for the problem of testing $f = 0$ against the alternative $f \in \mathcal{E}\backslash\{0\}$. 

5. Proofs

Along the proofs we repeatedly use the following inequalities on the quantiles of noncentral $\chi^2$ random variables which are due to Birgé (2001). For all $u \in ]0, 1[$, $z \geq 0$, $d \geq 1$

\begin{align*}
q_{z,d}(u) & \leq z + d + 2\sqrt{(2z + d)\log(1/u)} + 2\log(1/u) \quad (19) \\
q_{z,d}(1-u) & \geq z + d - 2\sqrt{(2z + d)\log(1/u}). \quad (20)
\end{align*}

In the sequel, $\Pi_m$ for $m \in \mathcal{M}_n$ denotes the orthogonal projector onto $S_m$.

5.1. Proof of Theorems 2.1 and 3.2. Theorem 2.1 being a straightforward consequence of Theorem 3.2 by taking $\eta = 0$, we only prove Theorem 3.2.

Let us first prove (10). The result is clear for $m = n$ as by definition $\hat{\rho} \leq \rho_m$. Let us fix some $m \in \mathcal{M}_n \setminus \{n\}$. The mapping $m' \mapsto \rho_{m'}$ being one to one, we derive from the definition of $\hat{\rho}$ that

\begin{align*}
\mathbb{P}_{f,\sigma}[\hat{\rho} > \rho_m] & \leq \mathbb{P}_{f,\sigma}[m \not\in \mathcal{A}] \\
& = \mathbb{P}_{f,\sigma}[\|Y - \hat{f}_m\|^2 > q_{0,N_m}(\alpha)\tau^2] \\
& \leq \mathbb{P}_{f,\sigma}[\|Y - \hat{f}_m\|^2 > q_{0,N_m}(\alpha)\sigma^2],
\end{align*}

as $\tau \geq \sigma$. We conclude noting that for $f \in S_m$, $\|Y - \hat{f}_m\|^2/\sigma^2$ is distributed as a $\chi^2$ with $N_m$ degrees of freedom.

We shall now show something that is stronger than (4) namely that

\begin{equation*}
\mathbb{P}_{f,\sigma}\left[f \not\in \bigcap_{m \in \mathcal{A}} B(\hat{f}_m, \rho_m)\right] \leq \beta.
\end{equation*}

For all $f \in \mathbb{R}^n$,

\begin{align*}
\mathbb{P}_{f,\sigma}\left[f \not\in \bigcap_{m \in \mathcal{A}} B(\hat{f}_m, \rho_m)\right] & \\
& = \mathbb{P}_{f,\sigma}\left[\exists m \in \mathcal{A}, \|f - \hat{f}_m\| > \rho_m\right] \\
& \leq \sum_{m \in \mathcal{M}_n} \mathbb{P}_{f,\sigma}\left[\|f - \hat{f}_m\| > \rho_m, \ \hat{m} \in \mathcal{A}\right] \\
& = \sum_{m \in \mathcal{M}_n} \mathbb{P}_{f,\sigma}\left[\|f - \hat{f}_m\| > \rho_m, \ \|Y - \hat{f}_m\|^2 \leq q_{0,N_m}(\alpha)\tau^2\right].
\end{align*}
Since $\sum_{m \in \mathcal{M}_n} \beta_m = \beta$, it is enough to prove that for each $m \in \mathcal{M}_n$, the probability
\[
P_{f,\sigma}(m) = P_{f,\sigma}\left[\|f - \hat{f}_m\| > \rho_m, \|Y - \hat{f}_m\|^2 \leq q_{0,n_m}(\alpha)\tau^2\right]
\]
is not larger than $\beta_m$. If $m = n$ this is clear since for $\tau^2 \geq \sigma^2$,
\[
P_{f,\sigma}(n) = P_{f,\sigma}\left[\|f - \hat{f}_n\| > \rho_n\right]
= P_{f,\sigma}\left[\sigma^2\|\varepsilon\|^2 > q_{0,n}(\beta_n)\tau^2\right]
\leq \beta_n.
\]
Let us now prove the inequality when $D_m = 0$. In this case $\hat{f}_m = 0$. If $\|f\| \leq \rho_m$ we have $P_{f,\sigma}(m) = 0$ and thus the inequality is true. Otherwise $\|f\| > \rho_m$ and as for all $u > 0$, $z \rightarrow \chi^2_{\tau^2}(u)$ is nondecreasing with $z$ we get
\[
P_{f,\sigma}(m) = P\left[\|Y\|^2 \leq q_{0,n}(\alpha)\tau^2\right]
= \chi^2_{\|f\|^2/\sigma^2,n}(q_{0,n}(\alpha)\tau^2/\sigma^2)
\leq \chi^2_{\rho^2_n/\sigma^2,n}(q_{0,n}(\alpha)\tau^2/\sigma^2)
\leq \beta_m
\]
by definition of $\rho_m$.

Let us now fix some $m \in \mathcal{M}_n \setminus \{n\}$ such that $D_m \neq 0$ and set $z = \|f - \Pi_m f\|^2/\sigma^2$. Note that the random variables
\[
\|f - \hat{f}_m\|^2 - \|f - \Pi_m f + \sigma \Pi_m \varepsilon\|^2 = z + \|\Pi_m \varepsilon\|^2
\]
and
\[
\|Y - \hat{f}_m\|^2 - \|f - \Pi_m f + \sigma (\varepsilon - \Pi_m \varepsilon)\|^2
\]
are independent and that the second one is distributed as a noncentral $\chi^2$ with noncentrality parameter $z$ and $N_m$ degrees of freedom. Therefore, we get
\[
P_{f,\sigma}(m) = \left(1 - \chi^2_{\rho^2_m/\sigma^2,n}(\rho^2_m/\sigma^2 - z)\right) \chi^2_{\tau^2, N_m} \left(q_{0,n_m}(\alpha)\tau^2/\sigma^2\right). \quad (21)
\]
We deduce from the definition of $\rho_m$ that for all $\sigma \in \mathcal{I}$ and $z \geq 0$, the right-hand side of (21) is not larger than $\beta_m$ which leads to the result.
5.2. **Proof of Theorem 3.1.** The principle of the proof is due to Lepski. However, the following nonasymptotic inequalities are to our knowledge new. In the sequel we set $N_m = n - D_m$. Let us now fix some $m \in \mathcal{M}_n$, we divide the proof into consecutive claims.

**Claim 1:** If $\alpha + \beta < 1 - \exp(-1/36)$ then

$$r_m^2 \geq \left( \frac{D_m}{27} - \sqrt{L_1 D_m} \right) \sigma^2$$

where $L_1 = -4 \log(1 - \alpha - \beta)/81$.

Note that the claim is clear when $D_m = 0$, we shall thus restrict ourself to the case $D_m \geq 1$. The proof relies on two lemmas. In the first one, we show that under the assumption of Theorem 3.1, with probability close to one the Euclidean distance between $f \in S_m$ and its estimator $\hat{f}$ is not larger than $r_m$.

**Lemma 5.1.** Let the pair $(\hat{f}, \hat{r})$ satisfies the assumption of Theorem 3.1. Then, for all $m \in \mathcal{M}_n$ and $f \in S_m$

$$\mathbb{P}_{f, \sigma} \left[ \|f - \hat{f}\| > r_m \right] \leq \alpha + \beta. \quad (22)$$

**Proof.** For all $f \in S_m$,

$$\mathbb{P}_{f, \sigma} \left[ \|f - \hat{f}\| > r_m \right]$$

$$\leq \mathbb{P}_{f, \sigma} \left[ \|f - \hat{f}\| > r_m, \ r_m \geq \hat{r} \right] + \mathbb{P}_{f, \sigma} \left[ \|f - \hat{f}\| > r_m, \ \hat{r} > r_m \right]$$

$$\leq \mathbb{P}_{f, \sigma} \left[ \|f - \hat{f}\| > \hat{r} \right] + \mathbb{P}_{f, \sigma} \left[ \hat{r} > r_m \right]$$

and we conclude thanks to (7) and (8). \qed

The second lemma shows that such a property of the estimator $\hat{f}$ is possible only if $r_m$ is large enough.

**Lemma 5.2.** Let $S$ be a linear subspace of $\mathbb{R}^n$ of dimension $D \geq 1$ and $\delta$ a positive number such that $\delta < 1 - \exp[-D/36]$. If $\hat{f}$ is an estimator of $f$ in (1) which satisfies for all $f \in S$

$$\mathbb{P}_{f, \sigma} \left[ \|f - \hat{f}\| > v_D(\delta) \right] \leq \delta, \quad (23)$$

then

$$v_D^2(\delta) \geq \left( \frac{D}{27} - \frac{2}{9} \sqrt{D \log(1/(1 - \delta))} \right) \sigma^2.$$
At the light of Lemma 5.1, the claim derives from Lemma 5.2 by taking \( S = S_m \) and \( \delta = \alpha + \beta \). Let us now turn to the proof of Lemma 5.2.

**Proof.** The Gaussian law being invariant by orthogonal transformation, with no loss of generality, we assume that \( S \) is the linear span generated by \( e_1, \ldots, e_D \), the \( D \) first vectors of the canonical basis of \( \mathbb{R}^n \). Moreover, by homogeneity, we assume that \( \sigma^2 = 1 \). Let \( v(\delta) \) be some positive number satisfying

\[
v^2(\delta) < \frac{D}{27} - \frac{2}{9} \sqrt{D \log(1 - \delta)}.
\]

(24)

Note that the right-hand side of (24) is positive for \( \delta < 1 - \exp[-D/36] \).

We prove Lemma 5.2 by showing that for all estimator \( \hat{f} \) with values in \( \mathbb{R}^n \),

\[
\inf_{f \in S, \ P_{f,1}} \mathbb{P}_{f,1} \left[ \| f - \hat{f} \|^2 \leq v^2(\delta) \right] < 1 - \delta.
\]

Let \( \xi_1, \ldots, \xi_D \) be Rademacher random variables (i.e. \( \mathbb{P}[\xi_i = \pm 1] = 1/2 \)) which are independent of \( Y \) and set \( f(\xi) = \lambda \sum_{i=1}^D \xi_i e_i \) where \( \lambda \) denotes for some positive number to be chosen later on. Using that

\[
\frac{d\mathbb{P}_{f(\xi),1}}{d\mathbb{P}_{0,1}}(y) = \exp \left( -\lambda^2 D/2 + \lambda \sum_{i=1}^D \xi_i y_i \right),
\]

and the fact that \( f(\xi) \in S \), we have

\[
\inf_{f \in S, \ P_{f,1}} \mathbb{P}_{f,1} \left[ \| f - \hat{f} \|^2 \leq v^2(\delta) \right]
\]

\[
\leq \mathbb{P}_{f(\xi),1} \left[ \sum_{i=1}^D (\lambda \xi_i - \hat{f}_i)^2 \leq v^2(\delta) \right]
\]

\[
= \mathbb{E}_{0,1} \left[ \mathbb{I} \left\{ \sum_{i=1}^D (\lambda \xi_i - \hat{f}_i(Y))^2 \leq v^2(\delta) \right\} \exp \left( -\lambda^2 D/2 + \lambda \sum_{i=1}^D \xi_i Y_i \right) \right].
\]

Note that \( \hat{f} = \hat{f}(Y) \) satisfies

\[
\sum_{i=1}^D (\lambda \xi_i - \hat{f}_i)^2 \geq \lambda^2 \sum_{i=1}^D \mathbb{I} \left\{ \xi_i \hat{f}_i(Y) \leq 0 \right\}
\]

and thus setting

\[
N(\xi, \hat{f}) = \lambda^2 \sum_{i=1}^D \mathbb{I} \left\{ \xi_i \hat{f}_i(Y) \leq 0 \right\},
\]
we derive

\[
\inf_{f \in S} \mathbb{P}_{f, \sigma} \left[ \| f - \hat{f} \|^2 \leq v^2(\delta) \right] \\
\leq \mathbb{E}_{0,1} \left[ \mathbb{I} \left\{ N(\xi, \hat{f}) \leq v^2(\delta) \right\} \exp \left( -\lambda^2 D/2 + \lambda \sum_{i=1}^{D} \xi_i Y_i \right) \right].
\]

By averaging with respect to \( \xi \) and using Fubini’s Theorem we get

\[
\inf_{f \in S} \mathbb{P}_{f, \sigma} \left[ \| f - \hat{f} \|^2 \leq v^2(\delta) \right] \\
\leq e^{-\lambda^2 D/2} \mathbb{E}_{0,1} \left[ \mathbb{E}_\xi \left[ \mathbb{I} \left\{ N(\xi, \hat{f}) \leq v^2(\delta) \right\} \exp \left( \lambda \sum_{i=1}^{D} \xi_i Y_i \right) \right] \right] (25)
\]

By Cauchy-Scharwz’s inequality we have

\[
\mathbb{E}_\xi^2 \left[ \mathbb{I} \left\{ N(\xi, \hat{f}) \leq v^2(\delta) \right\} \exp \left( \lambda \sum_{i=1}^{D} \xi_i Y_i \right) \right] \\
\leq \mathbb{P}_\xi \left[ N(\xi, \hat{f}) \leq v^2(\delta) \right] \mathbb{E}_\xi \left[ \exp \left( 2\lambda \sum_{i=1}^{D} \xi_i Y_i \right) \right] \\
= \mathbb{P}_\xi \left[ N(\xi, \hat{f}) \leq v^2(\delta) \right] \prod_{i=1}^{D} \cosh (2\lambda Y_i),
\]

which together with (25) gives

\[
\inf_{f \in S} \mathbb{P}_{f, \sigma} \left[ \| f - \hat{f} \|^2 \leq v^2(\delta) \right] \\
\leq e^{-\lambda^2 D/2} \mathbb{E}_{0,1} \left[ \mathbb{P}_\xi^{1/2} \left[ N(\xi, \hat{f}) \leq v^2(\delta) \right] \prod_{i=1}^{D} \cosh^{1/2} (2\lambda Y_i) \right]. (26)
\]

Conditionally to \( Y \), the random variable \( N(\xi, \hat{f})/\lambda^2 \) is a sum of \( D \) independent random variables with values in \( \{0, 1\} \). Thus by Hoeffding’s inequality we obtain that for all \( t \geq 0 \),

\[
\mathbb{P}_\xi \left[ N(\xi, \hat{f}) \leq \mathbb{E} \left[ N(\xi, \hat{f}) \right] - \lambda^2 \sqrt{Dt} \right] \leq e^{-2t}.
\]
Taking \( t = \lambda^2 D/2 - \log(1 - \delta) \) and noting that \( \mathbb{E}_f \left[ N(\xi, \tilde{f}) \right] \geq \lambda^2 D/2 \) we get from (24) that
\[
\mathbb{E}_f \left[ N(\xi, \tilde{f}) \right] - \lambda^2 \sqrt{D} \leq \mathbb{E}_f \left[ N(\xi, \tilde{f}) \right] - \lambda^2 \sqrt{D} t \geq \lambda^2 \left( \frac{D}{2} - \sqrt{\frac{\lambda^2 D^2}{2} - D \log(1 - \delta)} \right) \geq \left( \frac{\lambda^2}{2} - \frac{\lambda^3}{\sqrt{2}} \right) D - \lambda^2 \sqrt{-D \log(1 - \delta)}
\]
and thus for \( \lambda = \sqrt{2}/3 \)
\[
\mathbb{E}_f \left[ N(\xi, \tilde{f}) \right] - \lambda^2 \sqrt{D} \leq \mathbb{E}_f \left[ N(\xi, \tilde{f}) \right] - \lambda^2 \sqrt{D} \leq v^2(\delta).
\]
Consequently,
\[
\mathbb{P}_f^{1/2} \left[ N(\xi, \tilde{f}) \leq v^2(\delta) \right] \leq e^{-t} = (1 - \delta) e^{-\lambda^2 D/2}.
\]
Using now that
\[
\mathbb{E}_{0,1} \left[ \prod_{i=1}^{D} \cosh^{1/2} (2\lambda Y_i) \right] = \prod_{i=1}^{D} \mathbb{E}_{0,1} \left[ \cosh^{1/2} (2\lambda Y_i) \right] < \mathbb{E}_{0,1}^{D/2} \left[ \cosh (2\lambda Y) \right] = \exp \left[ \lambda^2 D \right],
\]
we derive from (26) that
\[
\inf_{f \in \mathcal{S}} \mathbb{P}_{f,\sigma} \left[ \| f - \tilde{f} \| \leq v^2(\delta) \right] < 1 - \delta
\]
which concludes the proof.

Claim 2: If \( \alpha + 2\beta \leq 1 - \exp(-1/4) \) then
\[
r_m^2 \geq \max \{ \sqrt{\mathcal{L}^2 N_m \sigma^2}, \left( N_m - 2\sqrt{N_m \mathcal{L}} \right) \eta \tau^2 \}. \tag{27}
\]
with \( \mathcal{L} = 2 \log(1 + 4(1 - \alpha - 2\beta)^2) \) and \( \mathcal{L} = -\log(1 - \alpha - 2\beta) \).

The claim is clear when \( N_m = 0 \), thus we only consider the case where \( N_m \geq 1 \). Again, the proof relies on two lemmas. The first one shows that if the pair \( (\tilde{f}, \tilde{f}) \) satisfies the assumptions of Theorem 3.1 then it is possible to build a level \( (\alpha + \beta)-\)test of \( \left\{ f \in \mathcal{S}_m \right\} \) against \( \left\{ f \in \mathbb{R}^n \setminus \mathcal{S}_m \right\} \) which achieves the power \( 1 - \beta \) on the complementary of a ball of radius \( 3r_m \). Namely, the following holds:
Lemma 5.3. Let \((\tilde{f}, \tilde{r})\) be a pair of random variables with values in \(\mathbb{R}^n \times \mathbb{R}_+\) satisfying the assumptions of Theorem 3.1. The test of hypothesis \(f \in S_m\) against the alternative \(f \notin S_m\) associated to the critical region
\[
\mathcal{R} = \{\tilde{r} > r_m\} \cup \left\{\|\tilde{f} - \Pi_m \tilde{f}\| > 2\tilde{r}\right\}
\]
has the following properties: for all \(\sigma \in I\)
\[
\sup_{f \in S_m} \mathbb{P}_{f, \sigma} [\mathcal{R}] \leq \alpha + \beta,
\]
and for all \(f\) satisfying \(\|f - \Pi_m f\| > 3r_m\)
\[
\mathbb{P}_{f, \sigma} [\mathcal{R}] \geq 1 - \beta.
\]

Proof. Let us show (29). First note that for all \(f \in S_m\)
\[
\|\tilde{f} - \Pi_m \tilde{f}\| \leq \|f - \tilde{f}\| + \|f - \Pi_m \tilde{f}\| \leq 2\|f - \tilde{f}\|.
\]
By (7), (8) and (31), for all \(f \in S_m\) we have
\[
\mathbb{P}_{f, \sigma} [\mathcal{R}] \leq \mathbb{P}_{f, \sigma} [\tilde{r} > r_m]
+ \mathbb{P}_{f, \sigma} \left[\|\tilde{f} - \Pi_m \tilde{f}\| > 2r_m, \ r_m \geq \tilde{r}\right]
\leq \alpha + \mathbb{P}_{f, \sigma} \left[2\|f - \tilde{f}\| > 2\tilde{r}\right]
\leq \alpha + \beta.
\]
Let us now show (30). Let \(f \in \mathbb{R}^n\) such that \(\|f - \Pi_m f\| \geq 3r_m\). Since
\[
\|\tilde{f} - \Pi_m \tilde{f}\| \geq \|f - \Pi_m \tilde{f}\| - \|f - \tilde{f}\|
\geq 3r_m - \|f - \tilde{f}\|,
\]
we derive that
\[
\mathbb{P}_{f, \sigma} [\mathcal{R}^c] = \mathbb{P}_{f, \sigma} \left[\|\tilde{f} - \Pi_m \tilde{f}\| \leq 2\tilde{r}, \ \tilde{r} \leq r_m \right]
\leq \mathbb{P}_{f, \sigma} \left[\|\tilde{f} - \Pi_m \tilde{f}\| \leq 2r_m, \ \tilde{r} \leq r_m \right]
\leq \mathbb{P}_{f, \sigma} \left[\|f - \tilde{f}\| \geq r_m, r_m \geq \tilde{r}\right]
\leq \mathbb{P}_{f, \sigma} \left[\|f - \tilde{f}\| \geq \tilde{r}\right] \leq \beta.
\]
\[\square\]
We obtain the claim by proving that a test having the properties describe in the previous lemma exists only if \( r_m \) is large enough. The inequality

\[
 r_m^2 \geq \sqrt{\mathcal{L}_2 N_m \sigma^2},
\]

derives from Baraud (2000) (Proposition 1). For the second inequality

\[
 r_m^2 \geq \mathcal{L}_3 N_m \eta \tau^2
\]

we use the following lemma.

**Lemma 5.4.** Let \( S \) be a linear subspace of \( \mathbb{R}^n \) with \( \dim(S) = D \) (we set \( N = n - D \)) and \( \delta \) and \( \beta \) numbers verifying \( 0 < \beta + \delta < 1 - \exp(-N/4) \). Let \( \phi(Y) \) be a test function with values in \( \{0, 1\} \) satisfying for all \( \sigma \in I \),

\[
 \sup_{f \in S} \mathbb{P}_{f, \sigma} [\phi(Y) = 1] \leq \delta,
\]

and for all \( f \in \mathbb{R}^n \) such that \( \|f - \Pi_S f\|^2 \geq \Delta(D, \beta) \),

\[
 \mathbb{P}_{f, \sigma} [\phi(Y) = 1] \geq 1 - \beta.
\]

Then

\[
 \Delta(D, \beta) \geq \left( N - 2 \sqrt{-N \log(1 - \beta - \delta)} \right) \eta \tau^2.
\]

By applying this lemma with \( \delta = \alpha + \beta \), \( S = S_m \) and \( D = D_m \) and the test described in Lemma 5.3 we obtain the claim.

**Proof.** Let \( \mathcal{F} \) be the set defined by

\[
 \mathcal{F} = \{ f \in S^\perp, \|f\|^2 \geq \Delta \},
\]

where \( \Delta \) denotes some positive number. To obtain the desired result it is enough to show that for

\[
 \Delta < \left( N - 2 \sqrt{-N \log(1 - \beta - \delta)} \right) \eta \tau^2,
\]

we have

\[
 \inf_{f \in \mathcal{F}} \mathbb{P}_{f, \sigma} [\phi(Y) = 1] < 1 - \beta.
\]

Since the quantity \( \sigma^2 = (1 + \eta) \tau^2 \) satisfies (6), we have that for all vector \( Z \in S^\perp \),

\[
 \inf_{\sigma \in I} \inf_{f \in \mathcal{F}} \mathbb{P}_{f, \sigma} [\phi(Y) = 1] 
 \leq \mathbb{P}_{Z, \sigma} [\phi(Y) = 1] \mathbb{I} \{ \|Z\|^2 \geq \Delta \} + \mathbb{I} \{ \|Z\|^2 \leq \Delta \}.
\]

By taking \( Z \) as a random variables independent of \( Y \) distributed as \( \sqrt{\eta} \tau \Pi_{S^\perp} \), we obtain by averaging with respect to \( Z \) that

\[
 \inf_{\sigma \in I} \inf_{f \in \mathcal{F}} \mathbb{P}_{f, \sigma} [\phi(Y) = 1] \leq \mathbb{E} \left[ \mathbb{P}_{Z, \sigma} [\phi(Y) = 1] \right] + \mathbb{P} \left[ \|Z\|^2 \leq \Delta \right].
\]
For the first term of the right-hand side of this inequality, note that 
\[ \mathbb{E}[P_{Z, \sigma}^*] = P_{0, \tau}. \] 
As \( 0 \in S \) and \( \tau \in \Sigma \), we have
\[ \mathbb{E}[P_{Z, \sigma}^*, [\phi(Y) = 1]] \leq \delta. \]
For the second term, note that our upper bound on \( \Delta \) ensures that
\[ \Delta < q_0 N (1 - \beta - \delta) \eta \tau^2 \]
by using the lower bound on the quantiles of \( \chi^2 \) random variables (20). As the random variable \( \|Z\|^2 / (\eta \tau^2) \) is distributed as a \( \chi^2(N) \), we get
\[ \mathbb{P} \left[ \|Z\|^2 \leq \Delta \right] < 1 - \beta - \delta, \]
which concludes the proof. \( \square \)

**Conclusion:** By gathering the inequalities of the two claims together with the fact that \( \sigma^2 \geq (1 - \eta) \tau^2 \) we get that for some constant \( C \) depending on \( \alpha \) and \( \beta \) only,
\[ \tau_m^2 \geq C \max \{ N_m \eta, D_m (1 - \eta), \sqrt{N_m (1 - \eta)} \} \tau^2. \]

### 5.3. Proof of Theorem 4.1

Let us first check that with probability \( 1 - 2\gamma \), the pair \((\tau, \eta)\) satisfies (6). Note that the random variable \( \|Y - \Pi_{S} Y\|^2 / \sigma^2 \) is distributed as a noncentral \( \chi^2 \) with noncentrality parameter \( ||f - \Pi_{S} f||^2 / \sigma^2 \) and \( D \) degrees of freedom. On the one hand, since for and \( u \in [0, 1] \), the function \( z \mapsto q_{z, D}(u) \) is increasing on \( \mathbb{R}_+ \) we get that
\[ 1 - \gamma = \mathbb{P}_{f, \sigma} \left[ \|Y - \Pi_{S} Y\|^2 \geq q \left( \|f - \Pi_{S} f\|^2 / \sigma^2, D, 1 - \gamma \right) \sigma^2 \right], \]
\[ \leq \mathbb{P}_{f, \sigma} \left[ \|Y - \Pi_{S} Y\|^2 \geq q \left( 0, D, 1 - \gamma \right) \sigma^2 \right], \]
and therefore, with probability larger than (or equal to) \( 1 - \gamma \),
\[ \tau^2 = \frac{||Y - \Pi_{S} Y||^2}{q (D, 1 - \gamma)} \geq \sigma^2. \]

On the other hand,
\[ 1 - \gamma = \mathbb{P}_{f, \sigma} \left[ \|Y - \Pi_{S} Y\|^2 \leq q \left( \|f - \Pi_{S} f\|^2 / \sigma^2, D, \gamma \right) \sigma^2 \right], \]
\[ \leq \mathbb{P}_{f, \sigma} \left[ \|Y - \Pi_{S} Y\|^2 \leq q \left( b^2, D, \gamma \right) \sigma^2 \right], \]
\[ = \mathbb{P}_{f, \sigma} \left[ (1 - \eta) \tau^2 \leq \sigma^2 \right]. \]

We deduce that
\[ \mathbb{P}_{f, \sigma} \left[ (1 - \eta) \tau^2 \leq \sigma^2 \leq \tau^2 \right] \geq 1 - 2\gamma. \]

We now prove Theorem 4.1 by applying Theorem 3.2. Actually, we only prove (15), the prove of (16) being analogous. By a change-of-basis
argument we reduce to the case where $S$ is the linear span generated by the $D$ first vectors of the canonical basis.

Setting $\tilde{f}_D = (f_1, \ldots, f_D)'$, as $d^2(f, S) \geq b^2 \sigma^2$ we obtain from Pythagoras' theorem

$$\mathbb{P}_{f,\sigma} \left[ f \in B(\tilde{f}, \tilde{\rho}) \right] \geq \mathbb{P}_{f,\sigma} \left[ b^2 \sigma^2 + \| \tilde{f}_D - \tilde{f} \|^2 \leq b^2 \tau^2 + \tilde{\rho}^2 \right]$$

$$\geq \mathbb{P}_{f,\sigma} \left[ \| \tilde{f}_D - \tilde{f} \|^2 \leq \tilde{\rho}^2, \ 0 \leq \tau^2 - \sigma^2 \leq \eta \tau^2 \right].$$

By Cochran’s theorem, the random variables $\tau^2$ and $\Pi_S Y$ are independent and therefore, by conditioning with respect to the event $\{0 \leq \tau^2 - \sigma^2 \leq \eta \tau^2\}$ we derive from Theorem 3.2 that

$$\mathbb{P}_{f,\sigma} \left[ \| \tilde{f}_D - \tilde{f} \|^2 \leq \tilde{\rho}^2 | 0 \leq \tau^2 - \sigma^2 \leq \eta \tau^2 \right] \geq 1 - \beta$$

and deduce that

$$\mathbb{P}_{f,\sigma} \left[ f \in B(\tilde{f}, \tilde{\rho}) \right] \geq (1 - 2\gamma)(1 - \beta),$$

which concludes the proof.

5.4. **Proof of Proposition 2.1 and 3.1.** The result of the former proposition being a consequence of the latter by taking $\eta = 0$ we only prove Proposition 3.1. In the sequel we set $L_m = \log(1/\beta_m)$ and $L_\alpha = \log(1/\alpha)$. We distinguish between three cases.

**Case $m = n$:** We derive from (19),

$$\rho_n^2 \leq \left( n + 2 \sqrt{n L_n} + 2 L_n \right) \tau^2,$$

which leads to the result.

**Case $D_m \neq 0$, $m \neq n$:** Let us fix $\sigma \in I$. Since for $z$ verifying

$$\chi^2_{z, N_m} \left( q_{0, N_m}(\alpha) \tau^2 / \sigma^2 \right) \leq \beta_m,$$

we have

$$z + q_{0, D_m} \left( \frac{\beta_m}{\chi^2_{z, N_m} \left( q_{0, N_m}(\alpha) \tau^2 / \sigma^2 \right) \wedge 1} \right) = -\infty,$$

we bound from above the left-hand side of (35) for those $z$ verifying

$$\chi^2_{z, N_m} \left( q_{0, N_m}(\alpha) \tau^2 / \sigma^2 \right) > \beta_m.$$  \hspace{1cm} (36)

It follows from (20) that if $z$ satisfies (36) then

$$q_{0, N_m}(\alpha) \frac{\tau^2}{\sigma^2} \geq z + N_m - 2 \sqrt{(2z + N_m) L_m}$$
and as we have
\[ 2\sqrt{(2z + N_m)L_m} \leq 2\sqrt{2zL_m} + 2\sqrt{N_mL_m} \leq \frac{z}{2} + 2\sqrt{N_mL_m} + 4L_m \]
and
\[ q_{0,N_m}(\alpha) \leq N_m + 2\sqrt{N_mL_\alpha} + 2L_\alpha \]
from (19), we deduce that \( z \) satisfies
\[
\begin{align*}
zs^2 & \leq \left( 2 \left( \frac{q_{0,N_m}(\alpha)}{\sigma^2} - N_m \right) + 4\sqrt{N_mL_m + 8L_m} \right) \sigma^2 \\
& \leq 2N_m(\tau^2 - \sigma^2) + \left( 4\sqrt{N_m} \left( \sqrt{L_m + \sqrt{L_\alpha}} \right) + 8L_m + 4L_\alpha \right) \sigma^2 \\
& \leq \left( 2N_m\eta + 4\sqrt{N_m} \left( \sqrt{L_m + \sqrt{L_\alpha}} \right) + 8L_m + 4L_\alpha \right) \tau^2. 
\end{align*}
\]
Thanks to (19) and the facts that \( \chi_{\tau^2/N_m}^2(q_{0,N_m}(\alpha)\tau^2/\sigma^2) \leq 1 \) and \( D_m \leq N_m \), we deduce that for those \( z \)
\[
zs^2 + q_{0,D_m}\left( \frac{\beta_m}{\chi_{\tau^2/N_m}^2(q_{0,N_m}(\alpha)\tau^2/\sigma^2)} \wedge 1 \right) \sigma^2 \\
\leq \left( z + D_m + 2\sqrt{D_mL_m + 2L_m} \right) \sigma^2 \\
\leq \left( z + D_m + 2\sqrt{N_mL_m + 2L_m} \right) \sigma^2 \\
\leq \left( 2N_m\eta + D_m + 2\sqrt{N_m} \left( 3\sqrt{L_m + 2\sqrt{L_\alpha}} \right) + 2(5L_m + 2L_\alpha) \right) \tau^2,
\]
and consequently, that
\[
\rho_m^2 \leq \left( 2N_m\eta + D_m + 2\sqrt{N_m} \left( 3\sqrt{L_m + 2\sqrt{L_\alpha}} \right) + 2(5L_m + 2L_\alpha) \right) \tau^2.
\]
The result follows as \( N_m \leq n \).

**Case** \( D_m = 0 \): Arguing as above we have that for \( x \) verifying
\[
x \geq \left( 2N_m\eta + 4\sqrt{N_m} \left( \sqrt{L_m + \sqrt{L_\alpha}} \right) + 8L_m + 4L_\alpha \right) \tau^2
\]
we have that for all \( \sigma \in I \),
\[
\chi_{\tau^2/\sigma^2,n}^2(q_{0,n}(\alpha)\tau^2/\sigma^2) \leq \beta_m
\]
and therefore, by definition of \( \rho_m \),
\[
\rho_m^2 \leq \left( 2n\eta + 4\sqrt{n} \left( \sqrt{L_m + \sqrt{L_\alpha}} \right) + 8L_m + 4L_\alpha \right) \tau^2,
\]
which leads to the result.


**Acknowledgments.** The author would like to thank Lucien Birgé for helpful discussions.

**ÉCOLE NORMALE SUPÉRIEURE. DÉPARTEMENT DE MATHEMATIQUES ET APPLICATIONS, CNRS UMR 8553, 45 RUE D’ULM, 75230 PARIS CEDEX 05**

**E-mail address:** yannick.baraud@ens.fr