Report 99-001

A consistent nonparametric test of the convexity of regression based on least squares splines

Cheikh A.T.Diack

A Consistent Nonparametric Test of the Convexity of Regression Based on Least Squares Splines

Cheikh A. T. DIACK
EURANDOM, P.O. Box 513, 5600 Eindhoven,
The Netherlands. E-mail:diack@eurandom.tue.nl

January 15, 1999

Abstract

This paper provides a test of convexity of a regression function. The basic idea uses an interesting connection between (1) a hypothesis test of convexity of a nonparametric regression function based on a cubic splines estimator and (2) a hypothesis test for normal means constrained by linear inequalities. The test statistic is shown to be asymptotically of size equal to the nominal level, while diverging to infinity if the convexity is misspecified. Therefore, the test is consistent against all deviations from the null hypothesis. The behavior of the test under the local alternatives is studied.

1 INTRODUCTION

Tests of convexity of a regression function is one of the most important problems in econometrics. Indeed, "*The General Theory of Employment, Interest, and Money* emphasized the central importance of the consumption function and explicitly argued that the consumption function is concave" (Carroll & Kimball 1996). Economic theory predicts also the convexity of functions like for example Bernoulli utility function, cost function, production function, Engels curves, etc. Besides, the Human Capital theory argued that the relationships between the logarithm of wage and the experience is concave.
On the other hand, psychologists have worried for over a century about whether subjective reports about physical magnitudes like length, weight, area, luminance etc. have a convex or concave relationship to corresponding measurement. Also, this convexity problem is very closely connected to the order-restricted hypothesis testing problems described in references such as Robertson et al. (1988).

There are some papers in the statistics literature dealing with nonparametric hypothesis tests of convexity of the regression function. The work along this line includes Schlee (1980), Yatchew (1992), Diack (1996), Diack & Thomas (1998) and Diack (1998).
Schlee (1980) in a nonparametric regression model with random design used an estimator of the second derivative of the regression function. His test statistic requires computing the distribution of the supremum of this normalized estimator over an interval. But this method imposes some theorecal difficulties. To overcome the problem, he proposes a sequence of points from the interval and uses the theory of maximal deviation to obtain the distribution of the test statistic under the null

hypothesis. However, this work does not discuss asymptotic results or pratical implementation.
Yatchew's test (with semi-parametric model) is based on comparing the nonparametric sum of squared residuals with convexity constraints, with the nonparametric sum of squared residuals without contraints. Yatchew's approach relies on sample splitting which results in a loss of efficiency. He gives a heuristic proof of the consistency of the test. Diack (1998) adapts respectively Schlee's idea and Yatchew's idea in a nonparametric model with fixed design to construct two other tests of convexity for which he gives new asymptotic results of convergence.
Diack and Thomas (1998) use a least-squares splines estimator and develop in a nonparametric model with deterministic design, a non-convexity test which is consistent for some alternative hypothesis. A small simulation study in Diack(1996) shows that the test is satisfactory for finite sample sizes.

In this paper, we propose a new test of convexity of a regression function in a nonparametric model. Our test uses, as Diack and Thomas' test, a cubic spline estimator which allows us to formulate the convexity hypothesis in a very simple way. Hence, our problem becomes roughly, a problem to test a multivariate normal mean with composite hypotheses determined by linear inequalities.

The remainder of this paper is organized as follows. In **section 2**, we introduce the nonparametric regression model and the hypotheses to be tested. After, we recall some properties of the cubic spline estimator. **Section 3** describes our test of convexity of the regression function, and **section 4** is devoted to a discussion and demonstration of some properties of the test.

2 PRELIMINARIES

2.1 The model and the hypotheses

Consider the nonparametric regression model:

$$y_{ij} = f(x_i) + \varepsilon_{ij}, i = 1, ..., r, j = 1, ..., n_i \quad x_i \in (0,1), i = 1, ..., r.$$

At each deterministic design point $x_i, (i = 1, ..., r), n_i$ measurements are taken. The probability measure assigning mass $\mu_i = n_i/n$ to the point $x_i(\sum \mu_i = 1)$ is referred to as the design and will be denoted by μ^n. We assume that the random errors ε_{ij} are uncorrelated and identically distributed with mean zero. Their variance σ^2 will be assumed unknown. Finally f is an unknown smooth regression function.
In what follows, we will assume some regularity conditions on f.

The following class of functions were used by Diack and Thomas (1998) to construct a test of non-convexity.

For $l \in I\!\!N$ and $M > 0$, let

$$\mathcal{F}_{l,M} = \{f \in C^{l+1}(0,1) : \sup_{0 \leq x \leq 1} \mid f^{((l+1)\wedge 4)}(x) \mid \leq M\}.$$

We intend to contruct a test of H_0 : "f is convex" versus H_1 : "f is non-convex."

It would be interesting to know how the test behaves under the local alternatives. So, we might consider a sequence of local alternatives H_{1n} : "$f_n = f_0 + h_n L$" where f_0 is a fixed function in the null hypothesis and L is known and lies in $\mathcal{F}_{l,M}$.

Throughout this paper, a testing problem with null hypothesis H_0 and alternative H_1 is denoted by $[H_0, H_1]$.

We will use a cubic spline estimator and characterize convexity in the set of all polynomial cubic splines to transform our problem into a test of a multivariate normal mean with composite hypotheses determined by linear inequalities.

2.2 The Cubic Spline Estimator

Let p be a positive continuous density on (0,1). We assume that

$$\min_{0\leq x\leq 1} p(x) > 0.$$

Let $\eta_0 = 0 < \eta_1 < ... < \eta_{k+1} = 1$ be a subdivision of the interval $(0, 1)$ by k distinct points defined by

$$\int_0^{\eta_i} p(x)\, dx = i/(k+1), \quad i = 0, ..., k+1. \tag{1}$$

Let $\delta_k = \max_{0\leq i\leq k}(\eta_{i+1} - \eta_i)$.

For each fixed set of knots of the form (1), we define $\mathcal{S}(k, d)$ as the collection of all polynomial splines of order d (degree$\leq d-1$) having for knots $\eta_1 < ... < \eta_k$. The class $\mathcal{S}(k, d)$ of such splines is a linear space of functions of dimension $(k+d)$. A basis for this linear space is provided by the B-splines(see Schumaker 1981). Let $\{N_1, .., N_{k+d}\}$ denote the set of normalized B-splines associated with the following nondecreasing sequence $\{t_1, .., t_{k+2d}\}$:

$$\begin{cases} t_1\leq t_2\leq ...\leq t_d = 0 \\ t_{2d+k} \geq t_{2d+k-1} \geq ... \geq t_{d+k+1} = 1 \\ t_{d+l} = \eta_l \quad \text{for} \quad l = 1, .., k \end{cases}$$

The reader is referred to Schumaker (1981) for a discussion of these B-splines.

In what follows, we shall only work with the class of cubic splines: $\mathcal{S}(k, 4)$. It will be convenient to introduce the following notations:

$N(x) = (N_1(x), .., N_{k+4}(x))' \in I\!R^{k+4}$ and $F = (N(x_1), .., N(x_r)); (k+4) \times r$ matrix.

We will denote by $\widehat{f}_n$ the least squares spline estimator of f :

$$\widehat{f}_n(x) = \sum_{p=1}^{k+4} \widehat{\theta}_p N_p(x) \tag{2}$$

where

$$\widehat{\Theta} = (\hat{\theta}_1, ..., \hat{\theta}_{k+4})' = \arg \min_{\Theta\in I\!R^{k+4}} \sum_{i=1}^{r}\sum_{j=1}^{n_i} \left(y_{ij} - \sum_{p=1}^{k+4} \theta_p N_p(x_i) \right)^2 . \tag{3}$$

Let

$$\begin{aligned} \bar{y}_i &= \frac{1}{n_i}\sum_{j=1}^{n_i} y_{ij},\ \bar{\varepsilon}_i = \frac{1}{n_i}\sum_{j=1}^{n_i} \varepsilon_{ij}, \\ \bar{Y} &= (\bar{y}_1, .., \bar{y}_r)', \bar{\varepsilon} = (\bar{\varepsilon}_1, .., \bar{\varepsilon}_r)' \text{ and } f_\Delta = (f(x_1), .., f(x_r))' \end{aligned}$$

Let $\mathcal{D}(\mu^n)$ be the $r \times r$ diagonal matrix with diagonal elements $\mu_1, ..., \mu_r$, then, basic least squares arguments prove that:

$$\widehat{\Theta} = M^{-1}(\mu^n)F\mathcal{D}(\mu^n)\bar{Y} \quad \text{with} \quad M(\mu^n) = \sum_{i=1}^{r} N(x_i)N'(x_i)\mu_i = F\mathcal{D}(\mu^n)F'.$$

Asymptotic properties of this estimator have been established in Argarwal and Studden(1980).

Note that the first moment of $\widehat{f}_n$ is given by

$$I\!E\widehat{f}_n(x) = N(x)'M^{-1}(\mu^n)F\mathcal{D}(\mu^n)f_\Delta.$$

Thus, if f is a cubic spline function (that is to say there is Θ such that $f_\Delta = F'\Theta$) then $\widehat{f}_n$ is unbiased and $I\!E\widehat{\Theta} = \Theta$. We will use below this property to construct our test.

3 TEST STATISTIC

3.1 Convexity in $\mathcal{S}$(k,4)

First of all, we characterize convexity in the class $\mathcal{S}(k,4)$. Note that if a function g is a cubic spline, then its second derivative is a linear function between any pair of adjacent knots η_i and η_{i+1}, and it follows that g is a convex function in the interval $\eta_i \leq x \leq \eta_{i+1}$ if and only if $g''(\eta_i)$ and $g''(\eta_{i+1})$ are both non negative (this property was used by Dierckx (1980) to define a convex estimator).
For a function g in the class $\mathcal{S}(k,4)$, we can write:

$$g(x) = \sum_{p=1}^{k+4} \theta_p N_p(x) \quad \text{with} \quad \Theta = (\theta_1, ..., \theta_{k+4})' \in I\!R^{k+4}.$$

$$\text{Then:} \qquad g''(\eta_l) = \sum_{p=1}^{k+4} \theta_p N''_p(\eta_l) = \sum_{p=1}^{k+4} \theta_p d_{p,l},$$

where the coefficients $d_{p,l}$ are easily calculated from the knots (see Dierckx 1980)

$$\begin{cases} d_{p,l} = 0 \quad \text{if} \quad p \leq l \quad \text{or} \quad p \geq l+4 \\ d_{l+1,l} = \frac{6}{(t_{l+5}-t_{l+2})(t_{l+5}-t_{l+3})} \\ d_{l+3,l} = \frac{6}{(t_{l+6}-t_{l+3})(t_{l+5}-t_{l+3})} \\ d_{l+2,l} = -(d_{l+3,l} + d_{l+1,l}) \end{cases} \qquad \text{for} \quad l = 0, ..., k+1$$

Let $b_l = (0, 0, ..., 0, -d_{l+1,l}, -d_{l+2,l}, -d_{l+3,l}, 0, ..., 0)' \in I\!R^{k+4}$ and $\Theta = (\theta_1, ..., \theta_{k+4})'$, then

$$g''(\eta_l) = -b'_l\Theta.$$

Hence, we see that a cubic spline g is a convex function if and only if $b'_l\Theta \leq 0$ for all $l = 0, ..., k+1$.
The basic idea of our test goes as follows. Whenever f is a cubic spline function, then $I\!E\widehat{\Theta} = \Theta$ (we have already mentioned it in **section 2**). Therefore, a test for convexity can be written as $H_0' : b'_l\Theta \leq 0$ for all $l = 0, ..., k+1$ versus $H_1' : b'_l\Theta > 0$ for some $l \in \{0, ..., k+1\}$ and where Θ is the mean of the random vector $\widehat{\Theta}$.

On the other hand, Beatson (1982) shows that for a smooth and convex function $f \in \mathcal{C}^m(0,1)(0 \leq m \leq 3)$, the uniform distance between f and the set $\mathcal{S}_c(k,4)$ of convex functions of $\mathcal{S}(k,4)$ tends to zero when the mesh size δ_k tends to zero (see **lemma 8** in the **appendix**).

A testing problem in the form $[H_0', H_1']$ is related to the one-sided testing problem in multivariate analysis and has been studied by several authors (Bartholomew 1961, Kudô 1963, Nüesch 1966, Kudô and Choi 1975, Shapiro 1985 and more recently by Raubertas et al. 1986 and Robertson et al. 1988).

3.2 One-sided Test

Let Y be a random vector distributed as $\mathcal{N}_q(\Theta, \Sigma_q)$ ($q \in I\!N, \quad q > 0$) where Σ_q is a known nonsingular matrix.
We consider testing $H_0' : b'_l\Theta \leq 0 \quad (l = 0, ..., k+1)$ against $H_1' : b'_l\Theta > 0$ for some $l \in \{0, ..., k+1\}$.
In this paper we identify a hypothesis with the corresponding set of parameters. For example, we write $H_0' = \{\Theta \in I\!R^q : b'_l\Theta \leq 0\}$.
The likelihood function is $L = c_0 \exp\left(-\frac{1}{2} \parallel Y - \Theta \parallel^2_{\Sigma_q^{-1}}\right)$ where c_0 is a positive constant independent of Θ. Thus, the likelihood ratio for the problem $[H_0', H_1']$ is given by:

$$\lambda = \frac{\sup_{H_o'} L}{\sup_{H_o' \cup H_1'} L} = \exp\left(-\frac{1}{2} \inf_{x \in H_o'} \parallel Y - x \parallel^2_{\Sigma_q^{-1}}\right).$$

So, to determine the test statistic under the null hypothesis, we need to resolve the following nonlinear programming problem: $\inf_{x \in H_o'} \parallel Y - x \parallel^2_{\Sigma_q^{-1}}$.
It is worth noting that H_o' is a polyhedral cone and is thus closed and convex. Hence, for a given Y, this infinimum is attained at unique point denoted by $\Pi_{H_o'}(Y)$ and represents the squared distance from Y to H_o'.
Thus, the likelihood ratio test (LRT) rejects H_0' for large values of the test statistic

$$\bar{\chi}^2 = \inf_{x \in H_o'} \parallel Y - x \parallel^2_{\Sigma_q^{-1}} .$$

Shapiro (1985) showed, in a study of the distribution of a minimum discrepancy statistic, that if H_o' is a any convex cone and if $\Theta = 0$, then the distribution of $\bar{\chi}^2$ statistic, called chi-bar-squared statistics, is a mixture of chi-squared distributions. Raubertas et al.(1986) generalize the one-sided testing problem to allow hypotheses involving homogeneous linear inequality restrictions. This framework includes the hypotheses of monotonicity, nonnegativity, and convexity. Here we give an immediate consequence of theorem 3.1 of Shapiro (1985). For that purpose we shall use some geometrical properties of polyhedral cones.

3.2.1 Polyhedral cones

A polyhedral cone is a set of points that satisfies a finite set of homogeneous linear inequalities. Let
$\{a_1, ..., a_p\}$ be a set of vectors in $I\!R^q$ (with $p \leq q$) and let $A = (a_1|...|a_p)$ a q-by-p matrix. Then the polyhedral cone determined by A is

$$\mathcal{C}[A] = \{x \in I\!R^q \ : A'x \leq 0\}.$$

The polar cone $\mathcal{C}^\circ[A]$ of a cone $\mathcal{C}[A]$ is defined by

$$\mathcal{C}^\circ[A] = \{x \in I\!R^q \ : x'y \leq 0, \quad \forall y \in \mathcal{C}[A]\}.$$

It is easy to see that

$$C^{\circ}[A] = \{x \in I\!R^q \ : x = \sum_{i=1}^{p} \lambda_i a_i, \quad \lambda_i \geq 0, \quad i = 1, ..., p\}.$$

Note that $C^{\circ}[A]$ is also a polyhedral convex cone and $(C^{\circ}[A])^{\circ} = C[A]$.

We shall introduce some useful notations which are a restatement of notations of Raubertas et al. (1986).

Let J be a subset (possibly empty) of $\{1, ..., p\}$ and let $\overline{J}$ be its complement. A_J will be the matrix consisting of those columns of A indexed by the elements of J. The matrix $A_{\overline{J}}$ is defined analogously.

The faces Ψ_J of $C[A]$ are defined as follows:

$$\Psi_J = \{x \in I\!R^q \ : A'_J x = 0, A'_{\overline{J}} x \leq 0\}$$

Note that $\Psi_J = C[A]$ when J is empty.

For the definition and basic properties of faces and polar cones the reader is referred to Rockafeller (1970) or Stoer and Witzgall (1970). A column of A is called redundant (Sasabuchi, 1980) if its presence or absence makes no difference to the cone determined by A. In what follows, we will assume (without loss of generality) that A contains no redundant columns. In this case $\dim(\Psi_J) = q - \#J$ where $\#J$ is the cardinal of J. Recall that the $\dim(\Psi_J)$ is defined to be the dimension of the linear subspace spanned by Ψ_J.

To a face Ψ_J, we denote by P_J the symmetric idempotent matrix giving the orthogonal projection onto the space generated by Ψ_J. As above, for all x, $\Pi_{C[A]}$ represents the projection of x onto $C[A]$.

Now, we shall need the following result which in various forms has been used by several authors (Kudô 1963, Wynn 1975 and Shapiro 1985).

Lemma 1 *For all $x \in I\!R^q$: $\Pi_{C[A]}(x) \in \Psi_J$ if and only if $P_J(x) \in \Psi_J$ and $x - P_J(x) \in C^{\circ}[A]$. In this case $\Pi_{C[A]}(x) = P_J(x)$ and $x - P_J(x) = A_J(A'_J A_J)^{-1} A'_J x$*

3.2.2 The distribution of $\bar{\chi}^2$

Let us assume that $\bar{\chi}^2 = \inf_{x \in C[A]} ||X - x||^2$. Then, the following result is an immediate consequence of **theorem 3.1** of Shapiro (1985).

Theorem 2 *Let X be a random vector distributed as $N_q(0, I_q)$, then the random variable $\bar{\chi}^2$ is distributed as a mixture of chi-squared distribution, namely:*

$$P\left(\bar{\chi}^2 \geq s^2\right) = \omega_0 P\left(\chi_0^2 \geq s^2\right) + \sum_{q-p \leq j \leq q-1} \omega_j P\left(\chi_{q-j}^2 \geq s^2\right) \tag{4}$$

with $\omega_0 = P(X \in C[A]) = P(A'X \leq 0)$
and
$\omega_j = \sum_{q-\#J=j} P(P_J(X) \in \Psi_J) P\left(A_J(A'_J A_J)^{-1} A'_J x \in C^{\circ}[A]\right)$
Moreover, $\omega_0 + \sum_{q-p \leq j \leq q-1} \omega_j = 1.$

Now we come back to our test. We can write $H_o' = C[B]$ with $B(b_0|...|b_{k+1})$. Thus $p = k+2$ and $q = k+4$. It is easy to see that B contains no redundant columns. Recall that in this case

$$\bar{\chi}^2 = \inf_{x \in H_o'} \| Y - x \|^2_{\Sigma_q^{-1}} .$$

If $Y \sim \mathcal{N}_q(0, \Sigma_q)$, by a straightforward manipulation of results of **theorem 2**, we obtain

$$P(\bar{\chi}^2 \geq s^2) = \omega_0 P(\chi_0^2 \geq s^2) + \sum_{q-p \leq j \leq q-1} \omega_j P(\chi^2_{q-j} \geq s^2) \tag{5}$$

with $\omega_0 = P(X \in C[B]) = P(B'X \leq 0)$
and

$$\omega_j = \sum_{q-\#J=j} P\left[B'_{\bar{J}}Y - (B'_{\bar{J}}\Sigma_q B_J)(B'_J\Sigma_q B_J)^{-1}B'_JY \leq 0\right] P\left[(B'_J\Sigma_q B_J)^{-1}B'_JY \geq 0\right].$$

This result shows that the distribution of $\bar{\chi}^2$ when $\Theta = 0$, is a mixture of chi-squared distributions. So, to calculate the probabilities in the right-hand side of (5), the values of ω_j are needed. However, even for moderate $q(q > 3)$, good closed form expressions for these level probabilities have not found. Thus approximations are of interest. For this, one may use Monte Carlo method (see Diack 1998).
Note that the coefficients ω_j depend on the vector b_j matrices Σ_q. Hence, in what follows, we denoted $\bar{\chi}^2$ by $\bar{\chi}^2_{\Sigma_q}(p)$.
Questions concerning the determination of the distribution of $\bar{\chi}^2_{\Sigma_q}(p)$ for any point of null hypothesis are unresolved. However, Raubertas et al. (1986) generalize the result of Shapiro (1985) to obtain the distribution of $\bar{\chi}^2_{\Sigma_q}(p)$ for Θ in the lineality space of H_o' ($\{x \in I\!R^q \ : B'x = 0\}$ in this case). They show that the distribution of $\bar{\chi}^2_{\Sigma_q}(p)$ is the same for any Θ in the lineality space of H_o' and stochastically greatest among $\Theta \in H_o'$ when Θ is in the lineality space of H_o'. (See the second corollary to **theorem 3.6** on page 2822 of Raubertas et al. 1986).

Therefore (5) has the following consequence: the size-α likelihood ratio test with null hypothesis H_o' versus the alternative hypothesis H_1' is the test with reject the null hypothesis if

$$\bar{\chi}^2_{\Sigma_q}(p) \geq s^2_{\alpha,p}$$

where $s^2_{\alpha,p}$ is defined by

$$\sum_{q-p \leq j \leq q-1} \omega_j P(\chi^2_{q-j} \geq s^2_{\alpha,p}) = \alpha. \tag{6}$$

Hence $s^2_{\alpha,p}$ is a function of the weights ω_j.

It is easily seen that all these results are still valid if Y is asymptotically normally distributed.

Now, the following result gives a sufficient condition of convergence of the power of the test.

Theorem 3 *If* $\inf_{x \in H_o'} \| Y - x \|^2_{\Sigma_q^{-1}} / q \longrightarrow +\infty$ *then* $P\left(\bar{\chi}^2_{\Sigma_q}(p) \geq s^2_{\alpha,p}\right) \longrightarrow 1.$

Proof. Let T be a $q \times q$ nonsingular matrix such that $T\Sigma_q T' = I_q$, that is $\Sigma_q = T^{-1}(T^{-1})'$, and make the transformation

$$X = TY, \quad U = T\Theta.$$

Then X is a random vector distributed as $\mathcal{N}_q(U, I_q)$. Define the set of vectors $\{a_1, ..., a_p\}$ as

$$a'_j = b'_j T^{-1}, \quad (j = 1, ..., p).$$

We have $b'_j\Theta = a'_j U \quad (j = 1, ..., p)$, and hence the problem $[H_o^{'}, H_1^{'}]$ is transformed to the following problem $[H_o^{''}, H_1^{''}]$:
$H_o^{''} : b'_j U \leq 0 \quad (j = 1, ..., p)$ versus $H_1^{''} : a'_j U > 0$ for some $j \in \{1, ..., p\}$.
We can write

$$\bar{\chi}^2_{\Sigma_q}(p) = \min_{x \in H_o^{''}} \| X - x \|^2 = \| X - \Pi_{H_o^{''}}(X) \|^2 .$$

On the other hand,

$$\| U - \Pi_{H_o^{''}}(U) \| \leq \| U - \Pi_{H_o^{''}}(X) \| \leq \| X - U \| + \| X - \Pi_{H_o^{''}}(X) \| .$$

Hence, $P\left(\bar{\chi}^2_{\Sigma_q}(p) \geq s^2_{\alpha,p}\right) = P\left(\left\|X - \Pi_{H_o^{''}}(X)\right\| \geq s_{\alpha,p}\right)$
$\geq P\left(\|X - U\| \leq \left\|U - \Pi_{H_o^{''}}(U)\right\| - s_{\alpha,p}\right)$.

Hence, the result follows from Bienayme-Chebychev inequality applied to the random variable $\| X - U \|$ and the assumption $\left\|U - \Pi_{H_o^{''}}(U)\right\|^2 / q \longrightarrow +\infty$. ∎

The test statistic requires computing the projection $\Pi_{H_o^{'}}(Y)$ of Y. However, a good closed-form solution has not found. Hence, this problem requires extensive numerical work to obtain solution. We propose an algorithm based on successive projections which has been introduced by Dykstra (1983) (see also Boyle and Dykstra 1985). This algorithm determines the projection of a point X of any real Hilbert space onto the intersection $\mathcal{K}$ of convex sets $\mathcal{K}_j \quad (j = 1, ..., p)$ and it is meant for applications where projections onto the $\mathcal{K}_j$'s can be calculated relatively easily. Let $\mathcal{K}$ be a closed convex cone in $\mathbb{R}^q$. We suppose that $\mathcal{K}$ can be written as $\bigcap_{j=1}^{p} \mathcal{K}_j$ and each $\mathcal{K}_j$ is also convex cone. For all $X \in \mathbb{R}^q$, we denote by $X^\Gamma_{\mathcal{K}}$ the $\Gamma-$ projection onto $\mathcal{K}$, where Γ is a positive definite matrix. The algorithm consists of repeated cycles and every cycle contains p stages.
Let X^Γ_{mi} be the appproximation of $X^\Gamma_{\mathcal{K}}$ given by Dykstra's algorithm at the ith stage of mth cycle.
The following result (see Boyle and Dykstra 1985) proves that the algorithm converges correctly.

Theorem 4 *For any* $(1 \leq i \leq p)$, *the sequence* $\{X^\Gamma_{mi}\}$ *converges to* $X^\Gamma_{\mathcal{K}}$ *in the following sense:* $\|X^\Gamma_{mi} - X^\Gamma_{\mathcal{K}}\|_\Gamma \longrightarrow 0$ *as* $m \longrightarrow +\infty$.

<u>**Application**</u>: Let $\mathcal{K} = H_o^{'}$ be the null hypothesis of the problem $[H_o^{'}, H_1^{'}]$ and let $\mathcal{K}_i = \{x \in \mathbb{R}^q \; : b_i^{'} x \leq 0\}$. Let $\Gamma = \Sigma_q^{-1}$ be the covariance matrix of Y. For all $m \in \mathbb{N}, \quad m > 0$, we defined $\bar{\chi}^2_{\Sigma_q}(p, m)$ by

$$\bar{\chi}^2_{\Sigma_q}(p, m) = \| Y - Y_{mp}^{\Sigma_q^{-1}} \|^2_{\Sigma_q^{-1}}$$

where $Y_{mp}^{\Sigma_q^{-1}}$ is given by the pth stage of mth cycle of the Dykstra's algorithm.
We have then the following equality:

$$\bar{\chi}^2_{\Sigma_q}(p) = \| Y - Y_{H_o^{'}}^{\Sigma_q^{-1}} \|^2_{\Sigma_q^{-1}} = \| Y - Y_{mp}^{\Sigma_q^{-1}} \|^2_{\Sigma_q^{-1}} + \| Y_{mp}^{\Sigma_q^{-1}} - Y_{H_o^{'}}^{\Sigma_q^{-1}} \|^2_{\Sigma_q^{-1}}$$

$$+2 < Y - Y_{mp}^{\Sigma_q^{-1}}, Y_{mp}^{\Sigma_q^{-1}} - Y_{H_o^{'}}^{\Sigma_q^{-1}} >_{\Sigma_q^{-1}}$$

where $<,>_{\Sigma_q^{-1}}$ is the inner product on $I\!R^q$ defined by Σ_q^{-1}. Using now theorem 4, we see that

$$|| Y_{mp}^{\Sigma_q^{-1}} - Y_{H_o'}^{\Sigma_q^{-1}} ||^2_{\Sigma_q^{-1}} \rightarrow 0 \text{ a.s. as } \quad m \rightarrow +\infty.$$

In the same way, it can be shown that for fixed p and q

$$< Y - Y_{mp}^{\Sigma_q^{-1}}, Y_{mp}^{\Sigma_q^{-1}} - Y_{H_o'}^{\Sigma_q^{-1}} >_{\Sigma_q^{-1}} \rightarrow 0 \quad \text{a.s. as} \quad m \rightarrow +\infty.$$

Hence, $\bar{\chi}^2_{\Sigma_q}(p,m)$ converges almost surely to $\bar{\chi}^2_{\Sigma_q}(p)$ as m tends to infinity. Therefore, to implement the test, we will use $\bar{\chi}^2_{\Sigma_q}(p,m)$ instead of $\bar{\chi}^2_{\Sigma_q}(p)$.
We can now define our convexity test.

3.3 Definition of the test

Consider the problem $[H_o, H_1]$ where H_o means that the regression function f is convex and H_1 is the unrestricted alternative.
Let $\widehat{\Theta}$ be the solution of the quadratic programming problem (3). The number of knots will be a function of the sample size $k = k_n$. To define $\widehat{\Theta}_{m,k_n+2}$ by

$$\widehat{\Theta}_{m,k_n+2} = \widehat{\Theta}^{\Sigma_n}_{m,k_n+2}$$

with $\Sigma_n^{-1} = \frac{\sigma^2}{n} M^{-1}(\mu^n)$ and where $\widehat{\Theta}^{\Sigma_n}_{m,k_n+2}$ given by the $(k_n+2)^{th}$ stage of the m^{th} cycle of Dykstra's algorithm.
Like this, we will define our test of convexity by rejecting H_o when

$$\bar{\chi}^2_{\frac{\sigma^2}{n}M^{-1}(\mu^n)}(k_n+2,m) = \frac{n}{\sigma^2} || \widehat{\Theta} - \widehat{\Theta}_{m,k_n+2} ||^2_{M(\mu^n)} \geq s^2_{\alpha,k_n+2}, \qquad (7)$$

where s^2_{α,k_n+2} is defined by (6).

4 ASYMPTOTIC PROPERTIES

Note that the test procedure requires the knowledge of the variance σ^2. However, in practice, σ is unknown and we need a consistent estimate of it. This can be obtained in the case of the least squares estimator, using $\hat{\sigma}_n^2 = \frac{1}{n-(k+4)} \sum_{i=1}^r (\bar{y}_i - \hat{f}_n(x_i))^2$ or alternatively, any consistent estimator based on nonparametric regression techniques.
In what follows, we assume that μ^n converges to a design measure μ, where μ is an absolutely continuous measure. We denote by G_n and G the cumulative distribution function of μ^n and μ respectively. The critical region of the test is

$$\Lambda_{n,m} = \{\bar{\chi}^2_{\frac{\hat{\sigma}_n^2}{n}M^{-1}(\mu^n)}(k_n+2,m) = \frac{n}{\hat{\sigma}_n^2} || \widehat{\Theta} - \widehat{\Theta}_{m,k_n+1} ||^2_{\Sigma_n} \geq s^2_{\alpha,k_n+2}\}.$$

Now, we are ready for the main result of this section.

Theorem 5 *Let $f \in \mathcal{F}_{l,M}$ with $l \geq 3$. Let us consider the problem $[H_o,H_1]$. Then, under the following assumptions:*
(i) ε_{ij} $(i = 1,...,r\ ;\ j = 1,...n_i)$ i.i.d. with mean zero and finite variance σ^2.
(ii) $n_i \rightarrow +\infty$ and $\hat{\sigma}_n^2 \rightarrow \sigma^2$ (with convergence in probability)
(iii) $\sup_{0\leq x\leq 1} |G_n(x) - G(x)| = o(\delta_{k_n})$, as $k_n \longrightarrow +\infty$.
(iv) $\lim_{n\longrightarrow+\infty} rn\delta^8_{k_n} \left(\sup_{1\leq i\leq r} \mu_i\right) = 0$

The test is asymptotically size α*. More precisely,*

$$\limsup_{n\longrightarrow+\infty}\sup_{f\in H_o}\lim_{m\longrightarrow+\infty}P_f\left(\Lambda_{n,m}\right)=\alpha.$$

Moreover, if

$$(v)\ \lim_{n\longrightarrow+\infty}h_n^2 n\delta_{k_n}^6=+\infty$$

then, the test is consistent for all $f_n \in H_{1n}$ *i.e.:*

$$\lim_{n\longrightarrow+\infty}\lim_{m\longrightarrow+\infty}P_{f_n}\left(\Lambda_{n,m}\right)=1.$$

The proof of **theorem 5** is given in Appendix.

Remark 6 ***Theorem 5*** *gives the behavior of the test under the hypotheses* H_{1n} : $f_n = f + h_n L$*. The fixed alternative corresponds to* $h_n = 1$*. This formulation allows us to deal with some local alternatives. From* ***theorem 5,*** *the test statistic diverges to* $+\infty$ *under any fixed alternative to* H_o*. Moreover, the test has power to detect local alternatives of the type* H_{1n} *approaching the null at rate slower than* $n^{-1/2}\delta_{k_n}^{-3}$*.*

In the case of gaussian errors, assumption $n_i \longrightarrow +\infty$ *is not needed.*

Assumption (iii) is the same as in Aggarwal and Studden (1980)

It is easy to see that assumption (iv) implies that $\lim_{n\longrightarrow+\infty} k_n = +\infty$ *.*

For a uniform design, i.e. $\mu_i = 1/r, i = 1, \ldots, r$ *then (iv) is equivalent to* $\lim_{n\longrightarrow+\infty} n\delta_{k_n}^8 = 0$.

Discussion: We have proposed a consistent test of convexity of a regression function in a nonparametric model. While it appears difficult to impose properties such as concavity on nonparametric local averaging estimators, this restriction is readily introduced by using a cubic spline estimator. Hence, the idea of the test exploits the close connection between the convexity problem and the hypothesis testing problems concerning linear inequalites and normal means. The test is shown to be consistent against local alternatives approaching the null at rates slower than $n^{-1/2}\delta_{k_n}^{-3}$. It is reasonable to think that the test is more powerful than Yatchew's test . Indeed, in general the local convergence rate of h_n in which those tests have power to detect local alternatives is the square root of the rate at which the test statistics converge to infinity. Yatchew's test converges to infinity at rate $n^{1/2}$ (see Diack 1998). It has then a power to detect local alternatives approaching to the null at rate slower than $n^{-1/4}$. Therefore our test is more powerful in detecting local misspecification. In Diack (1998) one can see that Schlee's test has local properties similar to ours (see **theorem 3** Diack 1998).

The test is also easy to compute. A simulation study in Diack (1998) shows that the test has adequate size and its behavior for small sample is better than Yatchew's and Schlee's. However, further extensive study on the choice of the number of knots is necessary.

A test of monotonicity can be readily constructed paralleling the above convexity test with quadratic splines instead of cubic splines. This additional step is still under study.

Appendix For the proof of the **theorem 5**, we need to use some preliminary lemmas. **Lemma 7** which is obtained by a straightforward manipulation of results of Schumaker (1981), gives a sup norm error bound when approximating a smooth function with a cubic spline.

Lemma 7 *There is a constant c such that for all* $f \in \mathcal{F}_{l,M}$*, there is a function S in* $\mathcal{S}(k, 4)$ *such that:*

$$\sup_{0\leq x\leq 1}\left|f\left(x\right)-S\left(x\right)\right|\leq c\delta_{k_n}^4.$$

Lemma 8 is a consequence of results of Betason (1982). It gives a similar sup norm error bound when approximating a convex and smooth function with a convex cubic spline.

Lemma 8 *There is a constant c such that if $f \in \mathcal{F}_{l,M}$ is convex, then there exists a convex functions S in $\mathcal{S}(k,4)$ such that:*
$\sup_{0\leq x\leq 1} |f(x) - S(x)| \leq c\delta_{k_n}^4.$

Lemma 9 has the following consequence: if $\hat{\sigma}_n^2$ converges in probability to σ^2 then $\frac{n}{\hat{\sigma}_n^2} \| \widehat{\Theta} - \widehat{\Theta}_{m,k_n+1} \|_{\Sigma_n}^2$ has the same asymptotic distribution than $\frac{n}{\sigma^2} \| \widehat{\Theta} - \widehat{\Theta}_{m,k_n+2} \|_{M(\mu^n)}^2$. Reader is referred to Lehman (1986) for a proof of this lemma.

Lemma 9 *Let X_n be a random variable converging in distribution. Let a_n and b_n be random variables converging respectively in probability to a and b. Then, the random variable $a_nX_n + b_n$ converges in the same distribution as $aX_n + b$.*

Proof.

$$\widehat{\Theta} = M^{-1}(\mu^n)F\mathcal{D}(\mu^n)\bar{Y}.$$

From **Lemma 2**, there is a function S in $\mathcal{S}(k_n, 4)$, such that :

$$\sup_{0\leq x\leq 1} | f(x) - S(x) | \leq c\delta_{k_n}^4.$$

$S \in \mathcal{S}(k_n,4)$ hence, there exists $\Theta \in I\!R^{k_n+4}$ such that $S(x) = N'(x)\Theta$.
Let

$$S_\Delta = (S(x_1),..,S(x_r))' = (N'(x_1)\Theta,..,N'(x_r)\Theta)' = F'\Theta.$$

Then,

$$M^{-1}(\mu^n)F\mathcal{D}(\mu^n)S_\Delta = M^{-1}(\mu^n)F\mathcal{D}(\mu^n)F'\Theta = \Theta.$$

Let

$$_s\widehat{\Theta} = M^{-1}(\mu^n)F\mathcal{D}(\mu^n)(S_\Delta + \bar{\varepsilon}).$$

Then

$$I\!E_s\widehat{\Theta} = \Theta \qquad \text{and} \qquad {}_s\widehat{\Theta} \rightsquigarrow \mathcal{N}(\Theta, \frac{\sigma^2}{n}M^{-1}(\mu^n)).$$

(with $\rightsquigarrow$ meaning convergence in distribution).
Therefore, we can write

$$\widehat{\Theta} = {}_s\widehat{\Theta} + B_n \qquad \text{and} \qquad I\!E\widehat{\Theta} = \Theta + B_n$$

with $B_n = M^{-1}(\mu^n)F\mathcal{D}(\mu^n)(f_\Delta - S_\Delta)$.
Now, let us recall that the test statistic is given by

$$\bar{X}^2_{\frac{\sigma^2}{n}M^{-1}(\mu^n)}(k_n+2,m) = \frac{n}{\sigma^2} \| \widehat{\Theta} - \widehat{\Theta}_{m,k_n+2} \|_{M(\mu^n)}^2.$$

But, for m sufficiently large and for fixed n, (see **section 3.2.3**) we have:
$\bar{X}^2_{\frac{\sigma^2}{n}M^{-1}(\mu^n)}(k_n+2,m)$ converges in probability to

$$\bar{X}^2_{\frac{\sigma^2}{n}M^{-1}(\mu^n)}(k_n+2) = \frac{n}{\sigma^2} \| \widehat{\Theta} - P(\widehat{\Theta}) \|_{M(\mu^n)}^2$$

where $P(\widehat{\Theta})$is the $M(\mu^n)-$projection of $\widehat{\Theta}$ onto the polyhedral cone

$$\mathcal{K}=\{x\in \mathbb{R}^{k_n+4} \ : b'_l x \leq 0, \quad l=0,...,k_n+1\}.$$

Besides,

$$\bar{\chi}^2_{\frac{\sigma^2}{n}M^{-1}(\mu^n)}(k_n+2)=\frac{n}{\sigma^2}\parallel(\widehat{\Theta}-{}_s\widehat{\Theta})+({}_s\widehat{\Theta}-P({}_s\widehat{\Theta}))-P(\widehat{\Theta})+P({}_s\widehat{\Theta})\parallel^2_{M(\mu^n)}.$$

We can rewrite this in the following form:

$$\bar{\chi}^2_{\frac{\sigma^2}{n}M^{-1}(\mu^n)}(k_n+2)=\frac{n}{\sigma^2}\parallel {}_s\widehat{\Theta}-P({}_s\widehat{\Theta})\parallel^2_{M(\mu^n)}$$

$$+\frac{n}{\sigma^2}\parallel B_n\parallel^2_{M(\mu^n)}+\frac{n}{\sigma^2}\parallel P({}_s\widehat{\Theta})-P({}_s\widehat{\Theta}+B_n)\parallel^2_{M(\mu^n)}$$

$$+\frac{2n}{\sigma^2}< {}_s\widehat{\Theta}-P({}_s\widehat{\Theta}),B_n>_{M(\mu^n)}+\frac{2n}{\sigma^2}< {}_s\widehat{\Theta}-P({}_s\widehat{\Theta}),P({}_s\widehat{\Theta})-P({}_s\widehat{\Theta}+B_n)>_{M(\mu^n)}$$

$$+\frac{2n}{\sigma^2}< B_n,P({}_s\widehat{\Theta})-P({}_s\widehat{\Theta}+B_n)>_{M(\mu^n)}$$

where $<,>_{M(\mu^n)}$ is the scalar product defined by the metric $M(\mu^n)$.
It is easily seen, as in Diack&Thomas (1998) (see **formula 3.5**) that

$$\sup_{f\in\mathcal{F}_{l,M}}\sqrt{\frac{n}{\sigma^2}\parallel B_n\parallel^2\parallel M(\mu^n)\parallel}=\mathcal{O}\left(r^{1/2}n^{1/2}\delta_{k_n}{}^4\left(\sup_{1\leq i\leq r}\mu_i\right)^{1/2}\right)$$

It follows that

$$\limsup_{n\longrightarrow+\infty}\sup_{f\in\mathcal{F}_{l,M}}\sqrt{\frac{n}{\sigma^2}\parallel B_n\parallel^2_{M(\mu^n)}}=0.$$

Using now the fact that projections onto closed convex cones are contracting maps as are projections onto linear subspaces, we obtain

$$\frac{n}{\sigma^2}\parallel P({}_s\widehat{\Theta})-P({}_s\widehat{\Theta}+B_n)\parallel^2_{M(\mu^n)}\leq\frac{n}{\sigma^2}\parallel B_n\parallel^2_{M(\mu^n)}.$$

Then

$$\lim\sup_{n\longrightarrow+\infty}\sup_{f\in\mathcal{F}_{l,M}}\frac{n}{\sigma^2}\parallel P({}_s\widehat{\Theta})-P({}_s\widehat{\Theta}+B_n)\parallel^2_{M(\mu^n)}=0\quad a.s.$$

It follows that

$$\lim\sup_{n\longrightarrow+\infty}\sup_{f\in\mathcal{F}_{l,M}}\frac{2n}{\sigma^2}< B_n,P({}_s\widehat{\Theta})-P({}_s\widehat{\Theta}+B_n)>_{M(\mu^n)}=0\quad a.s.$$

On the other hand

$$\mid\frac{2n}{\sigma^2}< {}_s\widehat{\Theta}-P({}_s\widehat{\Theta}),B_n>_{M(\mu^n)}\mid\leq\frac{2n}{\sigma^2}\parallel {}_s\widehat{\Theta}-P({}_s\widehat{\Theta})\parallel_{M(\mu^n)}\parallel B_n\parallel_{M(\mu^n)}$$

And also

$$\frac{2n}{\sigma^2} \quad < \quad {}_s\widehat{\Theta} - P({}_s\widehat{\Theta}), P({}_s\widehat{\Theta}) - P({}_s\widehat{\Theta} + B_n) >_{M(\mu^n)}$$
$$\leq \frac{2n}{\sigma^2} \quad | \quad | {}_s\widehat{\Theta} - P({}_s\widehat{\Theta}) ||_{M(\mu^n)} || B_n ||_{M(\mu^n)}$$

Then, there is a sequence a_n of reals converging to zero such that

$$\mathcal{P}_f \left(\bar{\chi}^2_{\frac{\sigma^2}{n} M^{-1}(\mu^n)}(k_n + 2) \geq s^2_{\alpha, k_n+2} \right) \leq \mathcal{P}_f \left((1 + a_n) \frac{n}{\sigma^2} || {}_s\widehat{\Theta} - P({}_s\widehat{\Theta}) ||^2_{M(\mu^n)} \geq s^2_{\alpha, k_n+2} \right)$$

Now, under the null hypothesis, S is convex. Thus $\Theta \in \mathcal{K}$. Then, since the distribution of $\bar{\chi}^2_{\frac{\sigma^2}{n} M^{-1}(\mu^n)}(k_n + 2)$ is stochastically greatest among $\Theta \in \mathcal{K}$ when $\Theta = 0$, we see that

$$|| {}_s\widehat{\Theta} - P({}_s\widehat{\Theta}) ||_{M(\mu^n)} \leq || {}_s\widehat{\Theta} - \Theta - P({}_s\widehat{\Theta} - \Theta) ||_{M(\mu^n)} .$$

Moreover,

$${}_s\widehat{\Theta} - \Theta \rightsquigarrow \mathcal{N}(0, \frac{\sigma^2}{n} M^{-1}(\mu^n))$$

Hence,

$$P(\frac{n}{\sigma^2} || {}_s\widehat{\Theta} - \Theta - P({}_s\widehat{\Theta} - \Theta) ||^2_{M(\mu^n)} \geq s^2) = \sum_{2 \leq j \leq k_n+3} \omega_j P\left(\chi^2_{q-j} \geq s^2\right)$$

Therefore, we can write that

$$\limsup_{n \to +\infty} \lim_{m \to +\infty} \sup_{f \in H_0} \mathcal{P}_f(\Lambda_{n,m}) = \limsup_{n \to +\infty} \sup_{f \in H_0} \mathcal{P}_f \left(\bar{\chi}^2_{\frac{\sigma^2}{n} M^{-1}(\mu^n)}(k_n + 2) \geq s^2_{\alpha, k_n+2} \right)$$

$$= \limsup_{n \to +\infty} \sup_{f \in H_0} \mathcal{P}_f \left(\frac{n}{\sigma^2} || {}_s\widehat{\Theta} - P({}_s\widehat{\Theta}) ||^2_{M(\mu^n)} \geq s^2_{\alpha, k_n+2} \right)$$

$$\leq \limsup_{n \to +\infty} \sup_{f \in H_0} \mathcal{P}_f \left(\frac{n}{\sigma^2} || {}_s\widehat{\Theta} - \Theta - P({}_s\widehat{\Theta} - \Theta) ||^2_{M(\mu^n)} \geq s^2_{\alpha, k_n+2} \right)$$

$$= \limsup_{n \to +\infty} \lim_{m \to +\infty} \mathcal{P}_0(\Lambda_{n,m}) = \alpha.$$

Then the test is asymptotically of size α. It remains to be proven that the test is consistent against local alternatives. From **theorem 3**, it suffices to show that

$$\frac{n}{\sigma^2} || \mathbb{E}\widehat{\Theta} - P(\mathbb{E}\widehat{\Theta}) ||^2_{M(\mu^n)} \delta_{k_n} \longrightarrow +\infty.$$

As above, we have

$$\frac{n}{\sigma^2} || \mathbb{E}\widehat{\Theta} - P(\mathbb{E}\widehat{\Theta}) ||^2_{M(\mu^n)} = \frac{n}{\sigma^2} || B_n + \Theta - P(B_n + \Theta) ||^2_{M(\mu^n)} .$$

In other words

$$\frac{n}{\sigma^2}\left\|\mathbb{E}\widehat{\Theta}-P(\mathbb{E}\widehat{\Theta})\right\|^2_{M(\mu^n)}=\frac{n}{\sigma^2}\|\Theta-P(\Theta)\|^2_{M(\mu^n)}+\frac{n}{\sigma^2}\|B_n\|^2_{M(\mu^n)}$$

$$+\frac{n}{\sigma^2}\|P(\Theta)-P(\Theta+B_n)\|^2_{M(\mu^n)}+\frac{2n}{\sigma^2}<\Theta-P(\Theta),B_n>_{M(\mu^n)}$$

$$+\frac{2n}{\sigma^2}<\Theta-P(\Theta)+B_n,P(\Theta)-P(\Theta+B_n)>_{M(\mu^n)}$$

In the same way as above, we see that

$$\frac{n}{\sigma^2}\parallel B_n\parallel^2_{M(\mu^n)}\to 0.$$

$$\frac{n}{\sigma^2}\parallel P(\Theta)-P(\Theta+B_n)\parallel^2_{M(\mu^n)}\to 0.$$

Hence

$$\frac{2n}{\sigma^2}\,|<\Theta-P(\Theta),B_n>_{M(\mu^n)}|=\epsilon_n\left(\frac{n}{\sigma^2}\parallel\Theta-P(\Theta)\parallel^2_{M(\mu^n)}\right)$$

and

$$\frac{2n}{\sigma^2}\,|<\Theta-P(\Theta)+B_n,P(\Theta)-P(\Theta+B_n)>_{M(\mu^n)}|$$

$$=\epsilon_n^{'}\left(\frac{n}{\sigma^2}\parallel\Theta-P(\Theta)\parallel^2_{M(\mu^n)}\right)$$

where ϵ_n and $\epsilon_n^{'}$ are nonnegative reals converging to zero. Therefore, the consistency of the test will be established when we show that

$$\frac{n}{\sigma^2}\parallel\Theta-P(\Theta)\parallel^2_{M(\mu^n)}\delta_{k_n}\longrightarrow+\infty.$$

If $f_n\in H_{1n}$ then S is also non-convex and therefore, $\Theta\notin\mathcal{K}$. Hence, there is a face Ψ_J of $\mathcal{K}$ such that $P(\Theta)$ lies in Ψ_J with Ψ_J defined by

$$\Psi_J=\{x\in\mathbb{R}^{k_n+4}\ :B'_Jx=0,\quad B'_{\bar{J}}x\leq 0\}.$$

Therefore, from **lemma 1**, $P(\Theta)$ is also the orthogonal projection of Θ onto the subspace generated by Ψ_J and $\Theta-P(\Theta)$ is also in the polar cone of $\mathcal{K}$. That is to say

$$\frac{n}{\sigma^2}\parallel\Theta-P(\Theta)\parallel^2_{M(\mu^n)}=\frac{n}{\sigma^2}(B'_J\Theta)'\,(B'_JM^{-1}(\mu^n)B_J)^{-1}\,(B'_J\Theta)$$

And

$$\left\{\begin{array}{c}(B'_JM^{-1}(\mu^n)B_J)^{-1}B'_J\Theta\geq 0.\\ B'_{\bar{J}}\Theta-\left(B'_{\bar{J}}M^{-1}(\mu^n)B_J\right)(B'_JM^{-1}(\mu^n)B_J)^{-1}B'_J\Theta\leq 0\end{array}\right. \tag{8}$$

Besides,

$$(B'_J\Theta)'\,(B'_JM^{-1}(\mu^n)B_J)^{-1}\,(B'_J\Theta)\geq\frac{\|B'_J\Theta\|^2}{\|B'_JM^{-1}(\mu^n)B_J\|}\geq\frac{\|B'_J\Theta\|^2}{\|B_J\|^2\,\|M^{-1}(\mu^n)\|}.$$

Since now, the sequence of knots is quasi-uniform, it is easy to see that

$$\sup_{0\leq j\leq k_n+1} \| b_j \| = \mathcal{O}\big(\inf_{0\leq j\leq k_n+1} \| b_j \|\big) = \mathcal{O}\left(\delta_{k_n}^{-2}\right).$$

On the other hand, $B'_J B_J$ is a band matrix. Thus, $\|B_J\|^2 = \mathcal{O}\left(\delta_{k_n}^{-4}\right)$.

Hence, we see that (using the fact that $\| M(\mu^n) \| = \mathcal{O}(\delta_{k_n})$ and $\| M^{-1}(\mu^n) \| = \mathcal{O}\left(\delta_{k_n}^{-1}\right)$ see Diack& Thomas 1998) for n sufficiently large,

$$\frac{n}{\sigma^2} \| \Theta - P_2(\Theta) \|^2_{M(\mu^n)} \geq \frac{n}{\sigma^2} \left(\sum_{j\in J} \left(f''_n(\eta_j)\right)^2 \right) \delta_{k_n}^5 .$$

Since $f''_n = f'' + h_n L''$ we will be finished when we show that: for n sufficiently large $\sum_{j\in J} \left(f''_n(\eta_j)\right)^2 > \epsilon$ where ϵ is a positive real.

First, it is clear that $\sum_{j\in J} \left(f''_n(\eta_j)\right)^2 \geq \sum_{j\in J} \left(f''_n(\eta_j)\right)^2 1_{[f''_n(\eta_j)<0]}$.

On the other hand, from (8) we have

$$B'_{\bar{J}}\Theta \leq \left(B'_{\bar{J}} M^{-1}(\mu^n) B_J\right) \left(B'_J M^{-1}(\mu^n) B_J\right)^{-1} B'_J \Theta.$$

Then for n sufficiently large,

$$\left\| \left(B'_{\bar{J}} M^{-1}(\mu^n) B_J\right) \left(B'_J M^{-1}(\mu^n) B_J\right)^{-1} B'_J \Theta \right\|^2 \geq \sum_{j\in \bar{J}} \left(f''_n(\eta_j)\right)^2 1_{[f''_n(\eta_j)<0]}.$$

Therefore,

$$\left\| \left(B'_{\bar{J}} M^{-1}(\mu^n) B_J\right) \right\|^2 \left\| \left(B'_J M^{-1}(\mu^n) B_J\right)^{-1} \right\|^2 \sum_{j\in J} \left(f''_n(\eta_j)\right)^2 \geq \sum_{j\in \bar{J}} \left(f''_n(\eta_j)\right)^2 1_{[f''_n(\eta_j)<0]}.$$

We have $\left\| \left(B'_{\bar{J}} M^{-1}(\mu^n) B_J\right) \right\|^2 \leq \|B\|^4 \left\|M^{-1}(\mu^n)\right\|^2$.

Moreover

$$\begin{aligned} \frac{1}{\left\|\left(B'_J M^{-1}(\mu^n) B_J\right)^{-1}\right\|} &= \min_x \frac{(B_J x)' M^{-1}(\mu^n) (B_J x)}{\|x\|^2} \leq \min_y \frac{y' M^{-1}(\mu^n) y}{\|y\|^2} \|B_J\|^2 \\ &\leq \frac{\|B\|^2}{\|M(\mu^n)\|}. \end{aligned}$$

Then

$$\left\| \left(B'_{\bar{J}} M^{-1}(\mu^n) B_J\right) \right\|^2 \left\| \left(B'_J M^{-1}(\mu^n) B_J\right)^{-1} \right\|^2 = \mathcal{O}(1)$$

Hence,

$$\sum_{j\in J} \left(f''_n(\eta_j)\right)^2 \geq \mathcal{O}(1) \sum_{j\in \bar{J}} \left(f''_n(\eta_j)\right)^2 1_{[f''_n(\eta_j)<0]}.$$

Because the knots are dense in $(0,1)$ one can deduce that

$$\sum_{j\in J} \left(f''_n(\eta_j)\right)^2 \geq c \sup_{x\in(0,1)} \left[(f''(x))^2 1_{[(f''(x)\leq 0)]}\right] > \epsilon$$

where c et ϵ are positive constants. We have

$$\frac{n}{\sigma^2} \| \Theta - P(\Theta) \|^2_{M(\mu^n)} \delta_{k_n} \to +\infty.$$

The consistency of the test for local alternatives follows. ■

References

[1] G.G. Agarwal and W.J Studden. Asymptotic integrated mean square error using least squares and bias minimizing splines. *The Annals of Statistics*, 8(6):1307–1325, 1980.

[2] R. E. Barlow, D.J. Bartholomew, J.M. Bremner, and H.D. Brunk. *Statistical inference under order restrictions.* John Wiley α Sons Ltd., 1972.

[3] D.J. Bartholomew. Ordered tests in the analysis of variance. *Biometrika*, 48:325–332, 1961.

[4] R.K. Beatson. Monotone and convex approximation by splines: error estimates and a curve fitting algorithm. *SIAM J. of Math. Analysis*, 19(4), 1982.

[5] J.P. Boyle and R.L. Dykstra. A method for finding projections onto the intersection of convex sets in hilbert spaces. *Lectures notes in mathematics*, 37:28–47, 1985.

[6] C.D. Carroll and M.S. Kimball. Notes and comments on the concavity of the comsumption function. *Econometrica*, 64(4):981–992, 1996.

[7] C.A.T. Diack. *Testing convexity.* Proceedings in Computational Statistics: Physica Verlag, 1996.

[8] C.A.T. Diack. Sur la convergence des tests de schlee et yatchew. *submitted*, 1998.

[9] C.A.T. Diack and C. Thomas-Agnan. A nonparametric test of the non-convexity of regression. *Journal of Nonparametric Statistics*, 49(4):335–362, 1998.

[10] H. Dierckx. An algorithm for cubic spline fitting with convexity constraints. *Computing*, 24:349–371, 1980.

[11] R.L. Dykstra. An algorithm for restricted least squares regression. *Journal of american statistical association*, 78(384), December 1983.

[12] A. Kudo. A multivariate analogue of the one-sided test. *Biometrika*, 50(3 and 4):403–418, 1963.

[13] A. Kudo and J.R. Choi. A generalized multivariate analogue of the one-sided test. *Mem. Fac. Sci., Kyushu Univ. A.*, 29:303–328, 1975.

[14] E.L. Lehmann. *Testing statistical hypotheses, second edition.* John Wiley α Sons, 1986.

[15] P.E. Nuesch. On the problem of testing location in multivariate populations for restricted alternatives. *Ann. Math. Statist.*, 37:113–119, 1966.

[16] R.F. Raubertas, C.I.C. Lee, and E.V. Nordheim. Hypothesis tests for normal means constrained by linear inequalities. *Comm. Statist. Theor. Meth.*, 15:2809–2833, 1986.

[17] T. Robertson, F.T. Wright, and R.L. Dykstra. *Order restricted statistical inference.* Wiley New-York, 1988.

[18] R.T.. Rockafeller. *Convex Analysis.* Princeton, New Jersey: Princeton University Press, 1970.

[19] S. Sasabuchi. A test of a multivariate normal mean with composite hypotheses determined by linear inequalities. *Biometrika*, 67(2):429–439, 1980.

[20] W. Schlee. Nonparametric test of the monotony and convexity of regression. *Nonparametric Statistical Inference*, 2:823–836, 1980.

[21] L. Schumaker. *Spline function: Basic theory.* John Wiley, New York, 1981.

[22] A. Shapiro. Asymptotic distribution of test statistics in the analysis of moment structures under inequality constraints. *Biometrika*, 72(1):133–144, 1985.

[23] H.P. Wynn. Integrals for one-sided confidence bounds: a general result. *Biometrika*, 62:393–396, 1975.

[24] A.J. Yatchew. Nonparametric regression tests based on least squares. *Econometrics Theory*, 8:435–451, 1992.