Testing convex hypotheses on the mean of a Gaussian vector. Application to testing qualitative hypotheses on a regression function

By Y. Baraud, S. Huet, B. Laurent

Ecole Normale Supérieure, DMA
INRA, Laboratoire de Biométrie.
Université Paris-Sud.

In this paper we propose a general methodology, based on multiple testing, for testing that the mean of a Gaussian vector in $\mathbb{R}^n$ belongs to a convex set. We show that the test achieves its nominal level, and characterize a class of vectors over which the tests achieve a prescribed power. In the functional regression model, this general methodology is applied to test some qualitative hypotheses on the regression function. For example, we test that the regression function is positive, increasing, convex, or more generally, satisfies a differential inequality. Uniform separation rates over classes of smooth functions are established and a comparison with other results in the literature is provided. A simulation study evaluates some of the procedures for testing monotonicity.

1. Introduction.

1.1. *The statistical framework.* We consider the following regression model:

$$Y_i = F(x_i) + \sigma\varepsilon_i,\ i = 1, ..., n \tag{1}$$

where $x_1 < x_2 < \ldots < x_n$ are known deterministic points in $[0, 1]$, σ is an unknown positive number and $(\varepsilon_i)_{i=1,\ldots,n}$ is a sequence of i.i.d. unobservable standard Gaussian random variables. From the observation of $\boldsymbol{Y} = (Y_1, \ldots, Y_n)'$, we consider the problem of testing that the regression function F belongs to one of the following functional sets $\mathcal{K}$:

$$\mathcal{K}_{\geq 0} = \{F : [0, 1] \to \mathbb{R},\ F \text{ is non-negative } \} \tag{2}$$

$$\mathcal{K}_{\nearrow} = \{F : [0, 1] \to \mathbb{R},\ F \text{ is non-decreasing } \} \tag{3}$$

$$\mathcal{K}_{\smile} = \{F : [0, 1] \to \mathbb{R},\ F \text{ is non-concave } \} \tag{4}$$

$$\mathcal{K}_{r,R} = \left\{F : [0, 1] \to \mathbb{R},\ \forall x \in [0, 1], \frac{d^r}{dx^r}[R(x)F(x)] \geq 0\right\}. \tag{5}$$

In the above definition of $\mathcal{K}_{r,R}$, r denotes a positive integer and R a smooth, non-vanishing function from $[0, 1]$ into $\mathbb{R}$. Choosing the function R equal to 1 leads to

Received

AMS 2000 subject classifications. Primary 62G10; Secondary 62G20.

Key words and phrases. tests of qualitative hypotheses, non-parametric test, test of positivity, test of monotonicity, test of convexity, rate of testing, Gaussian regression

test that the derivative of order r is positive. Taking $r = 1$ and choosing a suitable function R leads to test that a positive function F is decreasing at some prescribed rate. It is also possible to test that F belongs to some classes of smooth functions. These testing hypotheses will be detailed in Section 3.
The problem is therefore to test some qualitative hypothesis on F. We shall show that it actually reduces to testing that the mean of the Gaussian vector $\boldsymbol{Y}$ belongs to a suitable convex subset of $\mathbb{R}^n$. Denoting by $<,>$ the inner product of $\mathbb{R}^n$, this convex subset takes the form

$$\mathcal{C} = \{\boldsymbol{f} \in \mathbb{R}^n,\ \forall j \in \{1, ..., p\}\ \ < \boldsymbol{f}, \boldsymbol{v}_j > \leq 0\},$$

where the vectors $\{\boldsymbol{v}_1, ..., \boldsymbol{v}_p\}$ are linearly independent in $\mathbb{R}^n$. The aim of this paper is to present a general methodology for the problem of testing that $\boldsymbol{f}$ belongs to $\mathcal{C}$ and to characterize a class of vectors over which the tests achieve a prescribed power. This general methodology is applied to test that the regression function F belongs to one of the sets $\mathcal{K}$. For the procedures we propose, the least-favorable distribution under the null hypothesis is achieved for $F = 0$ and $\sigma = 1$. Consequently, by carrying out simulations, we easily obtain tests that achieve their nominal level for fixed values of n. Moreover, we show that these tests have good properties under smooth alternatives.

For the problem of testing positivity, monotonicity and convexity, we obtain tests based on the comparison of local means of consecutive observations. A precise description of these tests is given in Section 2. For the problem of testing monotonicity, our methodology also leads to tests based on the slopes of regression lines on short intervals, as explained in Section 3.1. These procedures, based on "running gradients", are akin to those proposed by Hall and Heckman (2000). For the problem of testing that F belongs to $\mathcal{K}_{r,R}$ with a non-constant function R we refer the reader to Section 3.2. We have delayed the description of the general methodology for testing that $\boldsymbol{f}$ belongs to $\mathcal{C}$ to Section 4. Simulations studies for testing monotonicity are shown in Section 5. The proofs are postponed to Sections 6 to 10.

1.2. *An overview of the literature.* In the literature, tests of monotonicity have been widely studied in the regression model. The test proposed by Bowman et al. (1998) is based on a procedure described in Silverman (1981) for testing unimodality of a density. This test is not powerful when the regression is flat or nearly flat as underlined by Hall and Heckman (2000). Hall and Heckman (2000) proposed a procedure based on "running gradients" over short intervals for which the least-favorable distribution under the null, when σ is known, corresponds to the case where F is identically constant. The test proposed by Gijbels et al. (2000) is based on the signs of differences between observations. The test offers the advantage not to depend on the error distribution when it is continuous. Consequently, the nominal level of the test is guaranteed for all continuous error distributions. In the functional regression model with random x_i's, the procedure proposed by Ghosal et al. (2001) is based on a locally weighted version of Kendall's tau. The procedure uses a kernel smoothing with a particular choice of the bandwidth and as in Gijbels et al. (2000) depends on the signs of the quantities $(Y_j - Y_i)(x_i - x_j)$. They show that for

certain local alternatives the power of their test tends to 1. Some comments on the power of our test under those alternatives can be found in Section 3.3. In Baraud, Huet and Laurent (2003) we propose a procedure which aims at detecting discrepancies with respect to the $\mathbb{L}^2(\mu_n)$-distance where $\mu_n = n^{-1}\sum_{i=1}^n \delta_{x_i}$. This procedure generalizes that proposed in Baraud, Huet and Laurent (2003) for linear hypotheses. A common feature of the present paper with these two lies in the fact that the proposed procedures achieve their nominal level and a prescribed power over a set of vectors we characterize. In the Gaussian white noise, Juditsky and Nemirovski (2000) propose to test that the signal belongs to the cone of non-negative, non-decreasing or non-concave functions. For a given $r \in [1, +\infty[$, their tests are based on the estimation of the $\mathbb{L}^r$-distance between the signal and the cone. However, this approach requires that the signal have a known smoothness under the null. In the Gaussian white noise model, other tests of such qualitative hypotheses are proposed by Dümbgen and Spokoiny (2001). Their procedure is based on the supremum over all bandwidths of the distance in sup-norm between a kernel estimator and the null hypothesis. They adopt a minimax point of view to evaluate the performances of their tests and we adopt the same in Sections 2 and 3.

1.3. *Uniform separation rates and optimality.* Comparing the performances of tests naturally arises in the problem of hypothesis testing. In this paper, we shall mainly describe the performances of our procedures in terms of uniform separation rates over classes of smooth functions. Given β in $]0, 1[$, a class of smooth functions $\mathcal{F}$ and a "distance" $\Delta(.)$ to the null hypothesis, we define the uniform separation rate of a test Φ over $\mathcal{F}$, denoted by $\rho(\Phi, \mathcal{F}, \Delta)$, as the smallest number ρ such that the test guarantees a power not smaller than $1 - \beta$ for all alternatives F in $\mathcal{F}$ at distance ρ from the null. More precisely,

$$\rho(\Phi, \mathcal{F}, \Delta) = \inf\left\{\rho > 0,\ \forall F \in \mathcal{F},\ \Delta(F) \geq \rho \Rightarrow \mathbb{P}_F\left(\Phi \text{ rejects }\right) \geq 1 - \beta\right\}. \tag{6}$$

In the regression or Gaussian white noise model, the word "rate" refers to the asymptotics of $\rho(\Phi, \mathcal{F}, \Delta) = \rho_\tau(\Phi, \mathcal{F}, \Delta)$ with respect to a scaling parameter τ (the number of observations n in the regression model, the level of the noise in the Gaussian white noise). Comparing the performances of two tests of the same level amounts to comparing their uniform separation rates (the smaller the better). A test is said to be optimal if there exists no better test. The uniform separation rate of an optimal test is called *the minimax separation rate*. In the sequel, we shall enlarge this notion of optimality by saying that a test is rate-optimal over $\mathcal{F}$ if its uniform separation rate differs from the minimax one by a bounded function of τ. Unfortunately, not much is known on the uniform separations rates of the tests mentioned in Section 1.2. The only exception we are aware of concerns the tests proposed by Dümbgen and Spokoiny (2001) and Juditsky and Nemirovski (2000) in the Gaussian white noise model (with $\tau = 1/\sqrt{n}$), and Baraud, Huet and Laurent (2003) in the regression model. The rates obtained by Juditsky and Nemirovski (2000) are established for the problem of testing that F belongs to $\mathcal{K} \cap \mathcal{H}$ where $\mathcal{H}$ is a class of smooth functions. In contrast, in the paper by Baraud, Huet and Laurent (2003) and Dümbgen and Spokoiny (2001), the null hypothesis is not restricted to those smooth functions belonging to $\mathcal{K}$. For the problem of testing positivity and monotonicity,

Baraud, Huet and Laurent (2003) established separation rates with respect to the $\mathbb{L}^2(\mu_n)$-distance to the null. For the problem of testing positivity, monotonicity and convexity, Dümbgen and Spokoiny (2001) considered the problem of detecting a discrepancy to the null in sup-norm. For any $L > 0$, their procedures are proved to achieve the optimal rate $(L\log(n)/n)^{1/3}$ over the class of Lipschitz functions

$$\mathcal{H}_1(L) = \{F,\ \forall x, y \in [0,1],\ |F(x) - F(y)| \leq L|x-y|\}\,.$$

The optimality of this rate derives from the lower bounds established by Ingster (1993)[Section 2.4] for the more simple problem of testing $F = 0$ against $F \neq 0$ in sup-norm. More generally, it can easily be derived from Ingster's results (see Proposition 2) that the minimax separation rate (in sup-norm) over Hölderian balls

$$\mathcal{H}_s(L) = \{F,\ \forall x, y \in [0,1],\ |F(x) - F(y)| \leq L|x-y|^s\} \quad \text{with } s \in]0,1] \tag{7}$$

is bounded from below (up to a constant) by $\left(L^{1/s}\log(n)/n\right)^{s/(1+2s)}$. In the regression setting, we propose tests of positivity, monotonicity and convexity whose uniform separation rates over $\mathcal{H}_s(L)$ achieve this lower bound whatever the value of $s \in]0,1]$ and $L > 0$. We only discuss the optimality in the minimax sense over the Hölderian balls $\mathcal{H}_s(L)$ with $s \in]0,1]$ and $L > 0$. To our knowledge, the minimax rates over smoother classes of functions remains unknown and it is beyond the scope of this paper to describe them.
For the problem of testing monotonicity or convexity, other choices of distances to the null are possible. For example the distance in sup-norm between the first (respectively the second) derivative of F and the set of non-negative functions. For such choices, Dümbgen and Spokoiny also provided uniform separation rates for their tests. In the regression setting, the uniform separation rates we get coincide with their separation rates on the classes of functions they considered. We do not know whether these rates are optimal or not neither in the Gaussian white noise model nor in the regression one.

2. Tests based on local means for testing positivity, monotonicity and convexity. We consider the regression model given by (1) and propose tests of positivity, monotonicity and convexity for the function F. We first introduce some partitions of the design points and notations that will be used throughout the paper.

2.1. *Partition of the design points and notations.* We first define an almost regular partition of the set of indices $\{1,\ldots,n\}$ into ℓ_n sets as follows: for each k in $\{1,\ldots,\ell_n\}$ we set

$$J_k = \left\{ i \in \{1,\ldots,n\}, \frac{k-1}{\ell_n} < \frac{i}{n} \leq \frac{k}{\ell_n} \right\}$$

and define the partition as

$$\mathcal{J}^{\ell_n} = \{J_k, k = \{1,...,\ell_n\}\}\,.$$

Then for each $\ell \in \{1, \dots, \ell_n\}$, we make a partition of $\{1, \dots, n\}$ into ℓ sets by gathering consecutive sets J_k. This partition is defined by

$$\mathcal{J}^\ell = \left\{ J_j^\ell = \bigcup_{\frac{j-1}{\ell} < \frac{k}{\ell_n} \leq \frac{j}{\ell}} J_k,\ j = 1, \dots, \ell \right\}. \tag{8}$$

We shall use the following notations.

- We use a bold type style for denoting the vectors of $\mathbb{R}^n$. We endow $\mathbb{R}^n$ with its Euclidean norm denoted by $\| \ \|$.
- For $\boldsymbol{v} \in \mathbb{R}^n$, let $\|\boldsymbol{v}\|_\infty = \max_{1 \leq i \leq n} |v_i|$.
- For a linear subspace V of $\mathbb{R}^n$, Π_V denotes the orthogonal projector onto V.
- For $a \in \mathbb{R}_+$, $D \in \mathbb{N} \backslash \{0\}$ and $u \in [0, 1]$, $\bar{\Phi}^{-1}(u)$ and $\bar{\chi}^{-1}_{D,a^2}(u)$ denote the $1 - u$ quantile of respectively a standard Gaussian random variable and a non-central χ^2 with D degrees of freedom and non-centrality parameter a^2.
- For $x \in \mathbb{R}$, $[x]$ denotes the integer part of x.
- For each $\mathbb{R}^n$-vector $\boldsymbol{v}$ and subset J of $\{1, ..., n\}$, we denote by $\boldsymbol{v}_J$ the $\mathbb{R}^n$-vector whose coordinates coincide with those of $\boldsymbol{v}$ on J and vanish elsewhere. We denote by $\bar{\boldsymbol{v}}_J$ the quantity $\sum_{i \in J} v_i / |J|$.
- We denote by $\mathbf{1}$ the $\mathbb{R}^n$-vector $(1, ..., 1)'$ and by $\boldsymbol{e}_i$ the ith vector of the canonical basis.
- We define $V_{n,\text{cste}}$ as the linear span of $\left\{\mathbf{1}_J, J \in \mathcal{J}^{\ell_n}\right\}$. Note that the dimension of $V_{n,\text{cste}}$ equals ℓ_n.
- The vector $\boldsymbol{\varepsilon}$ denotes a standard Gaussian variable in $\mathbb{R}^n$.
- We denote by $\mathbb{P}_{\boldsymbol{f},\sigma}$ the law of the Gaussian vector in $\mathbb{R}^n$ with expectation $\boldsymbol{f}$ and covariance matrix $\sigma^2 I_n$, where I_n is the $n \times n$ identity matrix. We denote by $\mathbb{P}_{F,\sigma}$ the law of Y under the model defined by Equation (1).
- The level α of all our tests is chosen in $]0, 1/2[$.

2.2. *Test of positivity.* We propose a level α-test for testing that F belongs to $\mathcal{K}_{\geq 0}$ defined by (2). The testing procedure is based on the fact that if F is non-negative, then for any subset J of $\{1, \dots, n\}$, the expectation of $\bar{\boldsymbol{Y}}_J$ is non-negative. For $\ell \in \{1, ..., \ell_n\}$, let $T_1^\ell(\boldsymbol{Y})$ be defined as

$$T_1^\ell(\boldsymbol{Y}) = \max_{J \in \mathcal{J}^\ell} \frac{-\sqrt{|J|}\, \bar{\boldsymbol{Y}}_J}{\|\boldsymbol{Y} - \Pi_{V_{n,\text{cste}}} \boldsymbol{Y}\|} \sqrt{n - \ell_n},$$

and let $q_1(\ell, u)$ be the $1 - u$ quantile of the random variable $T_1^\ell(\boldsymbol{\varepsilon})$. We introduce the test statistic

$$T_{\alpha,1} = \max_{\ell \in \{1, \dots, \ell_n\}} \left\{ T_1^\ell(\boldsymbol{Y}) - q_1(\ell, u_\alpha) \right\}, \tag{9}$$

where u_α is defined as

$$u_\alpha = \sup\left\{u \in]0,1[, \mathbb{P}\left(\max_{\ell\in\{1,\dots,\ell_n\}} \left\{T_1^\ell(\boldsymbol{\varepsilon}) - q_1(\ell,u)\right\} > 0\right) \le \alpha\right\}. \tag{10}$$

We reject that F belongs to $\mathcal{K}_{\geq 0}$ if $T_{\alpha,1}$ is positive.
Comment. When ℓ increases from 1 to ℓ_n, the cardinality of the sets $J \in \mathcal{J}^\ell$ decreases. We thus take into account local discrepancies to the null hypothesis for various scales.

2.3. *Testing monotonicity.* We now consider the problem of testing that F belongs to $\mathcal{K}_{\nearrow}$ defined by (3). The testing procedure lies on the following property: if I and J are two subsets of $\{1,2,\dots,n\}$ such that I is on the left of J and if $F \in \mathcal{K}_{\nearrow}$, then the expectation of the difference $\bar{\boldsymbol{Y}}_I - \bar{\boldsymbol{Y}}_J$ is non-positive. For $\ell \in \{2,...,\ell_n\}$, let $T_2^\ell(\boldsymbol{Y})$ be defined as

$$T_2^\ell(\boldsymbol{Y}) = \max_{1\le i<j\le \ell} N_{ij}^\ell \frac{\bar{\boldsymbol{Y}}_{J_i^\ell} - \bar{\boldsymbol{Y}}_{J_j^\ell}}{\|\boldsymbol{Y} - \Pi_{V_{n,\mathrm{cste}}}\boldsymbol{Y}\|}\sqrt{n-\ell_n},$$

where

$$N_{ij}^\ell = \left(\frac{1}{|J_i^\ell|} + \frac{1}{|J_j^\ell|}\right)^{-1/2},$$

and let $q_2(\ell,u)$ be the $1-u$ quantile of the random variable $T_2^\ell(\boldsymbol{\varepsilon})$. We introduce the test statistic

$$T_{\alpha,2} = \max_{\ell\in\{2,\dots,\ell_n\}} \left\{T_2^\ell(\boldsymbol{Y}) - q_2(\ell,u)\right\}, \tag{11}$$

where u_α is defined as

$$u_\alpha = \sup\left\{u \in]0,1[, \mathbb{P}\left(\max_{\ell\in\{2,\dots,\ell_n\}} \left\{T_2^\ell(\boldsymbol{\varepsilon}) - q_2(\ell,u)\right\} > 0\right) \le \alpha\right\}. \tag{12}$$

We reject that F belongs to $\mathcal{K}_{\nearrow}$ if $T_{\alpha,2}$ is positive.

2.4. *Testing convexity.* We now consider the problem of testing that F belongs to $\mathcal{K}_{\smile}$ defined by (4). The testing procedure is based on the following property: if I, J and K are three subsets of $\{1,2,\dots,n\}$, such that J is between I and K and if $F \in \mathcal{K}_{\smile}$, then we find a linear combination of $\bar{\boldsymbol{Y}}_I$, $\bar{\boldsymbol{Y}}_J$, and $\bar{\boldsymbol{Y}}_K$ with non-positive expectation. Let $\boldsymbol{x} = (x_1,\dots,x_n)'$ and for each $\ell \in \{3,...,\ell_n\}$, $1 \le i < j < k \le \ell$, let

$$\lambda_{ijk}^\ell = \frac{\bar{\boldsymbol{x}}_{J_k^\ell} - \bar{\boldsymbol{x}}_{J_j^\ell}}{\bar{\boldsymbol{x}}_{J_k^\ell} - \bar{\boldsymbol{x}}_{J_i^\ell}},$$

and

$$N_{ijk}^\ell = \left(\frac{1}{|J_j^\ell|} + (\lambda_{ijk}^\ell)^2\frac{1}{|J_i^\ell|} + (1-\lambda_{ijk}^\ell)^2\frac{1}{|J_k^\ell|}\right)^{-1/2}.$$

For $\ell \in \{3, ..., \ell_n\}$, let

$$T_3^\ell(\boldsymbol{Y}) = \max_{1\le i<j<k\le \ell} N_{ijk}^\ell \frac{\bar{\boldsymbol{Y}}_{J_j^\ell} - \lambda_{ijk}^\ell \bar{\boldsymbol{Y}}_{J_i^\ell} - (1-\lambda_{ijk}^\ell)\bar{\boldsymbol{Y}}_{J_k^\ell}}{\|\boldsymbol{Y} - \Pi_{V_{n,\mathrm{cste}}}\boldsymbol{Y}\|/\sqrt{n-\ell_n}},$$

and let $q_3(\ell, u)$ be the $1-u$ quantile of the random variable $T_3^\ell(\boldsymbol{\varepsilon})$. We introduce the test statistic

$$T_{\alpha,3} = \max_{\ell\in\{3,...,\ell_n\}} \left\{T_3^\ell(\boldsymbol{Y}) - q_3(\ell, u_\alpha)\right\}, \tag{13}$$

where u_α is defined as

$$u_\alpha = \sup\left\{u \in]0,1[, \mathbb{P}\left(\max_{\ell\in\{3,...,\ell_n\}} \left\{T_3^\ell(\boldsymbol{\varepsilon}) - q_3(\ell, u)\right\} > 0\right) \le \alpha\right\}. \tag{14}$$

We reject that F belongs to $\mathcal{K}_\smile$ if $T_{\alpha,3}$ is positive.

2.5. *Properties of the procedures.* In this section we evaluate the performances of the previous procedures under the null and under smooth alternatives.

Proposition 1. *Let $(T_\alpha, \mathcal{K})$ be either $(T_{\alpha,1}, \mathcal{K}_{\ge 0})$ or $(T_{\alpha,2}, \mathcal{K}_\nearrow)$ or $(T_{\alpha,3}, \mathcal{K}_\smile)$. We have*

$$\sup_{\sigma>0}\sup_{F\in\mathcal{K}} \mathbb{P}_{F,\sigma}(T_\alpha > 0) = \alpha.$$

Assume now that $x_i = i/n$ for all $i = 1, ..., n$ and $\ell_n = [n/2]$. Let us fix $\beta \in]0,1[$ and define for each $s \in]0,1]$ and $L > 0$

$$\rho_n = L^{1/(1+2s)}\left(\frac{\sigma^2\log(n)}{n}\right)^{s/(1+2s)}.$$

Then, for n large enough there exists some constant κ depending on α, β, s only such that for all $F \in \mathcal{H}_s(L)$ satisfying

$$\Delta(F) = \inf_{G\in\mathcal{K}} \|F - G\|_\infty \ge \kappa\rho_n \tag{15}$$

we have

$$\mathbb{P}_{F,\sigma}(T_\alpha > 0) \ge 1 - \beta.$$

Comment. This result states that our procedures are of size α. Moreover, following the definition of the uniform separation rate of a test given in Section 1.3, this result shows that the tests achieve the uniform separation rate ρ_n (in sup-norm) over the Hölderian ball $\mathcal{H}_s(L)$. In the following proposition, we show that this rate cannot be improved at least in the Gaussian white noise model for testing positivity and monotonicity. The proof can be extended to the case of testing convexity but is omitted here.

PROPOSITION 2. *Let Y be the observation from the Gaussian white noise model*

$$dY(t) = F(t)dt + \frac{1}{\sqrt{n}}dW(t), \quad \text{for } t \in [0,1], \tag{16}$$

where W is a standard Brownian motion. Let $\mathcal{K}$ be either the set $\mathcal{K}_{\geq 0}$ or $\mathcal{K}_{\nearrow}$ and $\mathcal{F}$ some class of functions. For the distance $\Delta(.)$ to $\mathcal{K}$ given by (15), we define

$$\rho_n(0, \mathcal{F}) = \inf \rho(\Phi, \mathcal{F}, \Delta)$$

where $\rho(\Phi, \mathcal{F}, \Delta)$ is given by (6) and where the infimum is taken over all tests Φ of level 3α for testing "$F = 0$". We define $\rho_n(\mathcal{K}, \mathcal{F})$ similarly by taking the infimum over all tests Φ of level α for testing "$F \in \mathcal{K}$". The following inequalities hold.

(*i*) *If $\mathcal{K} = \mathcal{K}_{\geq 0}$ then*

$$\rho_n(\mathcal{K}, \mathcal{F}) \geq \rho_n(0, \mathcal{F}).$$

If $\mathcal{K} = \mathcal{K}_{\nearrow}$ then for some constant κ depending on α and β only

$$\rho_n(\mathcal{K}, \mathcal{F}) \geq \frac{1}{2}\left[\rho_n(0, \mathcal{F}) - \kappa\frac{\sigma}{\sqrt{n}}\right].$$

(*ii*) *In particular, if $\mathcal{F} = \mathcal{H}_s(L)$, for n large enough there exists some constant κ' depending on α, β and s only such that*

$$\rho_n(\mathcal{K}, \mathcal{F}) \geq \kappa' L^{1/(1+2s)} \left(\frac{\log(n)}{n}\right)^{s/(1+2s)}. \tag{17}$$

The proof of the first part of the proposition extends easily to the regression framework. The second part (*ii*), namely Inequality (17), derives from (*i*) and the lower bound on $\rho_n(0, \mathcal{F})$ established by Ingster(1993).
For the problem of testing the positivity of a signal in the Gaussian white noise model, Juditsky and Nemirovski (2000) showed that the minimax separation rate with respect to the $\mathbb{L}^r$-distance ($r \in [1, +\infty[$) is of the same order as ρ_n up to a logarithmic factor.

3. Testing that F satisfies a differential inequality. In this section, we consider the problem of testing that F belongs to $\mathcal{K}_{r,R}$ defined by (5). Several applications of such hypotheses can be of interest. For example, by taking $r = 1$ and $R(x) = -\exp(ax)$ (for some positive number a) one can test that a positive function F is decreasing at rate $\exp(-ax)$, that is, satisfies

$$\forall x \in [0,1], 0 < F(x) \leq F(0)\exp(-ax).$$

Other kinds of decay are possible by suitably choosing the function R. Another application is to test that F belongs to the class of smooth functions

$$\left\{F : [0,1] \to \mathbb{R},\ \|F^{(r)}\|_\infty \leq L\right\}.$$

To tackle this problem, it is enough to test that the derivatives of order r of the functions $F_1(x) = -F(x) + Lx^r/r!$ and $F_2(x) = F(x) + Lx^r/r!$ are positive. This is easily done by considering a multiple testing procedure based on the data $-Y_i +$

$Lx_i^r/r!$ for testing that F_1 is positive, and on $Y_i + Lx_i^r/r!$ for testing that F_2 is positive.
In Section 3.1, we consider the case where the function R equals 1. The procedure then amounts to testing that the derivative of order r of F is non-negative. We turn to the general case in Section 3.2.

We first introduce the following notations.

- For $\boldsymbol{w} \in \mathbb{R}^n$, we set $R \star \boldsymbol{w}$ the vector whose i-th coordinate $(R \star \boldsymbol{w})_i$ equals $R(x_i)w_i$.
- For $k \in \mathbb{N} \setminus \{0\}$, we denote by $\boldsymbol{w}^k$, the $\mathbb{R}^n$-vector $(w_1^k, \ldots, w_n^k)$, and we set $\boldsymbol{w}^0 = \mathbf{1}$ by convention.
- For $J \subset \{1, ..., n\}$, let us define $\mathcal{X}_J$ as the space spanned by $\mathbf{1}_J, \boldsymbol{x}_J, \cdots, \boldsymbol{x}_J^{r-1}$.

3.1. *Testing that the derivative of order r of F is non-negative.* In this section, we take $R(x) = 1$ for all $x \in [0, 1]$. The procedure lies on the idea that if the derivative of order r of F is non-negative then on each subset J of $\{1, 2, \ldots, n\}$, the highest degree coefficient of the polynomial regression of degree r based on the pairs $\{(x_i, F(x_i)), i \in J\}$ is non-negative. For example, under the assumption that F is non-decreasing, the slope of the regression based on the pairs $\{(x_i, F(x_i)), i \in J\}$ is non-negative.

Let $\ell_n = [n/(2(r+1))]$, V_n be the linear span of $\left\{\mathbf{1}_J, \boldsymbol{x}_J, \cdots, \boldsymbol{x}_J^r,\ J \in \mathcal{J}^{\ell_n}\right\}$, and for each $J \subset \{1, \ldots, n\}$

$$\boldsymbol{t}_J^* = -\frac{\boldsymbol{x}_J^r - \Pi_{\mathcal{X}_J}\boldsymbol{x}_J^r}{\|\boldsymbol{x}_J^r - \Pi_{\mathcal{X}_J}\boldsymbol{x}_J^r\|}.$$

For each $\ell \in \{1, ..., \ell_n\}$, let $T^\ell(\boldsymbol{Y})$ be defined as

$$T^\ell(\boldsymbol{Y}) = \max_{J \in \mathcal{J}^\ell} \frac{< \boldsymbol{Y}, \boldsymbol{t}_J^* >}{\|\boldsymbol{Y} - \Pi_{V_n}\boldsymbol{Y}\|}\sqrt{n - d_n} \tag{18}$$

and let $q(\ell, u)$ denote the $1-u$ quantile of the random variable $T^\ell(\boldsymbol{\varepsilon})$. We introduce the following test statistic:

$$T_\alpha = \max_{\ell \in \{1, \ldots, \ell_n\}} \left\{T^\ell(\boldsymbol{Y}) - q(\ell, u_\alpha)\right\}, \tag{19}$$

where u_α is defined as

$$u_\alpha = \sup\left\{u \in]0, 1[, \mathbb{P}\left(\max_{\ell \in \{1, \ldots, \ell_n\}} \left\{T^\ell(\boldsymbol{\varepsilon}) - q(\ell, u)\right\} > 0\right) \leq \alpha\right\}. \tag{20}$$

We reject the null hypothesis if T_α is positive.

Comment. When $r = 1$, the procedure is akin to that proposed by Hall and Heckman (2000) where for all ℓ, $q(\ell, u_\alpha)$ is the $1-\alpha$ quantile of $\max_{\ell \in \{1, \ldots, \ell_n\}} T^\ell(\boldsymbol{\varepsilon})$.

3.2. *Extension to the general case.* The ideas underlying the preceding procedures extend to the case where $R \not\equiv 1$. In the general case, the test is obtained as follows.

Let ℓ_n be such that the dimension d_n of the linear space

$$V_n = \mathrm{Span}\left\{\mathbf{1}_J, \boldsymbol{x}_J, \cdots, \boldsymbol{x}_J^r, R \star \mathbf{1}_J, ..., R \star \boldsymbol{x}_J^r, J \in \mathcal{J}^{\ell_n}\right\} \tag{21}$$

is not larger than $n/2$. We define for each $J \subset \{1, \ldots, n\}$

$$\boldsymbol{t}_J^* = -\frac{R \star (\boldsymbol{x}_J^r - \Pi_{\mathcal{X}_J} \boldsymbol{x}_J^r)}{\gamma_J} \text{ where } \gamma_J = \|R \star (\boldsymbol{x}_J^r - \Pi_{\mathcal{X}_J} \boldsymbol{x}_J^r)\|. \tag{22}$$

We reject that F belongs to $\mathcal{K}_{r,R}$ if T_α defined by (19) is positive.

3.3. *Properties of the tests.* In this section, we describe the behavior of the procedure. We start with some notations.

- Let us define the function $\Lambda(F)$ as
$$\Lambda(F)(x) = \frac{d^r}{dx^r}[R(x)F(x)],$$
and let ω be its modulus of continuity defined for all $h > 0$ by
$$\omega(h) = \sup_{|x-y| \le h} |\Lambda(F)(x) - \Lambda(F)(y)|.$$
- For $J \in \bigcup_{\ell=1}^{\ell_n} \mathcal{J}^\ell$, let us denote by x_J^- (respectively x_J^+) the quantities $\min\{x_i,\ i \in J\}$ (respectively $\max\{x_i,\ i \in J\}$) and set $h_J = x_J^+ - x_J^-$.
- Let $\boldsymbol{f} = (F(x_1), \ldots, F(x_n))'$ and for each $\ell = 1, \ldots, \ell_n$ and $\beta \in]0,1[$, let
$$\nu_\ell(\boldsymbol{f}, \beta) = \left(\frac{q(\ell, u_\alpha)}{\sqrt{n - d_n}} \sqrt{\bar{\chi}^{-1}_{n-d_n, \|\boldsymbol{f} - \Pi_{V_n}\boldsymbol{f}\|^2/\sigma^2}(\beta/2)} + \bar{\Phi}^{-1}(\beta/2)\right)\sigma. \tag{23}$$
- For each $\rho > 0$, let
$$\mathcal{E}_{n,r}(\rho) = \left\{F : [0,1] \to \mathbb{R},\ F^{(r)} \in \mathcal{H}_s(L), -\inf_{x \in [0,1]} F^{(r)}(x) \ge \rho\right\}.$$

We have the following result.

Proposition 3. *Let T_α be the test statistic defined in Section 3.2. We have*

$$\sup_{\sigma > 0} \sup_{F \in \mathcal{K}_{r,R}} \mathbb{P}_{F,\sigma}(T_\alpha > 0) = \alpha.$$

For each $\beta \in]0,1[$, we have

$$\mathbb{P}_{F,\sigma}(T_\alpha > 0) \ge 1 - \beta,$$

if for some $\ell \in \{1, \ldots, \ell_n\}$ there exists a set $J \in \mathcal{J}^\ell$ such that either

$$-\inf_{i \in J} \Lambda(F)(x_i) \ge \nu_\ell(\boldsymbol{f}, \beta) \frac{r!\gamma_J}{\|\boldsymbol{x}_J^r - \Pi_{\mathcal{X}_J}\boldsymbol{x}_J^r\|^2} + \omega(h_J), \tag{24}$$

or

$$\inf_{x\in]x_J^-,x_J^+[} -\Lambda(F)(x) \geq \nu_\ell(\boldsymbol{f},\beta)\frac{r!\gamma_J}{\|\boldsymbol{x}_J^r - \Pi_{\mathcal{X}_J}\boldsymbol{x}_J^r\|^2}. \tag{25}$$

Moreover, if $R \equiv 1$ then there exists some constant κ depending on α, β, s and r only such that for n large enough and for all $F \in \mathcal{E}_{n,r}(\rho_{n,r})$ with

$$\rho_{n,r} = \kappa \left(\frac{\sigma^2 \log(n)}{n}\right)^{s/(1+2(s+r))} L^{(1+2r)/(1+2(s+r))}$$

we have

$$\mathbb{P}_{F,\sigma}(T_\alpha > 0) \geq 1 - \beta.$$

Comments.

1. In the particular case where $R \equiv 1$, let us give the orders of magnitude of the quantities appearing in the above proposition. Under the assumption that $\|\boldsymbol{f} - \Pi_{V_n}\boldsymbol{f}\|^2/n$ is smaller than σ^2, one can show that ν_ℓ is of order $\sqrt{\log(n)}$ (see Section 9.2). When $R \equiv 1$, we have $\gamma_J = \|\boldsymbol{x}_J^r - \Pi_{\mathcal{X}_J}\boldsymbol{x}_J^r\|$ and it follows from computations that will be detailed in the proofs that

$$\nu_\ell(\boldsymbol{f},\beta)\frac{r!\gamma_J}{\|\boldsymbol{x}_J^r - \Pi_{\mathcal{X}_J}\boldsymbol{x}_J^r\|^2} \leq C\sqrt{\frac{\log(n)}{nh_J^{1+2r}}} \tag{26}$$

   for some constant C which does not depend on J nor n.

2. In the particular case where $r = 1$, Inequality (26) allows to compare our result to the performances of the test proposed by Ghosal et al. (2001). For each $\delta \in]0, 1/3[$, they give a procedure (depending on δ) that is powerful if the function F is continuously differentiable and satisfies that for all x in some interval of length $n^{-\delta}$, $F'(x) < -M\sqrt{\log(n)}n^{-(1-3\delta)/2}$ for some M large enough.

   By using (25) and the upper bound in (26) with h_J of order $n^{-\delta}$, we deduce from Proposition 3 that our procedure is powerful too over this class of functions. Note that by considering a multiple testing procedure based on various scales ℓ, our test does not depend on δ and is therefore powerful for all δ simultaneously.

3. For $r = 1$ (respectively $r = 2$) and $s = 1$, Dümbgen and Spokoiny (2001) obtained the uniform separation rate $\rho_{n,r}$ for testing monotonicity (respectively convexity) in the Gaussian white noise model.

4. For the problem of testing monotonicity ($r = 1$ and $R \equiv 1$), it is possible to combine this procedure with that proposed in Section 2.3. More precisely, consider the test which rejects the null at level 2α if one of these two tests rejects. The so-defined test performs as well as the best of these two tests under the alternative.

4. A general approach. The problems we have considered previously reduce to testing that $\boldsymbol{f} = (F(x_1), \ldots, F(x_n))'$ belongs to a convex set of the form

$$\mathcal{C} = \{\boldsymbol{f} \in \mathbb{R}^n,\ \forall j \in \{1, ..., p\}\ \ < \boldsymbol{f}, \boldsymbol{v}_j > \leq 0\}\,, \tag{27}$$

where the vectors $\{\boldsymbol{v}_1, ..., \boldsymbol{v}_p\}$ are linearly independent in $\mathbb{R}^n$. For example, testing that the regression function F is non-negative or non-decreasing amounts to testing that the mean of $\boldsymbol{Y}$ belongs respectively to the convex subsets of $\mathbb{R}^n$

$$\mathcal{C}_{\geq 0} = \{\boldsymbol{f} \in \mathbb{R}^n,\ \forall i \in \{1, ..., n\}\ f_i \geq 0\}\,. \tag{28}$$

and

$$\mathcal{C}_{\nearrow} = \{\boldsymbol{f} \in \mathbb{R}^n,\ \forall i \in \{1, ..., n-1\}\ f_{i+1} - f_i \geq 0\}\,. \tag{29}$$

Clearly, these sets are of the form given by (27) by taking respectively $p = n$, $\boldsymbol{v}_j = -\boldsymbol{e}_j$ and $p = n-1$, $\boldsymbol{v}_j = \boldsymbol{e}_j - \boldsymbol{e}_{j+1}$. The following proposition extends this result to the general case. Note that one can also define the set $\mathcal{C}$ as

$$\mathcal{C} = \{\boldsymbol{f} \in \mathbb{R}^n,\ L_1(\boldsymbol{f}) \geq 0, \ldots, L_p(\boldsymbol{f}) \geq 0\}\,,$$

where the L_i's are p independent linear forms. We shall use this definition of $\mathcal{C}$ in the following.

Proposition 4. *For each $r \in \{1, \ldots, n-1\}$ and $i \in \{1, ..., n-r\}$ let $\phi_{i,r}$ be the linear form defined for $\boldsymbol{w} \in \mathbb{R}^n$ by*

$$\phi_{i,r}(\boldsymbol{w}) = \det \begin{pmatrix} 1 & x_i & \cdots & x_i^{r-1} & w_i \\ 1 & x_{i+1} & \cdots & x_{i+1}^{r-1} & w_{i+1} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_{i+r} & \cdots & x_{i+r}^{r-1} & w_{i+r} \end{pmatrix}.$$

If F belongs to $\mathcal{K}_{\smile}$ then $\boldsymbol{f} = (F(x_1), \ldots, F(x_n))'$ belongs to

$$\mathcal{C}_{\smile} = \{\boldsymbol{f} \in \mathbb{R}^n,\ \forall i \in \{1, ..., n-2\},\ \phi_{i,2}(\boldsymbol{f}) \geq 0\}\,. \tag{30}$$

If F belongs to $\mathcal{K}_{r,R}$ then $\boldsymbol{f}$ belongs to

$$\mathcal{C}_{r,R} = \{\boldsymbol{f} \in \mathbb{R}^n,\ \forall i \in \{1, ..., n-r\},\ \phi_{i,r}(R \star \boldsymbol{f}) \geq 0\}\,.$$

In view of keeping our notations as simple as possible, we omit the dependency of the linear forms $\phi_{i,r}$ with respect to r when there is no ambiguity. The remaining part of the section is organized as follows. In the next subsection, we present a general approach for the problem of testing that $\boldsymbol{f}$ belongs to $\mathcal{C}$. In the last subsection, we show how this approach applies to the problems of hypothesis testing considered in Sections 2 and 3.

4.1. *Testing that $\boldsymbol{f}$ belongs to $\mathcal{C}$.* We consider the problem of testing that the vector $\boldsymbol{f} = (f_1, \ldots, f_n)'$ involved in the regression model

$$Y_i = f_i + \sigma\varepsilon_i,\ i = 1, ..., n. \tag{31}$$

belongs to $\mathcal{C}$ defined by (27). Our aim is twofold. First, build a test which achieves its nominal level, and second, describe for each n a class of vectors over which this test is powerful.

The testing procedure. The testing procedure relies on the following idea: since under the assumption that $\boldsymbol{f}$ belongs to $\mathcal{C}$, the quantities $< \boldsymbol{f}, \sum_{j=1}^{p} \lambda_j \boldsymbol{v}_j >$ are non-positive for all non-negative numbers $\lambda_1, \cdots, \lambda_p$, we base our test statistic on random variables of the form $< \boldsymbol{Y}, \sum_{j=1}^{p} \lambda_j \boldsymbol{v}_j >$ for non-negative sequences of λ_j's. We denote by $\mathcal{T}$ the subset of $\mathbb{R}^n$ defined by

$$\mathcal{T} = \left\{ \boldsymbol{t} = \sum_{j=1}^{p} \lambda_j \boldsymbol{v}_j,\ \|\boldsymbol{t}\| = 1,\ \lambda_j \geq 0, \forall j = 1, ..., p \right\}. \tag{32}$$

Let $\mathcal{T}_n$ be a finite subset of $\mathcal{T}$ such that there exists some linear space V_n, with dimension $d_n < n$ containing the linear span of $\mathcal{T}_n$. Let $\{q_{\boldsymbol{t}}(\alpha), \boldsymbol{t} \in \mathcal{T}_n\}$ be a sequence of numbers satisfying

$$\mathbb{P}\left[\sup_{\boldsymbol{t} \in \mathcal{T}_n} \left(\sqrt{n - d_n} \frac{< \boldsymbol{\varepsilon}, \boldsymbol{t} >}{\|\boldsymbol{\varepsilon} - \Pi_{V_n} \boldsymbol{\varepsilon}\|} - q_{\boldsymbol{t}}(\alpha) \right) > 0 \right] = \alpha. \tag{33}$$

We reject the null hypothesis if the statistic

$$T_\alpha = \sup_{\boldsymbol{t} \in \mathcal{T}_n} \left(\sqrt{n - d_n} \frac{< \boldsymbol{Y}, \boldsymbol{t} >}{\|\boldsymbol{Y} - \Pi_{V_n} \boldsymbol{Y}\|} - q_{\boldsymbol{t}}(\alpha) \right) \tag{34}$$

is positive.

Properties of the test. For all $\beta \in]0, 1[$ and each $\boldsymbol{t} \in \mathcal{T}_n$ let

$$v_{\boldsymbol{t}}(\boldsymbol{f}, \beta) = \left(q_{\boldsymbol{t}}(\alpha) \frac{1}{\sqrt{n - d_n}} \sqrt{\bar{\chi}^{-1}_{n-d_n, \|\boldsymbol{f} - \Pi_{V_n} \boldsymbol{f}\|^2/\sigma^2}(\beta/2)} + \bar{\Phi}^{-1}(\beta/2) \right) \sigma. \tag{35}$$

The order of magnitude of $v_{\boldsymbol{t}}(\boldsymbol{f}, \beta)$ is proved to be $\sqrt{\log(n)}\sigma$ under the assumption that $\|f - \Pi_{V_n} f\|^2/n$ is smaller than σ^2 as it is shown in the proof of Proposition 1. We have the following result.

THEOREM 1. *Let T_α be the test statistic defined by (34). We have*

$$\sup_{\sigma > 0} \sup_{\boldsymbol{f} \in \mathcal{C}} \mathbb{P}_{\boldsymbol{f}, \sigma} \left(T_\alpha > 0 \right) = \mathbb{P}_{0,1} \left(T_\alpha > 0 \right) = \alpha. \tag{36}$$

Moreover, if there exists $\boldsymbol{t} \in \mathcal{T}_n$ such that $< \boldsymbol{f}, \boldsymbol{t} > \geq v_{\boldsymbol{t}}(\boldsymbol{f}, \beta)$ then

$$\mathbb{P}_{\boldsymbol{f}, \sigma}(T_\alpha > 0) \geq 1 - \beta.$$

Comments. The values of the $q_{\boldsymbol{t}}(\alpha)$'s that satisfy (33) can be easily obtained by simulations under $\mathbb{P}_{0,1}$. This property of our procedure lies in the fact that the least-favorable distribution under the null is $\mathbb{P}_{0,1}$. Note that we do not need to use bootstrap procedures to implement the test.

4.2. *How to apply these procedures to test qualitative hypotheses?* In the sequel, we give the choices of $\mathcal{T}_n$ and V_n leading to the tests presented in Sections 2 and 3.

For the test of positivity described in Section 2.2. We take $\mathcal{T}_n = \mathcal{T}_{n,1}$, with $\mathcal{T}_{n,1} = \bigcup_{\ell=1}^{\ell_n} \mathcal{T}_{n,1}^{\ell}$, where for all $\ell \in \{1, ..., \ell_n\}$

$$\mathcal{T}_{n,1}^{\ell} = \left\{ -\frac{1}{\sqrt{|J|}} \sum_{j \in J} \boldsymbol{e}_j, \ J \in \mathcal{J}^{\ell} \right\}.$$

We take $V_n = V_{n,\text{cste}}$. Note that $V_{n,\text{cste}}$ is also the linear span of $\mathcal{T}_{n,1}$.

For the test of monotonicity described in Section 2.3. Let us define for each $\ell \in \{2, ..., \ell_n\}$ and $1 \leq i < j \leq \ell$,

$$\boldsymbol{e}_{ij}^{\ell} = N_{ij}^{\ell} \left(\frac{1}{|J_i^{\ell}|} \sum_{l \in J_i^{\ell}} \boldsymbol{e}_l - \frac{1}{|J_j^{\ell}|} \sum_{l \in J_j^{\ell}} \boldsymbol{e}_l \right). \tag{37}$$

Note that N_{ij}^{ℓ} is such that $\|\boldsymbol{e}_{ij}^{\ell}\| = 1$. We take $\mathcal{T}_n = \mathcal{T}_{n,2}$, with $\mathcal{T}_{n,2} = \bigcup_{\ell=2}^{\ell_n} \mathcal{T}_{n,2}^{\ell}$, where

$$\mathcal{T}_{n,2}^{\ell} = \left\{ \boldsymbol{e}_{ij}^{\ell}, \ 1 \leq i < j \leq \ell \right\},$$

and we take $V_n = V_{n,\text{cste}}$. Note that V_n contains $\mathcal{T}_{n,2}$.

For the test of convexity presented in Section 2.4. Let us define for each $\ell \in \{3, ..., \ell_n\}$, $1 \leq i < j < k \leq \ell$,

$$\boldsymbol{e}_{ijk}^{\ell} = N_{ijk}^{\ell} \left(\frac{1}{|J_j^{\ell}|} \sum_{l \in J_j^{\ell}} \boldsymbol{e}_l - \lambda_{ijk}^{\ell} \frac{1}{|J_i^{\ell}|} \sum_{l \in J_i^{\ell}} \boldsymbol{e}_l - (1 - \lambda_{ijk}^{\ell}) \frac{1}{|J_k^{\ell}|} \sum_{l \in J_k^{\ell}} \boldsymbol{e}_l \right). \tag{38}$$

Note that N_{ijk}^{ℓ} is such that $\|\boldsymbol{e}_{ijk}^{\ell}\| = 1$. We take $\mathcal{T}_n = \mathcal{T}_{n,3}$, with $\mathcal{T}_{n,3} = \bigcup_{\ell=3}^{\ell_n} \mathcal{T}_{n,3}^{\ell}$, where

$$\mathcal{T}_{n,3}^{\ell} = \left\{ \boldsymbol{e}_{ijk}^{\ell}, \ 1 \leq i < j < k \leq \ell \right\},$$

and we take $V_n = V_{n,\text{cste}}$. Note that V_n contains $\mathcal{T}_{n,3}$.

For the test of $F \in \mathcal{K}_{r,R}$ presented in Section 3. We take

$$\mathcal{T}_{n,4} = \bigcup_{\ell=1}^{\ell_n} \mathcal{T}_n^{\ell} \ \text{ where } \ \mathcal{T}_n^{\ell} = \left\{ \boldsymbol{t}_J^*, \ J \in \mathcal{J}^{\ell} \right\}$$

and $V_n = V_{n,4}$ defined by (21). Note that V_n contains $\mathcal{T}_{n,4}$.

We justify these choices of $\mathcal{T}_n$ by the following proposition proved in Section 7.

PROPOSITION 5. *Let $\mathcal{C}$ and $\mathcal{T}_n$ be either $(\mathcal{C}_{\geq 0}, \mathcal{T}_{n,1})$, $(\mathcal{C}_{\nearrow}, \mathcal{T}_{n,2})$, $(\mathcal{C}_{\smile}, \mathcal{T}_{n,3})$ or $(\mathcal{C}_{r,R}, \mathcal{T}_{n,4})$. There exist $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_p$ for which $\mathcal{C}$ is of the form given by (27) and for which $\mathcal{T}$, defined by (32), contains $\mathcal{T}_n$.*

5. Simulation studies . In this section we describe how to implement the test for testing $F \in \mathcal{K}_{\nearrow}$ and we carry out a simulation study in order to evaluate the performances of our tests both when the errors are Gaussian and when they are not. We first describe how the testing procedure is performed, then we present the simulation experiment and finally discuss the results of the simulation study.

5.1. *The testing procedures.* We carry out the simulation study for the two testing procedures described in Sections 2.3 and 3.1. In the sequel, the procedure based on differences of local means and described in Section 2.3 is called LM and the procedure based on *local gradients* defined below (from the test statistic given in Section 3.1 with $r = 1$) is called LG.
In the case of the procedure LM, we set $T_{\mathrm{LM}} = T_{\alpha,2}$ defined in (11). For each ℓ, the quantiles $q_2(\ell, u_\alpha)$ are calculated as follows. For u varying among a suitable grid of values $u_1, \ldots, u_m$, we estimate by simulations the quantity

$$p(u_j) = \mathbb{P}\left(\max_{l=1,\ldots,\ell_n} \left\{T_2^\ell(\boldsymbol{\varepsilon}) - q_2(\ell, u_j)\right\} > 0\right),$$

$\boldsymbol{\varepsilon}$ being a n-sample of $\mathcal{N}(0,1)$, and we take u_α as $\max\{u_j, p(u_j) \leq \alpha\}$. Note that u_α does not depend on $(x_i, i = 1, \ldots, n)$, but only on the number of observations n.
In the case of the procedure LG, the test statistic is defined as follows. For each $\ell = 1, \ldots, \ell_n$ and for $J \in \mathcal{J}^\ell$, we take

$$\boldsymbol{t}_J^* = \frac{\bar{x}_J \mathbf{1}_J - \boldsymbol{x}_J}{\|\bar{x}_J \mathbf{1}_J - \boldsymbol{x}_J\|}.$$

The space V_n reduces to $V_{n,\mathrm{lin}}$ the linear space of dimension $2\ell_n$ generated by

$$\left\{\mathbf{1}_J, \boldsymbol{x}_J, J \in \mathcal{J}^{\ell_n}\right\}.$$

The test statistic T_α takes the form

$$T_{\mathrm{LG}} = T_{\alpha,4} = \max_{\ell=1,\ldots,\ell_n} \left\{T_4^\ell(\boldsymbol{Y}) - q_4(\ell, u_\alpha)\right\},$$

where for each $\ell \in \{1, \ldots, \ell_n\}$,

$$T_4^\ell(\boldsymbol{Y}) = \max_{J \in \mathcal{J}^\ell} \sqrt{n - 2\ell_n} \frac{< \boldsymbol{Y}, \boldsymbol{t}_J^* >}{\|\boldsymbol{Y} - \Pi_{V_{n,\mathrm{lin}}} \boldsymbol{Y}\|},$$

and $q_4(\ell, u_\alpha)$ denotes the $1 - u$ quantile of the random variable $T_4^\ell(\boldsymbol{\varepsilon})$.
The procedure for calculating $q_4(\ell, u_\alpha)$ for $\ell = 2, \ldots, \ell_n$ is the same as the procedure for calculating the $q_2(\ell, u_\alpha)$'s.

5.2. *The simulation experiment.* The number of observations n equals 100, $x_i = i/(n+1)$, for $i = 1, \ldots, n$ and ℓ_n is either equal to 15 or 25.
We consider three distributions of the errors ε_i, with expectation zero and variance 1.

1. The Gaussian distribution: $\varepsilon_i \sim \mathcal{N}(0,1)$.

2. The type I distribution: ε_i has density $sf_X(\mu + sx)$ where $f_X(x) = \exp\{-x - \exp(-x)\}$ and where μ and s^2 are the expectation and the variance of a variable X with density f_X. This distribution is asymmetrical.

3. The mixture of Gaussian distributions: ε_i is distributed as $\pi X_1 + (1-\pi)X_2$ where π is distributed as a Bernoulli variable with expectation 0.9, X_1 and X_2 are centered Gaussian variables with variance respectively equal to $2.43s$ and $25s$, π, X_1 and X_2 are independent. The quantity s is chosen such that the variance of ε_i equals 1. This distribution has heavy tails.

We consider several functions F that are presented below. For each of them, we simulate the observations $Y_i = F(x_i) + \sigma\varepsilon_i$. The values of σ^2 and of the distance in sup-norm between F and $\mathcal{K}_{\nearrow}$ are reported in Table 1:

$$d_\infty(F, \mathcal{K}_{\nearrow}) = \frac{1}{2} \sup_{0 \le s \le t \le 1} (F(s) - F(t)).$$

Let us comment the choice of the considered functions.

- $F_0(x) = 0$ corresponds to the case for which the quantiles $q(\ell, u_\alpha)$ are calculated.

- The function $F_1(x) = 15\mathbf{1}_{x \le 0.5}(x-0.5)^3 + 0.3(x-0.5) - \exp(-250(x-0.25)^2)$ presents a strongly increasing part with a pronounced dip around $x = 1/4$ followed by a nearly flat part on the interval $[1/2, 1]$.

- The decreasing linear function $F_2(x) = -ax$, the parameter a being chosen such that $a = 1.5\sigma$.

- The function $F_3(x) = -0.2\exp(-50(x-0.5)^2)$ deviates from F_0 by a smooth dip while the function $F_4(x) = 0.1\cos(6\pi x)$ deviates from F_0 by a cosine function.

- The functions $F_5(x) = 0.2x + F_3(x)$ and $F_6(x) = 0.2x + F_4(x)$ deviate from an increasing linear function in the same way as F_3 and F_4 do from F_0.

Let us mention that it is more difficult to detect that F_5 (respectively F_6) is non-increasing than to detect that F_3 (respectively F_4) is. Indeed, adding an increasing function to a function F reduces the distance in sup-norm between F and $\mathcal{K}_{\nearrow}$. This is the reason why the values of σ are smaller in the simulation study when we consider the functions F_5 and F_6.
In Figure 1 we have displayed the functions F_ℓ for $\ell = 1, \ldots, 6$ and for each of them one sample simulated with Gaussian errors. The corresponding values of the test statistics T_{LM} and T_{LG} for $\alpha = 5\%$ and $\ell_n = 25$ are given. For this simulated sample, it appears that the test based on the statistic T_{LM} leads to reject the null hypothesis in all cases, while the test based on T_{LG} reject in all cases except for functions F_2 and F_4.
The results of the simulation experiment based on 4000 simulations are presented in Tables 2 and 3.

F	σ^2	$d_\infty(F, \mathcal{K}_\nearrow)$
$F_0(x)$	0.01	0
$F_1(x)$	0.01	0.25
$F_2(x)$	0.01	0.073
$F_3(x)$	0.01	0.1
$F_4(x)$	0.01	0.1
$F_5(x)$	0.004	0.06
$F_6(x)$	0.006	0.08

TABLE 1

Testing monotonicity : simulated functions F, values of σ^2 and distance in sup-norm between F and $\mathcal{K}_\nearrow$.

FIG. 1. *For each function $F_\ell, \ell = 1, \ldots, 6$, the simulated data $Y_i = F_\ell(x_i) + \sigma\varepsilon_i$ for $i = 1, \ldots n$ are displayed. The errors ε_i are Gaussian normalized centered variables. The value of the test statistics T_{LM} and T_{LG}, with $\alpha = 5\%$, are given for each example.*

	$\ell_n = 15$		$\ell_n = 25$	
Errors Distribution	T_{LM}	T_{LG}	T_{LM}	T_{LG}
Gaussian	0.049	0.050	0.046	0.051
Type I	0.048	0.072	0.064	0.085
Mixture	0.064	0.117	0.093	0.180

TABLE 2

Testing monotonicity : levels of the tests based on T_{LM} and T_{LG}.

	$\ell_n = 15$		$\ell_n = 25$	
F	T_{LM}	T_{LG}	T_{LM}	T_{LG}
F_1	0.85	0.99	0.99	1.
F_2	0.96	0.96	0.99	0.99
F_3	0.99	0.73	1	0.98
F_4	0.89	0.71	0.99	0.94
F_5	0.99	0.69	0.99	0.87
F_6	0.87	0.79	0.98	0.93

TABLE 3

Testing monotonicity : powers of the tests based on T_{LM} and T_{LG} when the errors are Gaussian.

5.3. *Comments on the simulation study.* As expected, the estimated level of the test calculated for the function $F_0(x) = 0$ is (nearly) equal to α when the errors are distributed as Gaussian variables.
When $\ell_n = 25$, the estimated levels of the tests for the Mixture and Type I errors distributions are greater than α (see Table 2). Let us recall that when ℓ_n is large, we are considering statistics based on the average of the observations on sets J with small cardinality. Therefore, reducing ℓ_n improves the robustness to non-Gaussian errors distribution. This is what we get in Table 2 for $\ell_n = 15$. It also appears that the method based on the local means is more robust than the method based on the local gradients, and that both methods are more robust for the Type I distribution that is asymmetric but not heavy tailed, than for the Mixture distribution.
Except for the function F_1, the estimated power is greater for the procedure based on the local means than for the procedure based on the local gradients (see Table 3). For both procedures the power of the test is larger with $\ell_n = 25$ than with $\ell_n = 15$. However, except for the function F_1, the loss of power is less significant for the procedure based on the local means.

5.4. *Comparison with other works.* As expected, the power of our procedure T_{LG} for the function F_1 is similar to that obtained by Hall and Heckman (2000).
The decreasing linear function $F_2(x) = -ax$ has already been studied by Gijbels et al. (2000) with $a = 3\sigma$. They get an estimated power of 77%.
Gijbels et al. (2000) studied the function $0.075F_3/0.2$ with $\sigma = 0.025$ and obtained a simulated power of 98%. With the same function and the same σ, we get a power equal to one, for both procedures and for $\ell_n = 15$ and $\ell_n = 25$.
Gijbels et al. (2000) and Hall and Heckman (2000) calculated the power of their test for the function $F_7(x) = 1 + x - a\exp(-50(x - 0.5)^2)$ for different values of a and σ. When $a = 0.45$ and $\sigma = 0.05$, we get a power equal to 1 as Gijbels et al do. When $a = 0.45$ and $\sigma = 0.1$, we get a power equal to 76% when using the procedure T_{LM} with $\ell_n = 25$ or $\ell_n = 15$. Gijbels et al. (2000) got 80% and Hall and Heckman (2000) a power larger than 87%.

6. Proof of Theorem 1

Level of the test: we first prove that for all $\boldsymbol{t} \in \mathcal{T}_n$, $q_{\boldsymbol{t}}(\alpha) > 0$. Indeed, thanks to (33), we have

$$\begin{aligned}
&\mathbb{P}\left[\sqrt{n-d_n}\frac{<\boldsymbol{\varepsilon},\boldsymbol{t}>}{\|\boldsymbol{\varepsilon}-\Pi_{V_n}\boldsymbol{\varepsilon}\|} - q_{\boldsymbol{t}}(\alpha) > 0\right] \\
&\leq \mathbb{P}\left[\sup_{\boldsymbol{t}\in\mathcal{T}_n}\left(\sqrt{n-d_n}\frac{<\boldsymbol{\varepsilon},\boldsymbol{t}>}{\|\boldsymbol{\varepsilon}-\Pi_{V_n}\boldsymbol{\varepsilon}\|} - q_{\boldsymbol{t}}(\alpha)\right) > 0\right] \\
&\leq \alpha < 1/2.
\end{aligned}$$

Since the random variable $\sqrt{n-d_n} < \boldsymbol{\varepsilon},\boldsymbol{t} > / \|\boldsymbol{\varepsilon} - \Pi_{V_n}\boldsymbol{\varepsilon}\|$ is symmetric (distributed as a Student with $n - d_n$ degrees of freedom) we deduce that $q_{\boldsymbol{t}}(\alpha)$ is positive. In

the sequel let us set

$$\hat{\sigma}_n = \|\boldsymbol{Y} - \Pi_{V_n}\boldsymbol{Y}\|/\sqrt{n-d_n}.$$

Since for all $\boldsymbol{f} \in \mathcal{C}$ and $j \in \{1, ..., p\}$, $< \boldsymbol{f}, v_j > \leq 0$ we have that for all $t \in \mathcal{T}_n$,

$$< \boldsymbol{f}, \boldsymbol{t} > = \sum_{j=1}^{p} \frac{\lambda_j < \boldsymbol{f}, v_j >}{\| \sum_{j=1}^{p} \lambda_j v_j \|} \leq 0.$$

Hence, $< \boldsymbol{Y}, \boldsymbol{t} > = < \boldsymbol{f}, \boldsymbol{t} > + \sigma < \boldsymbol{\varepsilon}, \boldsymbol{t} > \leq \sigma < \boldsymbol{\varepsilon}, \boldsymbol{t} >$ and therefore for all $\boldsymbol{f} \in \mathcal{C}$ and $\sigma > 0$,

$$\begin{aligned}\mathbb{P}_{\boldsymbol{f},\sigma}\left[T_\alpha > 0\right] &\leq \mathbb{P}_{\boldsymbol{f},\sigma}\left[\sup_{\boldsymbol{t}\in\mathcal{T}_n}\left(\frac{< \boldsymbol{\varepsilon}, \boldsymbol{t} >}{\hat{\sigma}_n/\sigma} - q_{\boldsymbol{t}}(\alpha)\right) > 0\right] \\ &\leq \mathbb{P}_{\boldsymbol{f},\sigma}\left[\frac{\hat{\sigma}_n}{\sigma} < \sup_{\boldsymbol{t}\in\mathcal{T}_n} \frac{< \boldsymbol{\varepsilon}, \boldsymbol{t} >}{q_{\boldsymbol{t}}(\alpha)}\right].\end{aligned}$$

We now use the following lemma for non-central χ^2-random variables:

Lemma 1. *For all $u > 0$, $\boldsymbol{f} \in \mathbb{R}^n$ and $\sigma > 0$*

$$\mathbb{P}_{\boldsymbol{f},\sigma}\left[\hat{\sigma}_n < \sigma u\right] \leq \mathbb{P}_{0,1}\left[\hat{\sigma}_n < u\right].$$

This lemma states that a non-central χ^2-random variable is stochastically larger than a χ^2-random variable with the same degrees of freedom. For a proof we refer to Lemma 1 in Baraud, Huet and Laurent (2003).
Since $\mathcal{T}_n \subset V_n$, the random variables $< \boldsymbol{\varepsilon}, \boldsymbol{t} >$ for $\boldsymbol{t} \in \mathcal{T}_n$ are independent of $\hat{\sigma}_n$ and thus by conditioning with respect to the $< \boldsymbol{\varepsilon}, \boldsymbol{t} >$'s and using Lemma 1 we get

$$\begin{aligned}\sup_{\sigma>0}\sup_{\boldsymbol{f}\in\mathcal{C}} \mathbb{P}_{\boldsymbol{f},\sigma}\left[T_\alpha > 0\right] &\leq \mathbb{P}_{0,1}\left[\hat{\sigma}_n < \sup_{\boldsymbol{t}\in\mathcal{T}_n} \frac{< \boldsymbol{\varepsilon}, \boldsymbol{t} >}{q_{\boldsymbol{t}}(\alpha)}\right] \\ &= \mathbb{P}_{0,1}\left[T_\alpha > 0\right] = \alpha.\end{aligned}$$

The reverse inequality being obvious, this concludes the proof of (36).

Power of the test: For any $\boldsymbol{f} \in \mathbb{R}^n$ and $\sigma > 0$

$$\mathbb{P}_{\boldsymbol{f},\sigma}\left(T_\alpha \leq 0\right) = \mathbb{P}_{\boldsymbol{f},\sigma}\left(\forall \boldsymbol{t} \in \mathcal{T}_n, < \boldsymbol{Y}, \boldsymbol{t} > \leq q_{\boldsymbol{t}}(\alpha)\hat{\sigma}_n\right).$$

Setting

$$x_n(\boldsymbol{f}, \beta) = \frac{\sigma}{\sqrt{n-d_n}}\sqrt{\bar{\chi}^{-1}_{n-d_n, \|\boldsymbol{f}-\Pi_{V_n}\boldsymbol{f}\|^2/\sigma^2}(\beta/2)},$$

we have

$$\mathbb{P}_{\boldsymbol{f},\sigma}\left(\hat{\sigma}_n > x_n(\boldsymbol{f}, \beta)\right) = \beta/2.$$

It follows that for all $\boldsymbol{f} \in \mathbb{R}^n$ and $\sigma > 0$,

$$\begin{aligned}\mathbb{P}_{\boldsymbol{f},\sigma}\left(T_\alpha \leq 0\right) &\leq \inf_{t\in\mathcal{T}_n} \mathbb{P}_{\boldsymbol{f},\sigma}\left(< \boldsymbol{Y}, \boldsymbol{t} > \leq q_{\boldsymbol{t}}(\alpha)x_n(\boldsymbol{f}, \beta)\right) + \beta/2 \\ &\leq \inf_{t\in\mathcal{T}_n} \mathbb{P}_{\boldsymbol{f},\sigma}\left(\sigma < \boldsymbol{\varepsilon}, \boldsymbol{t} > \leq q_{\boldsymbol{t}}(\alpha)x_n(\boldsymbol{f}, \beta) - < \boldsymbol{f}, \boldsymbol{t} >\right) + \beta/2.\end{aligned}$$

Since $\|t\| = 1$, $< \boldsymbol{\varepsilon}, \boldsymbol{t} >$ is distributed as a standard Gaussian variable, and therefore $\mathbb{P}_{\boldsymbol{f},\sigma}(T_\alpha \leq 0) \leq \beta$ as soon as there exists $t \in \mathcal{T}_n$ such that

$$q_{\boldsymbol{t}}(\alpha)x_n(\boldsymbol{f},\beta) - < \boldsymbol{f}, \boldsymbol{t} > \leq -\sigma\bar{\Phi}^{-1}(\beta/2).$$

This concludes the proof of Theorem 1.

7. Proof of Propositions 4 and 5 Let us denote by $\mathcal{I}_r$ the set of increasing sequences of $r+1$ indices in $\{1, ..., n\}$ that is

$$\mathcal{I}_r = \{(i_1, ..., i_{r+1}),\ i_1 < \cdots < i_{r+1},\ i_j \in \{1, ..., n\}\}. \tag{39}$$

For $\boldsymbol{i} = (i_1, ..., i_{r+1}) \in \mathcal{I}_r$ and $\boldsymbol{v} \in \mathbb{R}^n$ we set

$$\phi_{\boldsymbol{i}}(\boldsymbol{v}) = \det \begin{pmatrix} 1 & x_{i_1} & \cdots & x_{i_1}^{r-1} & v_{i_1} \\ 1 & x_{i_2} & \cdots & x_{i_2}^{r-1} & v_{i_2} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_{i_{r+1}} & \cdots & x_{i_{r+1}}^{r-1} & v_{i_{r+1}} \end{pmatrix}. \tag{40}$$

For $\boldsymbol{i} = (i, ..., i+r)$ $\phi_{\boldsymbol{i}}(\boldsymbol{v}) = \phi_i(\boldsymbol{v})$ where $\phi_i(\boldsymbol{v})$ is defined by Equation (30). For $\boldsymbol{w}^1, ..., \boldsymbol{w}^q$ q vectors of $\mathbb{R}^n$, we set

$$\mathrm{Gram}(\boldsymbol{w}^1, ..., \boldsymbol{w}^q) = \det(G) \text{ where } G = (< \boldsymbol{w}^i, \boldsymbol{w}^j >)_{1 \leq i,j \leq q}.$$

Let us define

$$\tilde{\mathcal{C}}_{r,R} = \{\boldsymbol{f} \in \mathbb{R}^n,\ \forall \boldsymbol{i} \in \mathcal{I}_r,\ \phi_{\boldsymbol{i}}(R \star \boldsymbol{f}) \geq 0\}. \tag{41}$$

The proofs of Propositions 4 and 5 rely on the following lemma.

Lemma 2. *The following equalities hold. First,*

$$\tilde{\mathcal{C}}_{r,R} = \mathcal{C}_{r,R}. \tag{42}$$

Assume that $\boldsymbol{f} = (F(x_1), ..., F(x_n))'$ where F is such that RF is r-th times differentiable. Then for each $\boldsymbol{i} \in \mathcal{I}_r$ there exists some $c_{\boldsymbol{i}} \in]x_{i_1}, x_{i_{r+1}}[$ such that

$$\phi_{\boldsymbol{i}}(R \star \boldsymbol{f}) = \frac{\Lambda(F)(c_{\boldsymbol{i}})}{r!}\phi_{\boldsymbol{i}}(\boldsymbol{x}^r). \tag{43}$$

For $J \subset \{1, \ldots, n\}$ let $\boldsymbol{t}_J^*$ be defined by (22). We have

$$- < \boldsymbol{f}, \boldsymbol{t}_J^* > = N_J^{-1} \sum_{\boldsymbol{i} \in \mathcal{I}_r \cap J^{r+1}} \phi_{\boldsymbol{i}}(R \star \boldsymbol{f})\phi_{\boldsymbol{i}}(\boldsymbol{x}^r). \tag{44}$$

where $N_J = \mathrm{Gram}(\mathbf{1}_J, \boldsymbol{x}_J, ..., \boldsymbol{x}_J^{r-1})\gamma_J$.

The proof of the lemma is delayed to the Appendix.

7.1. *Proof of Proposition 4.* The result concerning $\mathcal{K}_{\smile}$ is clear as a function F is non-concave on $[0,1]$ if and only if for all x, y, z in $[0,1]$ with $x < y < z$ one has

$$\det \begin{pmatrix} 1 & x & F(x) \\ 1 & y & F(y) \\ 1 & z & F(z) \end{pmatrix} \geq 0.$$

Let us now turn to the set $\mathcal{K}_{r,R}$. First note that the $n-r$ linear forms $\boldsymbol{f} \mapsto \phi_{i,r}(R\star\boldsymbol{f})$ are independent since the linear space

$$\{\boldsymbol{f} \in \mathbb{R}^n,\ \forall i \in \{1, ..., n-r\}\ \phi_{i,r}(R \star \boldsymbol{f}) = 0\},$$

which is generated by

$$\frac{1}{R} \star \mathbf{1}, \frac{1}{R} \star \boldsymbol{x}, ..., \frac{1}{R} \star \boldsymbol{x}^{r-1},$$

is of dimension r. Second, the fact that $\boldsymbol{f}$ belongs to $\tilde{\mathcal{C}}_{r,R}$ is a straightforward consequence of (43) since under the assumption that $F \in \mathcal{K}_{r,R}$, $\Lambda(F)(x) \geq 0$ for all x, and since the Vandermond determinants $\phi_{\boldsymbol{i}}(\boldsymbol{x}^r)$ are positive for all $\boldsymbol{i} \in \mathcal{I}_r$.

7.2. *Proof of Proposition 5.* The result is clear in the case where $\mathcal{C}_{\geq 0}$. For the other cases we use the following lemma.

Lemma 3. *Let W be the orthogonal of the linear space generated by the $\boldsymbol{v}_j$'s for $j = 1, ..., p$. If $\boldsymbol{t}^* \notin W$ satisfies for all $\boldsymbol{f} \in \mathcal{C}$*

$$< \boldsymbol{t}^* - \Pi_W \boldsymbol{t}^*, \boldsymbol{f} > \leq 0$$

then

$$\frac{\boldsymbol{t}^* - \Pi_W \boldsymbol{t}^*}{\|\boldsymbol{t}^* - \Pi_W \boldsymbol{t}^*\|} \in \mathcal{T}.$$

Proof. The vector $\boldsymbol{t}^* - \Pi_W \boldsymbol{t}^*$ belongs to the linear space generated by the $\boldsymbol{v}_j$'s and thus one can write $\boldsymbol{g}^* = \boldsymbol{t}^* - \Pi_W \boldsymbol{t}^* = \sum_{j=1}^p \lambda_j \boldsymbol{v}_j$. It remains to show that the λ_j's are non-negative. Let us fix $j_0 \in \{1, ..., p\}$ and choose $\boldsymbol{f}^{j_0}$ in $\mathbb{R}^n$ satisfying $< \boldsymbol{f}^{j_0}, \boldsymbol{v}_j >= 0$ for all $j \neq j_0$ and $< \boldsymbol{f}^{j_0}, \boldsymbol{v}_{j_0} >< 0$. Such a vector exists since the $\boldsymbol{v}_j$'s are linearly independent in $\mathbb{R}^n$. Clearly $\boldsymbol{f}^{j_0}$ belongs to $\mathcal{C}$ and therefore $< \boldsymbol{f}^{j_0}, \boldsymbol{g}^* >= \lambda_{j_0} < \boldsymbol{f}^{j_0}, \boldsymbol{v}_{j_0} > \leq 0$ which constraints λ_{j_0} to be non-negative. This concludes the proof of Lemma 3. □

Let us consider the case where $\mathcal{C} = \mathcal{C}_{\nearrow}$. We apply Lemma 3. In this case, W is the linear space generated by $\mathbf{1}$, we get that for all $\ell \in \{2, ..., \ell_n\}$ and $1 \leq i < j \leq \ell$, $\boldsymbol{e}_{ij}^\ell$ satisfies $\Pi_W \boldsymbol{e}_{ij}^\ell = 0$. Moreover $\|\boldsymbol{e}_{ij}^\ell\| = 1$ and

$$\forall \boldsymbol{f} \in \mathcal{C}_{\nearrow},\ < \boldsymbol{f}, \boldsymbol{e}_{ij}^\ell >= N_{ij}^\ell \left(\bar{f}_{J_i^\ell} - \bar{f}_{J_j^\ell} \right) \leq 0.$$

Let us consider the case where $\mathcal{C} = \mathcal{C}_{\smile}$. In this case, $p = n-2$ and for all $j = 1, \ldots, n-2$,

$$\boldsymbol{v}_j = (x_{j+1} - x_{j+2})\boldsymbol{e}_j + (x_{j+2} - x_j)\boldsymbol{e}_{j+1} + (x_j - x_{j+1})\boldsymbol{e}_{j+2}.$$

Since $\|\boldsymbol{e}_{ijk}^{\ell}\| = 1$, by Lemma 3 it is enough to prove that

(i) for all $\boldsymbol{f} \in W$, $< \boldsymbol{f}, \boldsymbol{e}_{ijk}^{\ell} >= 0$,

(ii) for all $\boldsymbol{f} \in \mathcal{C}_{\smile}$, $< \boldsymbol{f}, \boldsymbol{e}_{ijk}^{\ell} >\leq 0$.

First note that for all $\boldsymbol{f} \in \mathbb{R}^n$,

$$< \boldsymbol{f}, \boldsymbol{e}_{ijk}^{\ell} >= N_{ijk}^{\ell} \left(\bar{f}_{J_j^\ell} - \lambda_{ijk}^{\ell} \bar{f}_{J_i^\ell} - (1 - \lambda_{ijk}^{\ell}) \bar{f}_{J_k^\ell} \right). \tag{45}$$

Clearly if $\boldsymbol{f} = \mathbf{1}$ or $\boldsymbol{f} = \boldsymbol{x}$, $< \boldsymbol{f}, \boldsymbol{e}_{ijk}^{\ell} >= 0$ and since by definition of $\mathcal{C}_{\smile}$, W is the linear space generated by $\mathbf{1}$ and $\boldsymbol{x}$, (i) holds true. Let now $\boldsymbol{f} \in \mathcal{C}_{\smile}$. There exists some convex function F mapping $[x_1, x_n]$ into $\mathbb{R}$ such that $F(x_i) = f_i$ for all $i = 1, ..., n$ (take the piecewise linear function verifying this property for instance). Let $i < j < k$ and $l \in J_j^\ell$. We set

$$\mu_{ik}^l = \frac{\bar{\boldsymbol{x}}_{J_k^\ell} - x_l}{\bar{\boldsymbol{x}}_{J_k^\ell} - \bar{\boldsymbol{x}}_{J_i^\ell}}.$$

Note that $0 \leq \mu_{ik}^l \leq 1$ and that

$$x_l = \mu_{ik}^l \bar{\boldsymbol{x}}_{J_i^\ell} + (1 - \mu_{ik}^l) \bar{\boldsymbol{x}}_{J_k^\ell}.$$

Since F is convex on $[x_1, x_n]$ we have for all $l \in J_j^\ell$,

$$\begin{aligned} F(x_l) &\leq \mu_{ik}^l F(\bar{\boldsymbol{x}}_{J_i^\ell}) + (1 - \mu_{ik}^l) F(\bar{\boldsymbol{x}}_{J_k^\ell}) \\ &\leq \mu_{ik}^l \bar{\boldsymbol{f}}_{J_i^\ell} + (1 - \mu_{ik}^l) \bar{\boldsymbol{f}}_{J_k^\ell}. \end{aligned}$$

Note that $\sum_{l \in J_j^\ell} \mu_{ik}^l / |J_j^\ell| = \lambda_{ijk}^\ell$. We derive from the above inequality that

$$\bar{f}_{J_j^\ell} = \frac{1}{|J_j^\ell|} \sum_{l \in J_j^\ell} F(x_l) \leq \lambda_{ijk}^\ell \bar{f}_{J_i^\ell} + (1 - \lambda_{ijk}^\ell) \bar{f}_{J_k^\ell},$$

which, thanks to (45), leads to (ii).

Let us consider the case where $\mathcal{C} = \mathcal{C}_{r,R}$. By Lemma 2 we know that $\tilde{\mathcal{C}}_{r,R} = \mathcal{C}_{r,R}$ and therefore for each $\boldsymbol{i} \in \mathcal{I}_r$, the linear form $\boldsymbol{f} \mapsto \phi_{\boldsymbol{i}}(R \star \boldsymbol{f})$ is a linear combination of the linear forms $\boldsymbol{f} \mapsto \phi_i(R \star \boldsymbol{f})$ with $i = 1, \ldots, n-r$. Consequently, if $\boldsymbol{w} \in W$ then for all $\boldsymbol{i} \in \mathcal{I}_r$, $\phi_{\boldsymbol{i}}(R \star \boldsymbol{w}) = 0$. For each $J \subset \{1, \ldots, n\}$, $\boldsymbol{t}_J^*$ defined by (22) satisfies $\|\boldsymbol{t}_J^*\| = 1$. By applying (44) with $\boldsymbol{f} = \boldsymbol{w}$, we get $< \boldsymbol{w}, \boldsymbol{t}_J^* >\leq 0$ for all $\boldsymbol{w} \in \mathcal{C}_{r,R}$ and $< \boldsymbol{w}, \boldsymbol{t}_J^* >= 0$ for all $\boldsymbol{w} \in W$. Consequently, by Lemma 3, $\boldsymbol{t}_J^*$ belongs to $\mathcal{T}$.

8. Proof of Proposition 1

8.1. *Proof for* $(T_\alpha, \mathcal{C}) = (T_{\alpha,1}, \mathcal{C}_{\geq 0})$. We prove the proposition by applying Theorem 1. We decompose the proof into six steps.

Step 1 For all integer $N \geq 1$, let $\bar{T}_N^{-1}(u)$ denote the $1-u$ quantile of a Student random variable with N degrees of freedom. We have for all $u \in]0,1[$,

$$\bar{T}_N^{-1}(u) \leq 1 + C\left\{\log^{1/4}\left(\frac{1}{u}\right) + \log^{1/2}\left(\frac{1}{u}\right)\exp\left(\frac{2}{N}\log\left(\frac{1}{u}\right)\right)\right\} \tag{46}$$

for some absolute constant $C > 0$.

Proof of step 1 Let $\bar{F}_{1,N}^{-1}(u)$ denote the $1-u$ quantile of a Fisher variable with 1 and N degrees of freedom, then

$$\bar{T}_N^{-1}(u) = \sqrt{\bar{F}_{1,N}^{-1}(u)}.$$

It follows from Lemma 1 in Baraud, Huet and Laurent (2003) that for all $u \in]0,1[$, $N \geq 1$,

$$\bar{F}_{1,N}^{-1}(u) \leq 1 + 2\sqrt{2}\log^{1/2}(\frac{1}{u}) + \frac{3N}{2}\left\{\exp\left(\frac{4}{N}\log\left(\frac{1}{u}\right)\right) - 1\right\}.$$

Using the inequality $\exp(x) - 1 \leq x\exp(x)$ which holds for all $x > 0$, we obtain

$$\bar{F}_{1,N}^{-1}(u) \leq 1 + 2\sqrt{2}\log^{1/2}\left(\frac{1}{u}\right) + 6\log\left(\frac{1}{u}\right)\exp\left(\frac{4}{N}\log\left(\frac{1}{u}\right)\right)$$

and since $\sqrt{a+b} \leq \sqrt{a} + \sqrt{b}$ for all $a > 0$ and $b > 0$

$$\sqrt{\bar{F}_{1,N}^{-1}(u)} \leq 1 + C\left\{\log^{1/4}\left(\frac{1}{u}\right) + \log^{1/2}\left(\frac{1}{u}\right)\exp\left(\frac{2}{N}\log\left(\frac{1}{u}\right)\right)\right\}$$

for some absolute constant $C > 0$.

Step 2 For all $\ell \in \{1, ..., \ell_n\}$, $t \in \mathcal{T}_{n,1}^{\ell}$, we have

$$q_{\boldsymbol{t}}(\alpha) = q_1(\ell, u_\alpha) \leq C(\alpha)\sqrt{\log(n)}. \tag{47}$$

Proof of step 2 On the one hand, by definition of $q_1(\ell, .)$

$$\alpha = \mathbb{P}_{0,1}\left(T_{\alpha,1} > 0\right) \leq \sum_{\ell=1}^{\ell_n} \mathbb{P}\left(T_1^{\ell}(\boldsymbol{\varepsilon}) - q_1(\ell, u_\alpha) > 0\right) \leq \ell_n u_\alpha,$$

and thus

$$u_\alpha \geq \alpha/\ell_n. \tag{48}$$

On the other hand, for all $\ell \in \{1, \ldots, \ell_n\}$ and $J \in \mathcal{J}^{\ell}$, the random variables

$$U_J = \frac{-\sum_{i\in J}\varepsilon_i}{\|\boldsymbol{\varepsilon} - \Pi_{V_n}\boldsymbol{\varepsilon}\|}\sqrt{\frac{n-d_n}{|J|}}$$

being distributed as Student variables with $n - d_n$ degrees of freedom, we have that

$$\mathbb{P}\left(T_1^{\ell}(\boldsymbol{\varepsilon}) > \bar{T}_{n-d_n}^{-1}\left(\frac{u_\alpha}{|\mathcal{J}^{\ell}|}\right)\right) \leq \sum_{J\in\mathcal{J}^{\ell}} \mathbb{P}\left(U_J > \bar{T}_{n-d_n}^{-1}\left(\frac{u_\alpha}{|\mathcal{J}^{\ell}|}\right)\right) \leq u_\alpha \tag{49}$$

and thus $q_1(\ell, u_\alpha) \leq \bar{T}^{-1}_{n-d_n}\left(u_\alpha/|\mathcal{J}^\ell|\right)$. This inequality together with (48) and (46) leads to (47), as $|\mathcal{J}^\ell| \leq \ell_n \leq n/2$ and $n - d_n = n - \ell_n \geq n/2$.

Step 3 For all $\boldsymbol{f} = (F(x_1), ..., F(x_n))'$ with $F \in \mathcal{H}_s(L)$,

$$\frac{\|\boldsymbol{f} - \Pi_{V_{n,\mathrm{cste}}}\boldsymbol{f}\|^2}{n} \leq C(s)L^2 n^{-2s}. \tag{50}$$

Proof of step 3 Note that the vector

$$\tilde{\boldsymbol{f}} = \sum_{k=1}^{\ell_n} F(\bar{x}_{J_k})\mathbf{1}_{J_k}$$

belongs to $V_{n,\mathrm{cste}}$ and therefore

$$\begin{aligned}
\|\boldsymbol{f} - \Pi_{V_{n,\mathrm{cste}}}\boldsymbol{f}\|^2 &\leq \|\boldsymbol{f} - \tilde{\boldsymbol{f}}\|^2 \\
&= \sum_{k=1}^{\ell_n}\sum_{i\in J_k}\left(F(x_i) - F(\bar{x}_{J_k})\right)^2 \\
&\leq \sum_{k=1}^{\ell_n}\sum_{i\in J_k} L^2\ell_n^{-2s} \\
&= nL^2\ell_n^{-2s}.
\end{aligned}$$

Noting that $\ell_n = d_n \geq n/4$, we get (50).

Step 4 Assuming that $n \geq (L/\sigma)^{1/s}$, there exists some constant C depending on s and β only such that

$$\frac{\bar{\chi}^{-1}_{n-d_n,\|\boldsymbol{f}-\Pi_{V_{n,\mathrm{cste}}}\boldsymbol{f}\|^2/\sigma^2}(\beta/2)}{n-d_n} \leq C. \tag{51}$$

Using the inequality due to Birgé (2001) on the quantiles of non-central χ^2 we have that

$$\bar{\chi}^{-1}_{n-d_n,a^2}(\beta/2) \leq n - d_n + a^2 + 2\sqrt{(n-d_n+2a^2)\log(2/\beta)} + 2\log(2/\beta).$$

Setting $a = \|\boldsymbol{f} - \Pi_{V_{n,\mathrm{cste}}}\boldsymbol{f}\|/\sigma$ and using (50) we derive that

$$\bar{\chi}^{-1}_{n-d_n,a^2}(\beta/2)/(n-d_n) \leq C(\beta, s). \tag{52}$$

Step 5 Under the assumption of step 4, for all $\boldsymbol{t} \in \mathcal{T}_n$,

$$v_{\boldsymbol{t}}(\boldsymbol{f}, \beta) \leq \kappa^*\sqrt{\log(n)}\sigma,$$

for some constant κ^* depending on α, β, s only.

Proof of step 5 We recall that

$$v_{\boldsymbol{t}}(\boldsymbol{f},\beta) = \left(q_{\boldsymbol{t}}(\alpha) \frac{1}{\sqrt{n-d_n}} \sqrt{\bar{\chi}^{-1}_{n-d_n,\|\boldsymbol{f}-\Pi_{V_n}\boldsymbol{f}\|^2/\sigma^2}(\beta/2)} + \bar{\Phi}^{-1}(\beta/2) \right) \sigma.$$

We conclude by using the elementary inequality

$$\bar{\Phi}^{-1}(\beta/2) \leq \sqrt{2\log(2/\beta)},$$

by gathering (47) and (51).
We conclude the proof with this final step.

Step 6 There exists a constant κ depending on α, β and s only, such that if n large enough, and F satisfies

$$\min_{x\in[0,1]} F(x) \leq -\kappa\rho_n, \tag{53}$$

then, there exists $\boldsymbol{t}^* \in \mathcal{T}_n$ such that

$$< \boldsymbol{f}, \boldsymbol{t}^* > \geq v_{\boldsymbol{t}^*}(\boldsymbol{f},\beta). \tag{54}$$

Proof of step 6 Since $F \in \mathcal{H}_s(L)$, under Assumption (53), there exists $j \in \{1, 2, \ldots, n\}$ such that

$$F(j/n) \leq -\kappa\rho_n + Ln^{-s}.$$

For n large enough, $Ln^{-s} \leq \kappa\rho_n/2$, hence $F(j/n) \leq -\kappa\rho_n/2$.
Let us take κ satisfying

$$\frac{\kappa}{4} = (2\kappa^*)^{2s/(1+2s)}$$

where κ^* is defined at step 5.
Let us define

$$\ell(n) = \left[\left(\frac{4L}{\kappa\rho_n} \right)^{1/s} \right], \tag{55}$$

and J as the element of $\mathcal{J}^{\ell(n)}$ containing j. Note that for n large enough, $\ell(n) \in \{1, ..., \ell_n\}$.
Now, for all $k \in J$, since $F \in \mathcal{H}_s(L)$

$$\begin{aligned} f_k = F(x_k) &= -F(x_j) + F(x_j) + F(x_k) \\ &\leq -\kappa\rho_n/2 + L|x_k - x_j|^s \\ &\leq -\kappa\rho_n/2 + L\ell(n)^{-s} \\ &\leq -\kappa\rho_n/4 \end{aligned}$$

and thus, by taking $\boldsymbol{t}^* \in \mathcal{T}_{n,1}$ as

$$\boldsymbol{t}^* = -\frac{1}{\sqrt{|J|}} \sum_{i\in J} \boldsymbol{e}_i,$$

we derive that

$$\begin{aligned} < \boldsymbol{f}, \boldsymbol{t}^* > &= -\sqrt{|J|}\bar{f}_J \\ &\geq \sqrt{|J|}\kappa\rho_n/4. \end{aligned}$$

By construction of the partition of the data, we have for all positive integers $p \leq q \leq r$ that

$$\left[\frac{r}{q}\right] \leq |I^r_{p,q}| \leq \left[\frac{r}{q}\right] + 1. \tag{56}$$

For all $j \in \{1, \ldots, \ell(n)\}$, $J = J_j^{\ell(n)}$ (see (8)) is a union of $|I^{\ell_n}_{j,\ell(n)}| \geq [\ell_n/\ell(n)]$ disjoint sets of cardinality at least $[n/\ell_n]$. Hence

$$|J_j^{\ell(n)}| \geq \left[\frac{n}{\ell_n}\right]\left[\frac{\ell_n}{\ell(n)}\right] \geq \frac{n}{4\ell(n)}$$

since $[x] \geq x/2$ for all $x \geq 1$. Therefore we get

$$|J| \geq n/(4\ell(n)) \geq \frac{n}{4}\left(\kappa\frac{\rho_n}{4L}\right)^{1/s} \tag{57}$$

using (55).
This implies that

$$< \boldsymbol{f}, \boldsymbol{t}^* > \geq \frac{\sqrt{n}}{8}\left(\frac{\kappa\rho_n}{4L}\right)^{1/(2s)}\kappa\rho_n \geq \kappa^*\sigma\sqrt{\log(n)}$$

by definition of κ.

8.2. *Proof for* $(T_\alpha, \mathcal{C}) = (T_{\alpha,2}, \mathcal{C}_{\nearrow})$. We follow the proof of Theorem 1 for $(T_\alpha, \mathcal{C}) = (T_{\alpha,1}, \mathcal{C}_{\geq 0})$: the results of steps 1 to 5 still hold. The proof of step 2 differs in the following way: Equation (49) becomes

$$\begin{aligned} &\mathbb{P}\left(T_2^\ell(\boldsymbol{\varepsilon}) > \bar{T}^{-1}_{n-d_n}\left(\frac{u_\alpha}{|\mathcal{T}^\ell_{n,2}|}\right)\right) \\ &\leq \sum_{1\leq i<j\leq \ell} \mathbb{P}\left(\frac{< \boldsymbol{\varepsilon}, \boldsymbol{e}^\ell_{ij} >}{\|\boldsymbol{\varepsilon} - \Pi_{V_{n,\mathrm{cste}}}\boldsymbol{\varepsilon}\|/(n-\ell_n)} > \bar{T}^{-1}_{n-d_n}\left(\frac{u_\alpha}{|\mathcal{T}^\ell_{n,2}|}\right)\right) \leq u_\alpha. \end{aligned}$$

We conclude the proof of step 2 by noticing that for all $\ell \in \{1, ..., \ell_n\}$, $|\mathcal{T}^\ell_{n,2}|$ is bounded from above by $n^2/4$.

Step 6 For n large enough, under the assumption that

$$\inf_{G\in\mathcal{K}_{\nearrow}} \|F - G\|_\infty \geq \kappa\rho_n, \tag{58}$$

there exists $\boldsymbol{t}^* \in \mathcal{T}_{n,2}$, such that $< \boldsymbol{t}^*, f > \geq v_{\boldsymbol{t}^*}(\boldsymbol{f}, \beta)$.

Proof of step 6 Let us first remark that

$$\inf_{G\in\mathcal{K}_{\nearrow}} \|F-G\|_\infty \leq \sup_{0\leq x\leq y\leq 1} (F(x)-F(y)).$$

Indeed, let $G^* \in \mathcal{K}_{\nearrow}$ be defined as

$$G^*(y) = \sup_{0\leq x\leq y} F(x).$$

Then,

$$\inf_{G\in\mathcal{K}_{\nearrow}} \|F-G\|_\infty \leq \|F-G^*\|_\infty = \sup_{0\leq x\leq y\leq 1} (F(x)-F(y)).$$

Hence, under Assumption (58), there exists $x<y$ such that $F(x)-F(y)\geq \kappa\rho_n$. Since $F\in\mathcal{H}_s(L)$, if $|x_i-x|\leq 1/n$ and $|x_j-y|\leq 1/n$, then

$$F(x_i)-F(x_j)\geq \kappa\rho_n - 2Ln^{-s} \geq \kappa\rho_n/2$$

for n large enough. Hence, there exists $1\leq i<j\leq n$ such that $F(x_i)-F(x_j)\geq \kappa\rho_n/2$.
Let us set

$$\ell(n) = \left[\left(\frac{8L}{\kappa\rho_n}\right)^{1/s}\right],$$

which belongs to $\{1,\dots,\ell_n\}$ at least for n large enough. Let I and J be the elements of $\mathcal{J}^{\ell(n)}$ satisfying $i\in I$ and $j\in J$.
Arguing as in step 6 of Section 8.1, since $F\in\mathcal{H}_s(L)$,

$$\bar{f}_I \geq F(x_i) - L\ell(n)^{-s} \quad\text{and}\quad \bar{f}_J \leq F(x_j) + L\ell(n)^{-s}$$

and we deduce that

$$\bar{f}_I - \bar{f}_J \geq \kappa\rho_n/2 - 2L\ell(n)^{-s} \geq \kappa\rho_n/4.$$

This implies that there exists $1\leq i^*<j^*\leq \ell(n)$ with $I=J_{i^*}^{\ell(n)}$ and $J=J_{j^*}^{\ell(n)}$, such that

$$< \boldsymbol{e}_{i^*j^*}^{\ell(n)}, f >= N_{i^*j^*}^{\ell(n)}\left(\bar{f}_I-\bar{f}_J\right) \geq N_{i^*j^*}^{\ell(n)}\frac{\kappa\rho_n}{4}.$$

Using Inequality (56), and since $\ell_n=[n/2]$, we have that for all $K\in\mathcal{J}^{\ell(n)}$,

$$2\left[\frac{\ell_n}{\ell(n)}\right] \leq |K| \leq 3\left(\left[\frac{\ell_n}{\ell(n)}\right]+1\right),$$

which implies that

$$N_{i^*j^*}^{\ell(n)} = \sqrt{\frac{|I||J|}{|I|+|J|}} \geq C\sqrt{\frac{\ell_n}{\ell(n)}}.$$

We now conclude as in the proof of step 6 by taking $\boldsymbol{t}^* = \boldsymbol{e}_{i^*j^*}^{\ell(n)}$.

8.3. *Proof of Theorem 1 for* $(T_\alpha, \mathcal{C}) = (T_{\alpha,3}, \mathcal{C}_\smile)$. We follow the proof of Theorem 1 for the case $(T_\alpha, \mathcal{C}) = (T_{\alpha,1}, \mathcal{C}_{\geq 0})$: the results of steps 1 to 5 still hold. Nevertheless, the proof of step 2 differs in the following way: Equation (49) becomes

$$\mathbb{P}\left(T_3^\ell(\boldsymbol{\varepsilon}) > \bar{T}_{n-d_n}^{-1}\left(\frac{u_\alpha}{|\mathcal{T}_{n,3}^\ell|}\right)\right)$$
$$\leq \sum_{1\leq i<j<k\leq \ell} \mathbb{P}\left(\frac{<\boldsymbol{\varepsilon}, \boldsymbol{e}_{ijk}^\ell>}{\|\boldsymbol{\varepsilon} - \Pi_{V_{n,\mathrm{cste}}}\boldsymbol{\varepsilon}\|/(n-\ell_n)} > \bar{T}_{n-d_n}^{-1}\left(\frac{u_\alpha}{|\mathcal{T}_{n,3}^\ell|}\right)\right) \leq u_\alpha.$$

We conclude the proof of step 2 by noticing that for all $\ell \in \{1, ..., \ell_n\}$, $|\mathcal{T}_{n,3}^\ell|$ is bounded from above by $n^3/8$.

Step 6 For n large enough, under the assumption that

$$\inf_{G\in\mathcal{K}_\smile} \|F - G\|_\infty \geq \kappa\rho_n, \tag{59}$$

there exists $\boldsymbol{t}^* \in \mathcal{T}_{n,3}$ such that

$$<\boldsymbol{t}^*, f> \geq v_{\boldsymbol{t}^*}(\boldsymbol{f}, \beta).$$

Proof of step 6 We decompose the proof into three parts.

Part 1: For n large enough, and all $F \in \mathcal{H}_s(L)$ satisfying (59), we have

$$\inf_{\boldsymbol{g}\in\mathcal{C}_\smile} \|\boldsymbol{f} - \boldsymbol{g}\|_\infty \geq \kappa\rho_n/4,$$

with $\boldsymbol{f} = (F(x_1), \ldots, F(x_n))'$.

Proof of Part 1 We first prove the following inequality :

$$\inf_{G\in\mathcal{K}_\smile} \|F - G\|_\infty \leq 2Ln^{-s} + 3 \inf_{\boldsymbol{g}\in\mathcal{C}_\smile} \|\boldsymbol{f} - \boldsymbol{g}\|_\infty. \tag{60}$$

Part 1 derives obviously from this inequality.
For all $\boldsymbol{g} \in \mathcal{C}_\smile$, we consider the function $G_{\boldsymbol{g}} \in \mathcal{K}_\smile$ defined as the piecewise linear function such that for all i $G_{\boldsymbol{g}}(x_i) = g_i$ and such that $G_{\boldsymbol{g}}$ is affine on the interval $[0, x_2]$. Then $\inf_{G\in\mathcal{K}_\smile} \|F - G\|_\infty \leq \|F - G_{\boldsymbol{g}}\|_\infty$. Moreover, by setting $x_0 = 0$ and $g_0 = G_{\boldsymbol{g}}(0)$,

$$\|F - G_{\boldsymbol{g}}\|_\infty$$
$$= \sup_{i\in\{1,\ldots,n\}} \sup_{x\in[x_{i-1},x_i]} |F(x) - G_{\boldsymbol{g}}(x)|$$
$$\leq \sup_{i\in\{1,\ldots,n\}} \sup_{x\in[x_{i-1},x_i]} |F(x) - F(x_i) + F(x_i) - G_{\boldsymbol{g}}(x_i) + G_{\boldsymbol{g}}(x_i) - G_{\boldsymbol{g}}(x)|$$
$$\leq Ln^{-s} + \|f - g\|_\infty + \sup_{i\in\{1,\ldots,n\}} |g_{i-1} - g_i|$$

since $\sup_{x\in[x_{i-1},x_i]} |G_{\boldsymbol{g}}(x_i)-G_{\boldsymbol{g}}(x)| = |G_{\boldsymbol{g}}(x_i)-G_{\boldsymbol{g}}(x_{i-1})|$ (G is linear on $[x_{i-1},x_i]$). In addition, noticing that $|g_1-g_0| = |g_2-g_1|$,

$$\begin{aligned}\sup_{i\in\{1,\dots,n\}} |g_i - g_{i-1}| &\le \sup_{i\in\{2,\dots,n\}} |g_i - f_i + f_i - f_{i-1} + f_{i-1} - g_{i-1}| \\ &\le 2\|\boldsymbol{f}-\boldsymbol{g}\|_\infty + Ln^{-s}\end{aligned}$$

This concludes the proof of (60).

Part 2: for all $\boldsymbol{f}\in\mathbb{R}^n$,

$$\inf_{\boldsymbol{g}\in\mathcal{C}_\smile} \|\boldsymbol{f}-\boldsymbol{g}\|_\infty \le \max_{1\le i<j<k\le n} \left(f_j - \frac{x_k-x_j}{x_k-x_i}f_i - \frac{x_j-x_i}{x_k-x_i}f_k\right)_+, \tag{61}$$

where for $x\in\mathbb{R}$, $(x)_+ = x\mathbf{1}_{x>0}$ denotes the positive part of x.

Proof of Part 2 Let us define $\boldsymbol{g}^*\in\mathcal{C}_\smile$ as follows: $g_1^* = f_1$ and for $i=1,\dots,n-1$,

$$g_{i+1}^* = g_i^* + \inf\left\{\frac{f_k-g_i^*}{x_k-x_i},\ k>i\right\}(x_{i+1}-x_i).$$

In words, if F_{lin} denotes the piecewise linear function on $[x_1,x_n]$ taking the value f_i at x_i, then $\boldsymbol{g}^*$ is the vector $(G_{lin}^*(x_1),\dots,G_{lin}^*(x_n))'$ where G_{lin}^* is the largest convex function satisfying for all $u\in[x_1,x_n]$ $G_{lin}^*(u)\le F_{lin}(u)$. Note that the function G_{lin}^* is also piecewise linear and satisfies that for all $j\in\{1,\dots,n\}$ such that $F_{lin}(x_j)-G_{lin}^*(x_j)>0$, there exist $1\le i<j<k\le n$ such that

$$F_{lin}(x_j) - G_{lin}^*(x_j) = f_j - \frac{x_k-x_j}{x_k-x_i}f_i - \frac{x_j-x_i}{x_k-x_i}f_k.$$

Consequently,

$$\begin{aligned}\|\boldsymbol{f}-\boldsymbol{g}^*\|_\infty &= \max_{j=1,\dots,n}(F_{lin}(x_j)-G_{lin}^*(x_j)) \\ &\le \max_{1\le i<j<k\le n}\left(f_j - \frac{x_k-x_j}{x_k-x_i}f_i - \frac{x_j-x_i}{x_k-x_i}f_k\right)_+\end{aligned}$$

Part 3 Let $\kappa' = \kappa/4$. We set,

$$\ell(n) = 1 + \left[\left(\frac{6L}{\kappa'\rho_n}\right)^{1/s}\right].$$

If there exist $1\le i<j<k\le n$ such that

$$f_j - \frac{x_k-x_j}{x_k-x_i}f_i - \frac{x_j-x_i}{x_k-x_i}f_k \ge \kappa'\rho_n$$

then there exist $I = J_{i^*}^{\ell(n)}$, $J = J_{j^*}^{\ell(n)}$ and $K = J_{k^*}^{\ell(n)}$ with $i^*<j^*<k^*$, such that

$$\bar f_J - \frac{\bar x_K-\bar x_J}{\bar x_K-\bar x_I}\bar f_I - \frac{\bar x_J-\bar x_I}{\bar x_K-\bar x_I}\bar f_K \ge \kappa'\rho_n/4. \tag{62}$$

Proof of part 3 Note that

$$\ell(n) \geq \left(\frac{6L}{\kappa' \rho_n}\right)^{1/s} \tag{63}$$

and that for n large enough $\ell(n) \in \{1, ..., \ell_n\}$.
In the sequel, we shall use the following inequalities

$$\forall E \in \{I, J, K\},\ \max_{l,l' \in E} |x_l - x_{l'}| \leq \frac{1}{\ell(n)} \quad \text{and} \quad \max_{l \in E} |x_l - \bar{x}_E| \leq \frac{1}{2\ell(n)} \tag{64}$$

and the following notations

$$\lambda = \frac{x_k - x_j}{x_k - x_i},\ \bar{\lambda} = \frac{\bar{x}_K - \bar{x}_J}{\bar{x}_K - \bar{x}_I} \quad \text{and} \quad \Delta = \bar{f}_J - \bar{\lambda}\bar{f}_I - (1-\bar{\lambda})\bar{f}_K.$$

We bound Δ from below as follows:

$$\begin{aligned}
\Delta &= f_j - \lambda f_i - (1-\lambda) f_k \\
&\quad + \bar{f}_J - f_j + \lambda f_i - \bar{\lambda}\bar{f}_I + (1-\lambda) f_k - (1-\bar{\lambda})\bar{f}_K \\
&\geq \kappa\rho_n + \bar{f}_J - f_j + (\lambda - \bar{\lambda}) f_i - \bar{\lambda}(\bar{f}_I - f_i) + (\bar{\lambda} - \lambda) f_k - (1-\bar{\lambda})(\bar{f}_K - f_k) \\
&\geq \kappa\rho_n - 2\max\left\{\left|\bar{f}_I - f_i\right|, \left|\bar{f}_J - f_j\right|, \left|\bar{f}_K - f_k\right|\right\} - \left|\lambda - \bar{\lambda}\right| |f_i - f_k|.
\end{aligned}$$

Let us now bound from above the quantities

$$|f_i - f_k|,\ \max\left\{\left|\bar{f}_I - f_i\right|, \left|\bar{f}_J - f_j\right|, \left|\bar{f}_K - f_k\right|\right\},\ \left|\lambda - \bar{\lambda}\right|.$$

Since $F \in \mathcal{H}_s(L)$, we have that

$$|f_i - f_k| = |F(x_i) - F(x_k)| \leq L|x_k - x_i|^s \tag{65}$$

and by using (64) that

$$\max\left\{\left|\bar{f}_I - f_i\right|, \left|\bar{f}_J - f_j\right|, \left|\bar{f}_K - f_k\right|\right\} \leq L\ell(n)^{-s}. \tag{66}$$

For each $(l, E) \in \{(i, I), (j, J), (k, K)\}$, let

$$h_l = \bar{x}_E - x_l,$$

we have

$$\begin{aligned}
\bar{\lambda} &= \frac{x_k - x_j + h_k - h_j}{x_k - x_i + h_k - h_i} = \lambda\left(\frac{1 + (h_k - h_j)/(x_k - x_j)}{1 + (h_k - h_i)/(x_k - x_i)}\right) \\
&= \lambda\left(1 + \frac{(h_k - h_j)/(x_k - x_j) - (h_k - h_i)/(x_k - x_i)}{1 + (h_k - h_i)/(x_k - x_i)}\right)
\end{aligned}$$

and as from (64) $\max\{|h_k - h_j|, |h_k - h_i|\} \leq \ell(n)^{-1}$, we deduce that

$$\begin{aligned}
\left|\bar{\lambda} - \lambda\right| &= |\lambda| \left|\frac{(h_k - h_j)/(x_k - x_i) - (h_k - h_i)/(x_k - x_i)}{1 + (h_k - h_i)/(x_k - x_i)}\right| \\
&\leq \frac{2\delta}{|1 - \delta|},
\end{aligned} \tag{67}$$

where

$$\delta = \frac{1}{\ell(n)\,|x_k - x_i|}. \tag{68}$$

In order to bound δ from above, note that since $F \in \mathcal{H}_s(L)$,

$$\begin{aligned}\kappa'\rho_n &\le f_j - \lambda f_i - (1-\lambda) f_k \\ &= \lambda\left(F(x_j) - F(x_i)\right) + (1-\lambda)\left(F(x_j) - F(x_k)\right) \\ &\le L \max\left\{|x_j - x_i|^s, |x_k - x_j|^s\right\}\end{aligned}$$

and therefore

$$\begin{aligned}|x_k - x_i| &\ge \max\left\{|x_j - x_i|, |x_k - x_j|\right\} \\ &= \left\{\max\left\{|x_j - x_i|^s, |x_k - x_j|^s\right\}\right\}^{1/s} \\ &\ge \left(\frac{\kappa'\rho_n}{L}\right)^{1/s}.\end{aligned}$$

Thus, we deduce by (63) and the fact that $s \in]0,1]$ that

$$\delta \le \frac{L^{1/s}}{(\kappa'\rho_n)^{1/s}\ell(n)} \le \frac{1}{6}. \tag{69}$$

By gathering (65), (66), (67), we get

$$\Delta \ge \kappa'\rho_n - 2L\ell(n)^{-s} - 2L\frac{\delta}{1-\delta}|x_k - x_i|^s.$$

By using (68), (69) and (63) we finally get

$$\begin{aligned}\Delta &= \kappa'\rho_n - 2L\ell(n)^{-s} - 2L\ell(n)^{-s}\frac{\delta^{1-s}}{1-\delta} \\ &\ge \kappa'\rho_n\left\{1 - \frac{1}{3}\left(1 + \frac{1}{1-1/6}\right)\right\} \\ &\ge \kappa'\rho_n/4.\end{aligned}$$

Let us now conclude the proof of step 6. Under the assumption that

$$\inf_{\boldsymbol{g}\in\mathcal{C}_{\smile}} \|\boldsymbol{f} - \boldsymbol{g}\|_\infty \ge \kappa'\rho_n$$

we know from (61) that there exists $i < j < k$ such that

$$f_j - \frac{x_k - x_j}{x_k - x_i} f_i - \frac{x_j - x_i}{x_k - x_i} f_k \ge \kappa'\rho_n,$$

and from (62) that there exists $I = J_{i^*}^{\ell(n)}$, $J = J_{j^*}^{\ell(n)}$ and $K = J_{k^*}^{\ell(n)}$ with $i^* < j^* < k^*$ such that

$$< \boldsymbol{f}, \boldsymbol{e}_{i^*j^*k^*}^{\ell(n)} >= N_{i^*j^*k^*}^{\ell(n)}\left(\bar{f}_J - \lambda_{i^*j^*k^*}^{\ell(n)}\bar{f}_I - (1-\lambda_{i^*j^*k^*}^{\ell(n)})\bar{f}_K\right) \ge \frac{N_{i^*j^*k^*}^{\ell(n)}\kappa'\rho_n}{4}.$$

Noting that for all $E \in \{I, J, K\}$

$$|E| \geq 2\left[\frac{\ell_n}{\ell(n)}\right] \geq \frac{\ell_n}{\ell(n)} \geq \frac{n}{4\ell(n)},$$

and that $\|\boldsymbol{e}^{\ell(n)}_{i^*j^*k^*}\|^2 \leq 1/|I| + 1/|J| + 1/|K|$, we have that

$$N^{\ell(n)}_{i^*j^*k^*} \geq \sqrt{\frac{1}{|I|^{-1} + |J|^{-1} + |K|^{-1}}} \geq \sqrt{\frac{n}{12\ell(n)}}.$$

As $\ell(n) \leq 2(12L/(\kappa'\rho_n))^{1/s}$ at least for n large enough, we deduce that

$$N^{\ell(n)}_{i^*j^*k^*} \geq \sqrt{\frac{n(\kappa'\rho_n)^{1/s}}{8(12L)^{1/s}}}.$$

Consequently, we get

$$\begin{aligned} < \boldsymbol{f}, \boldsymbol{e}^{\ell(n)}_{i^*j^*k^*} > &\geq \sqrt{(\kappa'\rho_n)^{(1+2s)/s}\frac{n}{12^{1/s}128L^{1/s}}} \\ &\geq \kappa^*\sqrt{\log(n)}\sigma, \end{aligned}$$

for κ' suitably chosen. It remains to take $\boldsymbol{t}^* = \boldsymbol{e}^{\ell(n)}_{i^*j^*k^*} \in \mathcal{T}_{n,3}$ to complete the proof.

9. Proof of Proposition 3 . The proof of Proposition 3 is divided into two parts. In Section 9.1 we show that if (24) or (25) hold, then $\mathbb{P}_{F,\sigma}(T_\alpha > 0) \geq 1 - \beta$. The second part of the proposition is shown in Section 9.2.

9.1. *Proof of the first part of Proposition 3.* We only prove the result under Condition (24), the proof under Condition (25) being almost the same. By combining (43) and (44) we obtain that if F is such that RF is r-th times differentiable, then for all $J \subset \{1, ..., n\}$ there exists a sequence $\left\{c_{\boldsymbol{i}},\ \boldsymbol{i} \in \mathcal{I}_r \cap J^{r+1}\right\}$ verifying both $c_{\boldsymbol{i}} \in]\min_{j\in J} x_j, \max_{j\in J} x_j[$ and

$$- < \boldsymbol{f}, \boldsymbol{t}^*_J >= N_J^{-1} \sum_{\boldsymbol{i}\in\mathcal{I}_r\cap J^{r+1}} \frac{\Lambda(F)(c_{\boldsymbol{i}})}{r!}\phi^2_{\boldsymbol{i}}(\boldsymbol{x}^r), \tag{70}$$

where $N_J = \mathrm{Gram}(\mathbf{1}_J, \boldsymbol{x}_J, ..., \boldsymbol{x}_J^{r-1})\gamma_J$. Let $i^* \in J$ such that

$$\inf_{i\in J} \Lambda(F)(x_i) = \Lambda(F)(x_{i^*}).$$

We have for all $c \in]x_J^-, x_J^+[$,

$$\Lambda(F)(c) \leq \Lambda(F)(x_{i^*}) + \omega(h_J).$$

Besides, by taking $\boldsymbol{f} = (x_1^r/R(x_1), ..., x_n^r/R(x_n))'$ in (44) we get that

$$\frac{1}{N_J}\sum_{\boldsymbol{i}\in\mathcal{I}_r\cap J^{r+1}} \phi^2_{\boldsymbol{i}}(\boldsymbol{x}^r) = \frac{\|\boldsymbol{x}^r_J - \Pi_{\mathcal{X}_J}\boldsymbol{x}^r_J\|^2}{\gamma_J}.$$

Now, by using (70) and (24) we deduce that

$$\begin{aligned}
&< \boldsymbol{f}, \boldsymbol{t}_J^* > \\
&\geq -\frac{\Lambda(F)(x_{i^*}) + \omega(h_J)}{r!} \left(\frac{1}{N_J} \sum_{\boldsymbol{i} \in \mathcal{I}_r \cap J^{r+1}} \phi_{\boldsymbol{i}}^2(\boldsymbol{x}^r) \right) \\
&= -\left(\Lambda(F)(x_{i^*}) + \omega(h_J)\right) \frac{\|\boldsymbol{x}_J^r - \Pi_{\mathcal{X}_J} \boldsymbol{x}_J^r\|^2}{\gamma_J r!} \\
&\geq v_{\boldsymbol{t}_J^*}(\boldsymbol{f}, \beta),
\end{aligned}$$

and we conclude thanks to Theorem 1.

9.2. *Proof of the second part of Proposition 3.* In order to prove this second part, we apply the first part of the proposition.

Evaluation of $v_{\boldsymbol{t}_J^*}(\boldsymbol{f}, \beta)$. Let us prove that for all $J \in \cup_{\ell=1}^{\ell_n} \mathcal{J}^{(\ell)}$,

$$v_{\boldsymbol{t}_J^*}(\boldsymbol{f}, \beta) \leq \kappa^* \sqrt{\log(n)} \sigma$$

where κ^* depends on α, β, s, r only. We use Steps 1 to 5 in the proof of Proposition 1. For Steps 1,2 and 5 the proof is similar as in the proof of Proposition 1.

Step 3. For all $\boldsymbol{f} = (F(x_1), ..., F(x_n))'$ with $F^{(r)} \in \mathcal{H}_s(L)$,

$$\frac{\|\boldsymbol{f} - \Pi_{V_n} \boldsymbol{f}\|^2}{n} \leq C(s, r) L^2 n^{-2(s+r)}. \tag{71}$$

Proof of step 3 We recall that V_n is the linear space generated by

$$\left\{ \mathbf{1}_J, \boldsymbol{x}_J, ..., \boldsymbol{x}_J^r, J \in \mathcal{J}^{\ell_n} \right\}.$$

Note that the vector

$$\widetilde{\boldsymbol{f}} = \sum_{k=1}^{\ell_n} \left(F(\bar{x}_{J_k}) \mathbf{1}_{J_k} + \sum_{l=1}^{r} \frac{F^{(l)}(\bar{x}_{J_k})}{l!} (\boldsymbol{x}_{J_k} - \bar{x}_{J_k} \mathbf{1}_{J_k})^l \right)$$

belongs to V_n. Hence, using that $F^{(r)} \in \mathcal{H}_s(L)$,

$$\begin{aligned}
&\|\boldsymbol{f} - \Pi_{V_n} \boldsymbol{f}\|^2 \\
&\leq \|\boldsymbol{f} - \widetilde{\boldsymbol{f}}\|^2 \\
&= \sum_{k=1}^{\ell_n} \sum_{i \in J_k} \left(\int_{u_1 = \bar{x}_{J_k}}^{x_i} \int_{u_2 = \bar{x}_{J_k}}^{u_1} \cdots \int_{u_r = \bar{x}_{J_k}}^{u_{r-1}} \left(F^{(r)}(u_r) - F^{(r)}(\bar{x}_{J_k}) \right) du_r \ldots du_1 \right)^2 \\
&\leq \sum_{k=1}^{\ell_n} \sum_{i \in J_k} L^2 \ell_n^{-2(r+s)} \\
&\leq C(s, r) L^2 n^{1-2(r+s)}
\end{aligned}$$

since $\ell_n \geq n/(4(r+1))$ using that $[x] \geq x/2$ for $x \geq 1$.

Step 4. Assuming that $n \geq (L/\sigma)^{1/(r+s)}$, there exists some constant C depending on s, r and β only such that

$$\frac{\bar{\chi}^{-1}_{n-d_n, \|\boldsymbol{f}-\Pi_{V_n}\boldsymbol{f}\|^2/\sigma^2}(\beta/2)}{n-d_n} \leq C. \tag{72}$$

The proof is similar to the proof of Step 4 in Proposition 1 by using (71).

Evaluation of γ_J. Let us prove that there exists some constant C depending on r only such that, for J such that $|J| \geq r+1$,

$$\gamma_J^2 \geq C \frac{|J|^{2r+1}}{n^{2r}}.$$

Since for all i, $x_i = i/n$, by translation

$$\begin{aligned}\gamma_J^2 &= \|\boldsymbol{x}_J^r - \Pi_{\mathcal{X}_J}\boldsymbol{x}_J^r\|^2 \\ &= \frac{1}{n^{2r}} \min_{a_0,\ldots,a_{r-1}} \sum_{i=1}^{|J|} \left(i^r - a_0 - a_1 i - \ldots - a_{r-1} i^{r-1}\right)^2.\end{aligned}$$

By setting for all $j \in \{0, \ldots, r-1\}$ $a_j = b_j |J|^{r-j}$, we have

$$\begin{aligned}&\min_{a_0,\ldots,a_{r-1}} \sum_{i=1}^{|J|} \left(i^r - a_0 - a_1 i - \ldots - a_{r-1} i^{r-1}\right)^2 \\ &= |J|^{2r+1} \min_{b_0,\ldots,b_{r-1}} \frac{1}{|J|} \sum_{i=1}^{|J|} \left(\left(\frac{i}{|J|}\right)^r - b_0 - \ldots - b_{r-1}\left(\frac{i}{|J|}\right)^{r-1}\right)^2.\end{aligned}$$

Since

$$\min_{b_0,\ldots,b_{r-1}} \frac{1}{|J|} \sum_{i=1}^{|J|} \left(\left(\frac{i}{|J|}\right)^r - b_0 - \ldots - b_{r-1}\left(\frac{i}{|J|}\right)^{r-1}\right)^2$$

converges as $|J| \to \infty$ towards

$$\min_{b_0,\ldots,b_{r-1}} \int_0^1 \left(x^r - b_0 - \ldots - b_{r-1} x^{r-1}\right)^2 dx,$$

which is positive, we obtain that there exists some constant $C > 0$ such that for $|J|$ large enough,

$$\gamma_J^2 \geq C \frac{|J|^{2r+1}}{n^{2r}}.$$

Moreover, since for $|J| \geq r+1$, $\gamma_J^2 > 0$, the above inequality holds for $|J| \geq r+1$, possibly enlarging C.

Evaluation of $\omega(h_J)$. Let $J \in \mathcal{J}^{(\ell)}$. Since $F^{(r)} \in \mathcal{H}_s(L)$, and since h_J defined in Theorem 3 satisfies $0 < h_J \leq 1/\ell$,

$$\begin{aligned}\omega(h_J) &= \sup_{|x-y|\leq h_J} |F^{(r)}(x) - F^{(r)}(y)| \\ &\leq L\ell^{-s}.\end{aligned}$$

Conclusion. Let us prove in conclusion that if

$$\inf_{x\in[0,1]} F^{(r)}(x) \leq -\rho_{n,r} \tag{73}$$

then (24) holds for some $J \in \cup_{\ell=1}^{\ell_n} \mathcal{J}^{(\ell)}$.

Since, $F^{(r)} \in \mathcal{H}_s(L)$ under (73), there exists $j \in \{1, \ldots, n\}$ such that

$$F^{(r)}(x_j) \leq -\rho_{n,r} + Ln^{-s} \leq -\rho_{n,r}/2$$

for n large enough.
Let

$$\ell(n) = \left[\left(\frac{L^2 n}{\sigma^2 \log(n)}\right)^{1/(1+2r+2s)}\right].$$

For n large enough, $\ell(n) \in \{1, \ldots, \ell_n\}$. Let J be the element of $\mathcal{J}^{(\ell(n))}$ containing j. Note that $|J| \geq n/(2\ell(n))$ at least for n large enough. This implies that, for n large enough,

$$\gamma_J^2 \geq C\frac{|J|^{1+2r}}{n^{2r}} \geq C(r)n\left(\ell(n)\right)^{-1-2r}.$$

It follows that

$$\begin{aligned}v_{\boldsymbol{t}_J^*}(\boldsymbol{f}, \beta)\frac{r!}{\gamma_J} + \omega(h_J) &\leq \frac{\kappa^* r!}{\sqrt{C(r)}}\sigma\sqrt{\log(n)}\frac{(\ell(n))^{r+1/2}}{\sqrt{n}} + L(\ell(n))^{-s} \\ &\leq \kappa L^{\frac{1+2r}{1+2r+2s}}\left(\frac{\sigma^2\log(n)}{n}\right)^{\frac{s}{1+2s+2r}}\end{aligned}$$

for some constant κ depending on α, β, s, r. This concludes the proof of the proposition.

10. Appendix

10.1. *Proof of Lemma 2.*

Proof of (42). Clearly, on has $\tilde{\mathcal{C}}_{r,R} \subset \mathcal{C}_{r,R}$. We prove $\mathcal{C}_{r,R} \subset \tilde{\mathcal{C}}_{r,R}$ by using repeatedly the following claim.

Claim 1. *Let $0 \leq u_1 < u_2 < ... < u_{r+1} < u_{r+2} \leq 1$ be an increasing sequence of $r+2$ points of $[0,1]$. Let $v_1, ..., v_{r+2}$ be real numbers verifying that*

$$D_1(1, u_{r+2}, ..., u_{r+2}^{r-1}, v_{r+2}) = \det \begin{pmatrix} 1 & u_2 & \cdots & u_2^{r-1} & v_2 \\ 1 & u_3 & \cdots & u_3^{r-1} & v_3 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & u_{r+2} & \cdots & u_{r+2}^{r-1} & v_{r+2} \end{pmatrix} \geq 0$$

and

$$D_{r+2} = \det \begin{pmatrix} 1 & u_1 & \cdots & u_1^{r-1} & v_1 \\ 1 & u_2 & \cdots & u_2^{r-1} & v_2 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & u_{r+1} & \cdots & u_{r+1}^{r-1} & v_{r+1} \end{pmatrix} \geq 0.$$

Then for all $j \in \{2, ..., r+1\}$

$$D_j(1, u_{r+2}, ..., u_{r+2}^{r-1}, v_{r+2}) = \det \begin{pmatrix} 1 & u_1 & \cdots & u_1^{r-1} & v_1 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & u_{j-1} & \cdots & u_{j-1}^{r-1} & v_{j-1} \\ 1 & u_{j+1} & \cdots & u_{j+1}^{r-1} & v_{j+1} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & u_{r+2} & \cdots & u_{r+2}^{r-1} & v_{r+2} \end{pmatrix} \geq 0.$$

Proof of the Claim. For real numbers $t_1, ..., t_r$ we set $\mathrm{vand}(t_1, ..., t_r)$ the Vandermond determinant

$$\mathrm{vand}(t_1, ..., t_r) = \det \begin{pmatrix} 1 & t_1 & \cdots & t_1^{r-1} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & t_r & \cdots & t_r^{r-1} \end{pmatrix}$$

and for $j = 1, ..., r+2$ we denote by $\boldsymbol{u}_j$ the vector $(1, u_j, ..., u_j^{r-1}, v_j)'$. Let us fix $j \in \{2, ..., r+1\}$. By expanding the determinant

$$D_j(1, u_{r+2}, ..., u_{r+2}^{r-1}, v_{r+2})$$

by its last column we get that if $j \in \{2, ..., r\}$,

$$\begin{aligned} D_j(1, u_{r+2}, ..., u_{r+2}^{r-1}, v_{r+2}) &= v_{r+2}\mathrm{vand}(u_1, ..., u_{j-1}, u_{j+1}, ..., u_{r+1}) \\ &\quad + D_j(1, u_{r+2}, ..., u_{r+2}^{r-1}, 0), \end{aligned}$$

and if $j = r+1$

$$D_{r+1}(1, u_{r+2}, ..., u_{r+2}^{r-1}, v_{r+2}) = v_{r+2}\mathrm{vand}(u_1, ..., u_r) + D_{r+1}(1, u_{r+2}, ..., u_{r+2}^{r-1}, 0).$$

Since the u_i's are increasing, the Vandermond determinants are positive and therefore $D_j(1, u_{r+2}, ..., u_{r+2}^{r-1}, v_{r+2})$ is increasing with respect to v_{r+2}. On the other hand, since by assumption

$$D_1(1, u_{r+2}, ..., u_{r+2}^{r-1}, v_{r+2}) = v_{r+2}\text{vand}(u_2, ..., u_{r+1}) + D_1(1, u_{r+2}, ..., u_{r+2}^{r-1}, 0) \geq 0$$

we have that

$$v_{r+2} \geq -\frac{D_1(1, u_{r+2}, ..., u_{r+2}^{r-1}, 0)}{\text{vand}(u_2, ..., u_{r+1})} = v^*,$$

and deduce that

$$D_j(1, u_{r+2}, ..., u_{r+2}^{r-1}, v_{r+2}) \geq D_j(1, u_{r+2}, ..., u_{r+2}^{r-1}, v^*).$$

It remains to show that $D_j(1, u_{r+2}, ..., u_{r+2}^{r-1}, v_{r+2}, v^*) \geq 0$. When $v_{r+2} = v^*$ we have that $D_1(1, u_{r+2}, ..., u_{r+2}^{r-1}, v^*) = 0$ and therefore $\boldsymbol{u}^* = (1, u_{r+2}, ..., u_{r+2}^{r-1}, v^*)'$ is a linear combination of $\boldsymbol{u}_2$,...,$\boldsymbol{u}_{r+1}$. Let us denote by λ_k the coordinate of $\boldsymbol{u}^*$ on $\boldsymbol{u}_k$. By Cramer's formula we have that for $k \in \{3, ..., r\}$

$$\begin{aligned}\lambda_k &= \frac{\text{vand}(u_2, ..., u_{k-1}, u_{k+1}, ..., u_{r+2})}{\text{vand}(u_2, ..., u_{k-1}, u_{k+1}, ..., u_{r+1}, u_k)} \\ &= (-1)^{r-k+1}\frac{\text{vand}(u_2, ..., u_{k-1}, u_{k+1}, ..., u_{r+2})}{\text{vand}(u_2, ..., u_{r+1})},\end{aligned}$$

$$\lambda_2 = (-1)^{r-1}\frac{\text{vand}(u_3, ..., u_{r+2})}{\text{vand}(u_2, ..., u_{r+1})} \quad \text{and } \lambda_{r+1} = \frac{\text{vand}(u_2, ..., u_r, u_{r+2})}{\text{vand}(u_2, ..., u_{r+1})}.$$

Hence, the positivity of the Vandermond determinants implies that λ_j is of the sign of $(-1)^{r-j+1}$. Since $\boldsymbol{u}^* = \sum_{k=2}^{r+1} \lambda_k \boldsymbol{u}_k$, by linearity of the determinant

$$\begin{aligned}D_j(1, u_{r+2}, ..., u_{r+2}^{r-1}, v^*) &= \lambda_j D_j(1, u_j, ..., u_j^{r-1}, v_j) \\ &= (-1)^{r-j+1}\lambda_j D_{r+1}\end{aligned}$$

and thus, as $D_{r+1} \geq 0$, $D_j(1, u_{r+2}, ..., u_{r+2}^{r-1}, v^*) \geq 0$.

Proof of (43). For $x \in [x_{i_1}, x_{i_{r+1}}]$ let us set

$$\begin{aligned}h(x) = \det &\begin{pmatrix} 1 & x & \cdots & x^{r-1} & R(x)F(x) \\ 1 & x_{i_2} & \cdots & x_{i_2}^{r-1} & R(x_{i_2})F(x_{i_2}) \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_{i_{r+1}} & \cdots & x_{i_{r+1}}^{r-1} & R(x_{i_{r+1}})F(x_{i_{r+1}}) \end{pmatrix} \\ &-\lambda\det\begin{pmatrix} 1 & x & \cdots & x^{r-1} & x^r \\ 1 & x_{i_2} & \cdots & x_{i_2}^{r-1} & x_{i_2}^r \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_{i_{r+1}} & \cdots & x_{i_{r+1}}^{r-1} & x_{i_{r+1}}^r \end{pmatrix},\end{aligned}$$

where λ is such that $h(x_1) = 0$. Since h is r-times differentiable and satisfies $h(x_{i_1}) = h(x_{i_2}) = \cdots = h(x_{ir+1}) = 0$, there exists some $c_{\boldsymbol{i}} \in]x_{i_1}, x_{i_{r+1}}[$ such that

$$0 = h^{(r)}(c_{\boldsymbol{i}}) = \det \begin{pmatrix} 0 & 0 & \cdots & 0 & \Lambda(F)(c_{\boldsymbol{i}}) \\ 1 & x_{i_2} & \cdots & x_{i_2}^{r-1} & R(x_{i_2})F(x_{i_2}) \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_{i_{r+1}} & \cdots & x_{i_{r+1}}^{r-1} & R(x_{i_{r+1}})F(x_{i_{r+1}}) \end{pmatrix}$$

$$-\lambda \det \begin{pmatrix} 0 & 0 & \cdots & 0 & r! \\ 1 & x_{i_2} & \cdots & x_{i_2}^{r-1} & x_{i_2}^{r} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_{i_{r+1}} & \cdots & x_{i_{r+1}}^{r-1} & x_{i_{r+1}}^{r} \end{pmatrix},$$

leading to $\lambda = \Lambda(F)(c_{\boldsymbol{i}})/r!$. We get the result by substituting the expression of λ in the equality $h(x_1) = 0$.

Proof of (44). We start with the following claim.

Claim 2. *Let $\mathcal{W}$ be a linear subspace of $\mathbb{R}^k$ of dimension $q \in \{1, ..., k-1\}$ and $\{\boldsymbol{w}^1, ..., \boldsymbol{w}^q\}$ a basis of $\mathcal{W}$. Then for all $\boldsymbol{u}, \boldsymbol{v}$ in $\mathbb{R}^k$*

$$\begin{aligned} &\mathrm{Gram}(\boldsymbol{w}^1, ..., \boldsymbol{w}^q) < \boldsymbol{u}, (I - \Pi_{\mathcal{W}})\boldsymbol{v} > \\ &= \sum_{\boldsymbol{i} \in \mathcal{I}_{q+1}} \det \begin{pmatrix} w_{i_1}^1 & \cdots & w_{i_1}^q & u_{i_1} \\ \vdots & \vdots & \vdots & \vdots \\ w_{i_{q+1}}^1 & \cdots & w_{i_{q+1}}^q & u_{i_{q+1}} \end{pmatrix} \det \begin{pmatrix} w_{i_1}^1 & \cdots & w_{i_1}^q & v_{i_1} \\ \vdots & \vdots & \vdots & \vdots \\ w_{i_{q+1}}^1 & \cdots & w_{i_{q+1}}^q & v_{i_{q+1}} \end{pmatrix} \end{aligned} \tag{74}$$

where

$$\mathrm{Gram}(\boldsymbol{w}^1, ..., \boldsymbol{w}^q) = \det(G) \text{ with } G = (< \boldsymbol{w}^i, \boldsymbol{w}^j >)_{1 \le i,j \le q}.$$

We conclude thanks to the claim by taking $\boldsymbol{u} = R \star \boldsymbol{f}$, $\boldsymbol{v} = \boldsymbol{x}_J^r - \Pi_{\mathcal{X}_J} \boldsymbol{x}_J^r$, $\mathcal{W} = \mathcal{X}_J$ and $k = |J|$.

Proof of the Claim. For $\boldsymbol{z} \in \mathbb{R}^k$, let $B(\boldsymbol{z})$ the $k \times (q+1)$ matrix

$$B(\boldsymbol{z}) = \begin{pmatrix} w_1^1 & \cdots & w_1^q & z_1 \\ \vdots & \vdots & \vdots & \vdots \\ w_k^1 & \cdots & w_k^q & z_k \end{pmatrix}.$$

We obtain the result by computing

$$\det(B(\boldsymbol{u})' B(\boldsymbol{v})) = \det \begin{pmatrix} < \boldsymbol{w}^1, \boldsymbol{w}^1 > & \cdots & < \boldsymbol{w}^1, \boldsymbol{w}^q > & < \boldsymbol{w}_1, \boldsymbol{v} > \\ \vdots & \vdots & \vdots & \vdots \\ < \boldsymbol{w}^q, \boldsymbol{w}^1 > & \cdots & < \boldsymbol{w}^q, \boldsymbol{w}^q > & < \boldsymbol{w}_q, \boldsymbol{v} > \\ < \boldsymbol{u}, \boldsymbol{w}^1 > & \cdots & < \boldsymbol{u}, \boldsymbol{w}^q > & < \boldsymbol{u}, \boldsymbol{v} > \end{pmatrix}$$

by two different ways. The first way is direct: since $\Pi_W \boldsymbol{v}$ is a linear combination of the $\boldsymbol{w}^j$'s we have

$$\begin{aligned}
&\det\left(B(\boldsymbol{u})'B(\boldsymbol{v})\right)\\
&= \det\begin{pmatrix} <\boldsymbol{w}^1,\boldsymbol{w}^1> & \cdots & <\boldsymbol{w}^1,\boldsymbol{w}^q> & <\boldsymbol{w}^1,(I-\Pi_W)\boldsymbol{v}>\\ \vdots & \vdots & \vdots & \vdots\\ <\boldsymbol{w}^q,\boldsymbol{w}^1> & \cdots & <\boldsymbol{w}^q,\boldsymbol{w}^q> & <\boldsymbol{w}^q,(I-\Pi_W)\boldsymbol{v}>\\ <\boldsymbol{u},\boldsymbol{w}^1> & \cdots & <\boldsymbol{u},\boldsymbol{w}^q> & <\boldsymbol{u},(I-\Pi_W)\boldsymbol{v}> \end{pmatrix}\\
&= \det\begin{pmatrix} <\boldsymbol{w}^1,\boldsymbol{w}^1> & \cdots & <\boldsymbol{w}^1,\boldsymbol{w}^q> & 0\\ \vdots & \vdots & \vdots & \vdots\\ <\boldsymbol{w}^q,\boldsymbol{w}^1> & \cdots & <\boldsymbol{w}^q,\boldsymbol{w}^q> & 0\\ <\boldsymbol{u},\boldsymbol{w}^1> & \cdots & <\boldsymbol{u},\boldsymbol{w}^q> & <\boldsymbol{u},(I-\Pi_W)\boldsymbol{v}> \end{pmatrix}\\
&= \mathrm{Gram}(\boldsymbol{w}^1,...,\boldsymbol{w}^q) <\boldsymbol{u},(I-\Pi_W)\boldsymbol{v}> .
\end{aligned}$$

The other way is to use the Cauchy-Binet formula (see Horn and Johnson (1991)): we calculate $\det\left(B(\boldsymbol{u})'B(\boldsymbol{v})\right)$ as a function of the $(q+1)\times(q+1)$ minors of the matrix $B(\boldsymbol{u})$ and $B(\boldsymbol{v})$ which leads to the right-hand side of (74) and concludes the proof.

10.2. *Proof of Proposition 2.*
Case $\mathcal{K}=\mathcal{K}_{\geq 0}$. Let $\mathbb{P}_F$ be the law of Y. under the model defined by Equation (16). Let Φ be a test of level α of hypothesis $F\in\mathcal{K}_{\geq 0}$. Let us define the test Ψ of hypothesis "$F=0$" against "$F\neq 0$" which rejects the null if $\Phi(Y)=1$ or if $\Phi(-Y)=1$. Since $0\in\mathcal{K}_{\geq 0}$ and since

$$\mathbb{P}_0(\Phi(Y)=1)=\mathbb{P}_0(\Phi(-Y)=1)\leq\alpha,$$

the test Ψ is of level $2\alpha\leq 3\alpha$. Let $\rho_n(\Phi,\mathcal{F})$ be the Δ-uniform separation rate of Φ over $\mathcal{F}$. It is enough to show that

$$\rho_n(\Phi,\mathcal{F})\geq\rho_n(0,\mathcal{F}).$$

To do so, we show that the $\|\ \|_\infty$-uniform separation rate of Ψ over $\mathcal{F}$ is not larger than $\rho_n(\Phi,\mathcal{F})$ which means that for all $F\in\mathcal{F}$ such that $\|F\|_\infty\geq\rho_n(\Phi,\mathcal{F})$ we have $\mathbb{P}_F\left(\Psi(Y)=1\right)\geq 1-\beta$.
Let $F\in\mathcal{F}$. If $\|F\|_\infty\geq\rho_n(\Phi,\mathcal{F})$ then

$$\text{either}\ \ \Delta(F)=\sup_{x\in[0,1]}\left(-F(x)\mathbf{1}_{F(x)>0}\right)\geq\rho_n(\Phi,\mathcal{F})\ \ \text{or}\ \ \Delta(-F)\geq\rho_n(\Phi,\mathcal{F}).$$

In the first case, by definition of $\rho_n(\Phi,\mathcal{F})$ we have $\mathbb{P}_F\left(\Phi(Y)=1\right)\geq 1-\beta$ and consequently $\mathbb{P}_F\left(\Psi(Y)=1\right)\geq 1-\beta$. Note that in the other case the same is true since by symmetry of the law of $Y-F$

$$\mathbb{P}_F\left(\Phi(-Y)=1\right)=\mathbb{P}_{-F}\left(\Phi(Y)=1\right).$$

Case $\mathcal{K} = \mathcal{K}_{\nearrow}$. We argue similarly. Let Φ be a test of level α of hypothesis $F \in \mathcal{K}_{\nearrow}$. We also consider the test Φ' of level α of "$F = 0$" against "$F \neq 0$" which rejects the null when $\sqrt{n}\left|\int_0^1 dY(t)\right|$ is large enough (namely, larger than the $1-\alpha$ quantile of a standard Gaussian random variable). Finally, we define the test Ψ of hypothesis "$F = 0$" against "$F \neq 0$" which rejects the null if $\Phi(Y) = 1$ or $\Phi(-Y) = 1$ or $\Phi'(Y) = 1$. Since $0 \in \mathcal{K}_{\nearrow}$, we have that the so-defined test Ψ is of level 3α.
Some easy computations shows that there exists some constant κ depending on α and β only such that Φ' rejects the null with probability not smaller than $1-\beta$ as soon as $\left|\int_0^1 F(t)dt\right|$ is larger than $\kappa\sigma/\sqrt{n}$ (the sum of the β and $1-\alpha$ quantiles of a standard Gaussian suits for κ). On the other hand, note that

$$\Delta(F) = \frac{1}{2}\sup_{0\le s\le t\le 1}(F(s) - F(t))$$

and thus, by definition of the Δ-separation rate, $\rho_n(\Phi, \mathcal{F})$, of Φ over $\mathcal{F}$, Ψ rejects the null with probability not smaller than $1-\beta$ under all alternatives $F \in \mathcal{F}$ satisfying

$$\max\{\Delta(F), \Delta(-F)\} = \frac{1}{2}\sup_{0\le t,s\le 1}|F(t) - F(s)| > \rho_n(\Phi, \mathcal{F}).$$

Therefore, since

$$\begin{aligned}
\|F\|_\infty &\le \sup_{t\in[0,1]}\left|F(t) - \int_0^1 F(s)ds\right| + \left|\int_0^1 F(s)ds\right| \\
&\le \int_0^1 \sup_{t\in[0,1]}|F(t) - F(s)|\,ds + \left|\int_0^1 F(s)ds\right| \\
&\le \sup_{t,s\in[0,1]}|F(t) - F(s)| + \left|\int_0^1 F(s)ds\right|
\end{aligned}$$

Ψ rejects the null with probability larger than $1-\beta$ under all alternative $\mathcal{F}$ such that

$$\|F\|_\infty \ge 2\rho_n(\Phi, \mathcal{F}) + \kappa\sigma/\sqrt{n},$$

and the result follows.

REFERENCES

Baraud, Y., Huet, S., and Laurent, B. 2003. Adaptive tests of linear hypotheses by Model Selection. *Ann. Statist.*, Vol. 31, pp. 225–251.

Baraud, Y., Huet, S., and Laurent, B. 2003. Adaptive tests of qualitative hypotheses. *ESAIM: P & S.* , Vol. 7, pp. 147–159.

Birgé, L. (2001). An alternative point of view on Lepski's method. *State of the art in probability and statistics (Leiden, 1999)* (ed. Monogr., IMS Lecture Notes), pp. 113–133, vol. 36. Physica-Verlag, Beachwood.

Bowman, A.W., Jones, M.C., and Gijbels, I. (1998). Testing monotonicity of regression. *J. Comput. Graph. Statist.*, Vol. 7, pp. 489–500.

Dümbgen, L., and Spokoiny, V.G. (2001). Multiscale testing of qualitative hypotheses. *Ann. Statist.*, Vol. 29, pp. 124–152.

Ghosal, S, Sen, A. and Van der Vaart, A. (2001). Testing monotonicity of regression. *Ann. Statist.*, Vol. 28, pp. 1054–1082.

Gijbels, I., Hall, P., Jones, M.C., and Koch, I. (2000). Tests for monotonicity of a regression mean with guaranteed level. *Biometrika*, Vol. 87, pp. 663–673.

Hall, P., and Heckman, N.E. (2000). Testing for monotonicity of a regression mean by calibrating for linear functions. *The Annals of Statistics*, Vol. 28, pp. 20–39.

Horn, R.A., and Johnson, C.R. (1991). *Topics in matrix analysis.* Cambridge University Press, Cambridge.

Ingster, Y. (1993). Asymptotic minimax hypothesis testing for nonparametric alternatives I. *Math. Methods Statist.*, Vol. 2, pp. 85–114.

Juditsky, A., and Nemirovski, A. (2002). On nonparametric tests of positivity / monotonicity / convexity. *The Annals of Statistics*, Vol. 30.

Silverman, B.W. (1981). Using kernel density estimates to investigate multimodality. *J. R. Statist. Soc. B*, Vol. 43, pp. 97–9.

Acknowledgment: the authors would like to thank Claudine Picaronny for pointing us this proof of (74)