

# Kernel density estimation and goodness-of-fit test in adaptive tracking

Bernard Bercu, Bruno Portier

► **To cite this version:**

Bernard Bercu, Bruno Portier. Kernel density estimation and goodness-of-fit test in adaptive tracking. SIAM Journal on Control and Optimization, Society for Industrial and Applied Mathematics, 2008, 47, pp.2440-2457. <hal-00012682>

**HAL Id: hal-00012682**

**<https://hal.archives-ouvertes.fr/hal-00012682>**

Submitted on 26 Oct 2005

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# KERNEL DENSITY ESTIMATION AND GOODNESS-OF-FIT TEST IN ADAPTIVE TRACKING

BERNARD BERCU\* AND BRUNO PORTIER†

**Abstract.** We investigate the asymptotic properties of a recursive kernel density estimator associated with the driven noise of a linear regression in adaptive tracking. We provide an almost sure pointwise and uniform strong law of large numbers as well as a pointwise and multivariate central limit theorem. We also propose a goodness-of-fit test together with some simulation experiments.

**Key words.** Adaptive control, Kernel density estimation, Goodness-of-fit test

**AMS subject classifications.** 93C40, 62G07, 62G10

**1. Introduction.** Since the pioneer work of Aström and Wittenmark [1], a wide range of literature is available on parametric estimation and adaptive tracking for linear regression models [4], [5], [6] [9], [13], [14], [15], [16]. However, only few references may be found on nonparametric estimation in adaptive tracking [19], [20], [21],[23]. Our goal is to investigate the asymptotic properties of a kernel density estimator associated with the driven noise of a linear regression in adaptive tracking and to propose a goodness-of-fit test. Consider the parametric linear regression model given, for all  $n \geq 0$ , by

$$X_{n+1} = \theta^t \phi_n + U_n + \varepsilon_{n+1} \tag{1.1}$$

where  $X_n, U_n$  and  $\varepsilon_n$  are the  $d$ -dimensional system output, input and driven noise, respectively. The regression vector  $\phi_n$  is totally observable and  $\theta$  is the unknown parameter of the model. In all the sequel, we shall assume that  $(\varepsilon_n)$  is a sequence of centered independent and identically distributed random vectors with positive definite covariance matrix  $\Gamma$  and unknown probability density function denoted by  $f$ . The crucial role played by  $U_n$  is to regulate the dynamic of the process  $(X_n)$  by forcing  $X_n$  to track step by step a bounded predictable reference trajectory  $x_n$ . Via the certainty equivalence principle [1],  $U_n$  commonly called the adaptive control of the system, is given, for all  $n \geq 0$ , by

$$U_n = -\widehat{\theta}_n^t \phi_n + x_{n+1} \tag{1.2}$$

where  $\widehat{\theta}_n$  is the standard least squares estimator of  $\theta$  defined by

$$\widehat{\theta}_n = S_{n-1}^{-1} \sum_{k=0}^{n-1} \phi_k (X_{k+1} - U_k)^t \quad \text{with} \quad S_n = \sum_{k=0}^n \phi_k \phi_k^t + S$$

where a positive definite matrix  $S$  is added in order to avoid useless invertibility assumption. By substituting (1.2) into (1.1), we obtain the closed-loop system

$$X_{n+1} - x_{n+1} = \pi_n + \varepsilon_{n+1} \tag{1.3}$$

---

\*Laboratoire de Statistique et Probabilités, UMR C5583, Université Paul Sabatier, 31062 Toulouse Cedex 09, France ([bernard.bercu@math.ups-tlse.fr](mailto:bernard.bercu@math.ups-tlse.fr)).

†Laboratoire de Mathématiques, UMR C8628, Probabilités, Statistique et Modélisation, Université Paris-Sud, Bâtiment 425, 91405 Orsay Cedex, France, and IUT Paris, Université René-Descartes ([bruno.portier@math.u-psud.fr](mailto:bruno.portier@math.u-psud.fr)).

where  $\pi_n$  denotes the prediction error  $\pi_n = (\theta - \hat{\theta}_n)^t \phi_n$ .

In this paper, we shall study the asymptotic properties of a kernel density estimator (KDE) of  $f$ . This problem is well-known and has been widely investigated in time series. We refer the reader to [10], [11], [22] for some interesting books on density estimation for stationary processes. Let us now define the KDE of  $f$  in the context of the nonstationary model (1.1). When the sequence  $(\varepsilon_n)$  is observable, the traditional Parzen-Rosenblatt KDE of  $f$  is given, for all  $x \in \mathbb{R}^d$  and  $n \geq 1$ , by

$$f_n(x) = \frac{1}{nh_n^d} \sum_{i=1}^n K\left(\frac{\varepsilon_i - x}{h_n}\right)$$

where the kernel  $K$  is a chosen density function and the bandwidth  $(h_n)$  is a sequence of positive real numbers decreasing to zero. In our situation, the sequence  $(\varepsilon_n)$  is of course unobservable. However, when the tracking objective is fulfilled, the prediction error  $\pi_n$  is as closed as possible to zero. Consequently, via (1.3), we can choose  $X_n - x_n$  as a predictor of  $\varepsilon_n$ . Moreover, since we are in an adaptive tracking framework, it is more suitable to make use of a recursive kernel density estimator (RKDE) of  $f$  given, for all  $x \in \mathbb{R}^d$  and  $n \geq 1$ , by

$$\hat{f}_n(x) = \frac{1}{n} \sum_{i=1}^n \frac{1}{h_i^d} K\left(\frac{X_i - x_i - x}{h_i}\right). \quad (1.4)$$

Our purpose is first to show that  $\hat{f}_n$  behaves pretty well as a RKDE of  $f$  in adaptive tracking and secondly to provide a goodness-of-fit test for  $f$  based on  $\hat{f}_n$ . The paper is organized as follows. Section 2 is devoted to the asymptotic behavior of  $\hat{f}_n$ . We establish the almost sure pointwise and uniform convergence of  $\hat{f}_n$  to  $f$  as well as a pointwise law of iterated logarithm (LIL) and a pointwise multivariate central limit theorem (CLT). Section 3 is concerned with the goodness-of-fit test for  $f$ . Finally, some simulation experiments are given in Section 4. All technical proofs are postponed in appendices.

**2. Main results.** In all the sequel, we shall assume that the kernel  $K$  is non-negative function, bounded with compact support, such that

$$\int_{\mathbb{R}^d} K(t) dt = 1, \quad \int_{\mathbb{R}^d} K^2(t) dt = \tau^2$$

and, for any  $i = 1, 2, \dots, d$ ,

$$\int_{\mathbb{R}^d} t_i K(t) dt = 0.$$

For example, for some  $r > 0$  and some known positive constants  $a_r, b_r, c_r$ , one can make use of the uniform kernel on the sphere of  $\mathbb{R}^d$  with radius  $r$ ,  $K(t) = a_r \mathbb{I}_{(\|t\| \leq r)}$ , the Epanechnikov kernel with scaling factor  $r$ ,  $K(t) = b_r (1 - \|t\|^2/r^2) \mathbb{I}_{(\|t\| \leq r)}$ , and the Gaussian kernel with truncation level  $r$ ,  $K(t) = c_r \exp(-\|t\|^2/2) \mathbb{I}_{(\|t\| \leq r)}$ . Moreover, we shall assume that the bandwidth  $(h_n)$  is a sequence of positive real numbers, decreasing to zero, such that  $nh_n^d$  tends to infinity and satisfying the mild condition

$$\sum_{i=1}^n h_i^2 = O(nh_n^2).$$

This last condition, not really restrictive, is due to the recursive form of the KDE  $\widehat{f}_n$ . For example, one can choose  $h_n = n^{-\alpha}$  with  $\alpha \in ]0, 1/d[$ .

We shall now state our assumptions concerning the probability density function  $f$  and the prediction errors sequence  $(\pi_n)$ .

ASSUMPTION [A1]. *The density function  $f$  is strictly positive and belongs to  $C^2(\mathbb{R}^d)$ . In addition,  $f$  and its first and second derivatives are bounded.*

ASSUMPTION [A2]. *The prediction errors sequence  $(\pi_n)$  satisfies*

$$\sum_{i=0}^n \|\pi_i\|^2 = O(\log n) \quad \text{a.s.}$$

Assumption [A1] is classical in density nonparametric estimation while [A2] seems to be more restrictive. However, it is fulfilled in many situations in adaptive tracking. More precisely, assume that the driven noise  $(\varepsilon_n)$  has finite conditional moment of order  $> 2$ . Then, it follows from convergence (20) of [5] that [A2] holds for the ARX model

$$X_{n+1} = A_1 X_n + A_2 X_{n-1} + \dots + A_p X_{n-p+1} + U_n + \varepsilon_{n+1}$$

as soon as  $\sum_{i=1}^n \|x_i\|^2 = o(n)$  a.s. Moreover, we can also deduce from relation (11) of [6] that [A2] also holds for the scalar nonlinear autoregressive model

$$X_{n+1} = \theta \varphi(X_n, \dots, X_{n-p+1}) + U_n + \varepsilon_{n+1}$$

under suitable moment assumption on  $(\varepsilon_n)$  as soon as the function  $\varphi : \mathbb{R}^p \rightarrow \mathbb{R}$  does not increase to infinity faster than a polynomial of degree  $< 4$ .

We shall now present several asymptotic results for the RKDE  $\widehat{f}_n$  of  $f$ , the first one dealing with the almost sure convergence properties of  $\widehat{f}_n$ .

THEOREM 2.1. *Assume that [A1] and [A2] hold and that  $nh_n^d$  tends to infinity faster than  $(\log n)^2$ . Then for any  $x \in \mathbb{R}^d$ ,  $\widehat{f}_n(x)$  converges a.s. to  $f(x)$ . In addition, if  $h_n^d \log n = O(1)$  and  $nh_n^{d+4} = o(\log \log n)$ , we also have*

$$\limsup_{n \rightarrow \infty} \left( \frac{nh_n^d}{2\tau^2 \|f\|_\infty \log \log n} \right)^{1/2} \left| \widehat{f}_n(x) - f(x) \right| \leq 1 \quad \text{a.s.} \quad (2.1)$$

Moreover, assume that the kernel  $K$  is Lipschitz and that the bandwidth  $(h_n)$  is given by  $h_n = n^{-\alpha}$  with  $\alpha \in ]0, 1/d[$ . Then,  $\widehat{f}_n$  converges a.s. to  $f$ , uniformly on all compact sets of  $\mathbb{R}^d$  and, for any  $\beta \in ](1 + \alpha d)/2, 1[$ ,

$$\sup_{x \in \mathbb{R}^d} \left| \widehat{f}_n(x) - f(x) \right| = o(n^{\beta-1}) + O(n^{-2\alpha}) \quad \text{a.s.} \quad (2.2)$$

*Proof.* The proof is given in Appendix A.  $\square$

**Remark 1.** The bandwidth condition associated with the almost sure pointwise convergence result is really not restrictive and it is clearly satisfied when  $h_n = n^{-\alpha}$  with  $\alpha \in ]0, 1/d[$ . In that particular case, the first bandwidth condition required for the LIL is obviously satisfied and the second one holds as soon as  $\alpha \in ]1/(d+4), 1/d[$ .

**Remark 2.** It is possible to relax assumption [A1] by only assuming that  $f$  belongs to  $C^1(\mathbb{R}^d)$  with bounded gradient. In that case, we have to replace the bandwidth condition  $nh_n^{d+4} = o(\log \log n)$  by  $nh_n^{d+2} = o(\log \log n)$ . Moreover, in the expanded uniform strong law result, it is necessary to replace  $O(n^{-2\alpha})$  by  $O(n^{-\alpha})$  in (2.2).

Our second result is a pointwise and a multivariate CLT for  $\widehat{f}_n$ .

**THEOREM 2.2.** *Assume that [A1] and [A2] hold. Moreover, suppose that the bandwidth  $(h_n)$  satisfies  $nh_n^{d+4} = o(1)$  together with*

$$\lim_{n \rightarrow \infty} \frac{h_n^d}{n} \sum_{i=1}^n h_i^{-d} = \ell_h \quad (2.3)$$

for some finite constant  $\ell_h > 0$ . Then, for any  $x \in \mathbb{R}^d$ , we have the pointwise CLT

$$G_n(x) = \sqrt{nh_n^d} (\widehat{f}_n(x) - f(x)) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \tau^2 \ell_h f(x)) = G(x). \quad (2.4)$$

In addition, for  $q$  distinct points  $x_1, \dots, x_q$  of  $\mathbb{R}^d$ , we also have

$$(G_n(x_1), \dots, G_n(x_q)) \xrightarrow{\mathcal{L}} (G(x_1), \dots, G(x_q)) \quad (2.5)$$

where  $G(x_1), \dots, G(x_q)$  are independent.

*Proof.* The proof is given in Appendix B.  $\square$

**Remark 3.** Result (2.4) is identical to the one obtained by Dufflo [12] for stationary processes. Besides, it is worthless to require the bandwidth condition (2.3) for the nonrecursive KDE of  $f$  and  $\ell_h$  has to be replaced by 1 in (2.4). Finally, when  $h_n = n^{-\alpha}$  with  $\alpha \in ]0, 1/d[$ , it is necessary to assume that  $\alpha(d+4) > 1$  and we obviously have  $\ell_h = (1 + \alpha d)^{-1}$ .

**3. Application to a goodness-of-fit test.** We shall now propose a statistical test associated with the probability density function  $f$  based on the convergence results of Section 2. We wish to test

$$\mathcal{H}_0 : \langle\langle f = f_0 \rangle\rangle \quad \text{vs} \quad \mathcal{H}_1 : \langle\langle f \neq f_0 \rangle\rangle$$

where  $f_0$  is a given probability density function. It is well-known that such a goodness-of-fit test is very important and it has been widely investigated in time series analysis. Indeed, many statistical procedures require the assumption of normality for the driven white noise [3], [8]. Consequently, a goodness-of-fit test for the white noise density is of particular interest. For an independent and identically distributed sample, Bickel and Rosenblatt [7] have proposed a statistical test based on the integrated quadratic deviation between the true density and a KDE of  $f$ . This approach has been extended to the autoregressive framework by Lee and Na [17] and more recently by Bachmann and Dette [2]. However, no such a statistical test is available in the adaptive tracking framework although several situations require the normality assumption on the driven white noise. Our purpose is now to provide a goodness-of-fit test for  $f$  based on the multivariate CLT for  $\widehat{f}_n$  together with the LIL. Our statistical test consists on a

suitably normalized sum of the quadratic deviation between the true density and the RKDE  $\widehat{f}_n$  evaluated on  $q$  distinct points of  $\mathbb{R}^d$ . More precisely, it is defined by

$$T_n(q) = \frac{1}{\tau^2 \ell_h} \sum_{j=1}^q \frac{\left(\widehat{f}_n(x_j) - f_0(x_j)\right)^2}{\widehat{f}_n(x_j)}$$

where  $x_1, \dots, x_q$  are  $q$  distinct points of  $\mathbb{R}^d$ . We shall make use of

$$\sigma^2(q) = \frac{1}{\tau^2 \ell_h} \sum_{j=1}^q \frac{(f(x_j) - f_0(x_j))^2}{f(x_j)}, \quad \lambda^2(q) = \frac{1}{\tau^2 \ell_h} \sum_{j=1}^q \frac{(f^2(x_j) - f_0^2(x_j))^2}{f^3(x_j)}.$$

**THEOREM 3.1.** *Assume that [A1] and [A2] hold. Moreover, suppose that the bandwidth  $(h_n)$  shares the same assumptions as in Theorem 2.2 and is such that  $nh_n^d$  goes to infinity faster than  $(\log n)^2$ . Then, under  $\mathcal{H}_0$ ,*

$$nh_n^d T_n(q) \xrightarrow{\mathcal{L}} \chi^2(q). \quad (3.1)$$

Moreover, under  $\mathcal{H}_1$  and if one can find  $x \in \{x_1, x_2, \dots, x_q\}$  such that  $f(x) \neq f_0(x)$ , then  $T_n(q)$  converges a.s. towards  $\sigma^2(q)$ . In addition, if  $h_n^d \log n = O(1)$ , we also have

$$\sqrt{nh_n^d} (T_n(q) - \sigma^2(q)) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \lambda^2(q)). \quad (3.2)$$

**Remark 4.** According to these asymptotic results, it is possible to construct a goodness-of-fit test associated with  $f$ . On the one hand, under the null hypothesis  $\mathcal{H}_0$ , we can approximate for  $n$  large enough the distribution of  $nh_n^d T_n(q)$  by a  $\chi^2(q)$  one. On the other hand, under the alternative hypothesis  $\mathcal{H}_1$ , if  $\sigma^2(q)$  is positive,  $nh_n^d T_n(q)$  goes a.s. to infinity which guarantees that the asymptotic power of our test is equal to 1. From a practical point of view, the null hypothesis  $\mathcal{H}_0$  will be rejected at level  $\delta$  whenever  $nh_n^d T_q(n) > a_\delta$  where  $a_\delta$  stands for the  $(1 - \delta)$  quantile of the  $\chi^2(q)$  distribution. Finally, one can observe that the weak convergence (3.2) allows us to evaluate the probability of the type II error of our test.

**Remark 5.** It is also possible to make use of the test statistic  $T'_n(q)$  defined by

$$T'_n(q) = \frac{1}{\tau^2 \ell_h} \sum_{j=1}^q \frac{\left(\widehat{f}_n(x_j) - f_0(x_j)\right)^2}{f_0(x_j)}$$

In that case, Theorem 3.1 holds with

$$\sigma^2(q) = \frac{1}{\tau^2 \ell_h} \sum_{j=1}^q \frac{(f(x_j) - f_0(x_j))^2}{f_0(x_j)}, \quad \lambda^2(q) = \frac{4}{\tau^2 \ell_h} \sum_{j=1}^q \frac{(f(x_j) - f_0(x_j))^2 f(x_j)}{f_0^2(x_j)}.$$

This statistical test should improve the empirical level under  $\mathcal{H}_0$ , but it should certainly degrade the empirical power under  $\mathcal{H}_1$ . Nevertheless, it is easier to compute

than  $T_n(q)$  because it allows to avoid the division by  $\widehat{f}_n(x_j)$  which can be equal to zero due to the use of a compactly supported kernel.

*Proof.* The proof is straightforward by use of Theorem 2.1 together with Theorem 2.2. As a matter of fact, we have the decomposition

$$T_n(q) - \sigma^2(q) = A_n + B_n \quad (3.3)$$

where

$$A_n = \frac{1}{\tau^2 \ell_h} \sum_{j=1}^q \frac{(\widehat{f}_n(x_j) - f(x_j))^2}{\widehat{f}_n(x_j)},$$

$$B_n = \frac{1}{\tau^2 \ell_h} \sum_{j=1}^q \frac{(\widehat{f}_n(x_j) - f(x_j)) (f^2(x_j) - f_0^2(x_j))}{\widehat{f}_n(x_j) f(x_j)}.$$

We can deduce from (2.5) and the pointwise almost sure convergence of  $\widehat{f}_n$  to  $f$  that

$$\sqrt{\frac{nh_n^d}{\tau^2 \ell_h}} \left( \frac{\widehat{f}_n(x_1) - f(x_1)}{\sqrt{\widehat{f}_n(x_1)}}, \dots, \frac{\widehat{f}_n(x_q) - f(x_q)}{\sqrt{\widehat{f}_n(x_q)}} \right) \xrightarrow{\mathcal{L}} \mathcal{N}(0, I_q) \quad (3.4)$$

where  $I_q$  stands for the identity matrix of order  $q$ . Hence, it immediately follows from (3.4) that

$$nh_n^d A_n \xrightarrow{\mathcal{L}} \chi^2(q). \quad (3.5)$$

Consequently, we clearly obtain (3.1) from (3.3) together with (3.5) since, under the null hypothesis  $\mathcal{H}_0$ ,  $\sigma^2(q)$  and  $B_n$  vanish. Under the alternative hypothesis  $\mathcal{H}_1$ , it is straightforward to see that  $T_n(q)$  converges a.s. towards  $\sigma^2(q)$  via the almost sure pointwise convergence of  $\widehat{f}_n$  to  $f$ . Only convergence (3.2) remains to be proven. On the one hand, by the pointwise LIL, we infer that

$$|A_n| = O\left(\frac{\log \log n}{nh_n^d}\right) \quad \text{a.s.}$$

which implies that

$$\sqrt{nh_n^d} A_n = o(1) \quad \text{a.s.} \quad (3.6)$$

as  $nh_n^d$  goes to infinity faster than  $(\log n)^2$ . On the other hand, we can deduce from (3.4) that

$$\sqrt{nh_n^d} B_n \xrightarrow{\mathcal{L}} \mathcal{N}(0, \lambda^2(q)). \quad (3.7)$$

Finally, convergence (3.2) immediately follows from the conjunction of (3.3), (3.6) and (3.7) which completes the proof of Theorem 3.1.  $\square$

**4. Simulation experiments.** In this section, we shall investigate the finite sample properties of our statistical test both under  $\mathcal{H}_0$  and  $\mathcal{H}_1$  without some bootstrap procedure as it is usual in this context of nonparametric test. Since it has never been experimented, we shall not restrict ourselves to models of form (1.1) but we will also consider some closely related stationary models. Our goal is to show that our statistical test behaves pretty well in many different situations. The different models that we will study are given as follows.

$$(WN) \quad X_n = \varepsilon_n$$

$$(AR) \quad X_{n+1} = \theta X_n + \varepsilon_{n+1}$$

$$(ARX) \quad X_{n+1} = \theta X_n + U_n + \varepsilon_{n+1}$$

$$(NARX) \quad X_{n+1} = \theta X_n^2 + U_n + \varepsilon_{n+1}$$

where  $(\varepsilon_n)$  is a sequence of centered independent and identically distributed random variable with probability density function  $f$ . We choose  $\theta = 7/10$ ,  $\theta = 2$  and  $\theta = 1/2$  for the AR, ARX and NARX models, respectively. We consider three choices of noise distributions that we combine two by two in order to study the performances of our statistical test both under  $\mathcal{H}_0$  and  $\mathcal{H}_1$ . The first one is the standard normal distribution

$$f_0(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right).$$

The second one is the normalized double exponential distribution

$$f_1(x) = \frac{1}{\sqrt{2}} \exp\left(-\sqrt{2}|x|\right).$$

The last one is the standardized chi-square distribution with twelve degrees of freedom

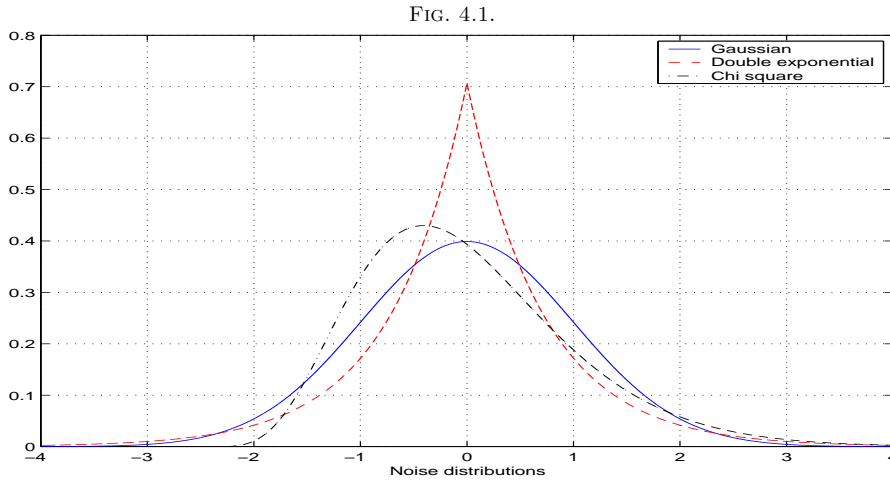
$$f_2(x) = \frac{9}{5}(x + \sqrt{6})^5 \exp\left(-\sqrt{6}(x + \sqrt{6})\right) \mathbb{I}_{(x \geq -\sqrt{6})}.$$

[Place Figure 4.1 about here]

For AR, ARX and NARX models, we estimate the unknown parameter  $\theta$  by use of the standard least squares estimator  $\hat{\theta}_n$ . For AR model, the probability density function  $f$  is estimated using the KDE given by (1.4) where  $X_n - x_n$  is replaced by  $X_n - \hat{\theta}_n X_{n-1}$ . For ARX and NARX models, the adaptive control  $U_n$  is given by  $U_n = -\hat{\theta}_n X_n$  and  $U_n = -\hat{\theta}_n X_n^2$ , respectively.

For each model and each test of  $\mathcal{H}_0$  against  $\mathcal{H}_1$ , we base our estimations on  $N = 800$  independent realizations of sample sizes  $n = 200, 500$  and  $1000$ . We are interested in the empirical level under  $\mathcal{H}_0$  to be compared with the theoretical level equal to 5% and the empirical power under  $\mathcal{H}_1$ , as well as the closeness between the simulated distribution of our statistical test and the corresponding theoretical





distribution. The implementation of our statistic test  $T_n(q)$  requires the choice of design points together with the specification of a bandwidth and a kernel for the RKDE  $\hat{f}_n$ . We use the design points selection rule proposed by Poggi and Portier in [18]. More precisely, we proceed as follows. Starting from an estimate of the distribution of the driven noise, we choose  $q$  equidistant points  $x_1 \dots, x_q$  so that the density at those points is not too small and in such a way that they are sufficiently distant to ensure sufficient accuracy in the use of the multivariate CLT. In other words, this last condition allows us to guarantee the asymptotic independence property in the multivariate CLT for small to moderate sample size. We choose  $q = 8, 13$  and  $22$  equidistant points for sample sizes  $n = 200, 500$  and  $1000$ , respectively. It should be noted that only a few number of points are needed to make a decision. The RKDE  $\hat{f}_n$  is constructed by use of the Epanechnikov kernel

$$K(t) = \frac{3}{4}(1 - t^2)\mathbb{I}_{(|t| \leq 1)}$$

and the bandwidth  $h_n = n^{-\alpha}$  with  $\alpha = 1/3$ . For the denominator of  $T_n(q)$ , we use the Gaussian kernel and the bandwidth  $h_n = n^{-\beta}$  with  $\beta = 1/5$ . Via this choice, we avoid a possible division by zero and we provide a smoother version for the estimation of  $f$ . Finally, for ARX and NARX models, we use a short learning period of  $\tau = 100$  time steps. This learning period allows us to forget the transitory phase.

The abbreviations  $\mathcal{G}f_0$ ,  $\mathcal{G}f_1$  and  $\mathcal{G}f_2$  mean that the driven noise ( $\varepsilon_n$ ) is generated with the normal  $f_0$  distribution, the double exponential  $f_1$  distribution and the chi-square  $f_2$  distribution, respectively while  $\mathcal{H}f_0$ ,  $\mathcal{H}f_1$  and  $\mathcal{H}f_2$  mean that we are testing the assumptions  $\mathcal{H}_0 : \langle f = f_0 \rangle$ ,  $\mathcal{H}_0 : \langle f = f_1 \rangle$  and  $\mathcal{H}_0 : \langle f = f_2 \rangle$ , respectively. Finally, as we have chosen a test level  $\alpha = 5\%$  and we have generated  $N = 800$  trials, the Kolmogorov-Smirnov fitting statistic in *italic* has to be compared with the value 0.048.

[Place Table 4.1 about here]

[Place Table 4.2 about here]

TABLE 4.1

**WN model.** Results under  $\mathcal{H}_0$  and  $\mathcal{H}_1$  with test level 5%. Empirical level in bold and percentage of correct decisions.

	$n = 200, q = 8$			$n = 500, q = 13$			$n = 1000, q = 22$		
	$\mathcal{H}f_0$	$\mathcal{H}f_1$	$\mathcal{H}f_2$	$\mathcal{H}f_0$	$\mathcal{H}f_1$	$\mathcal{H}f_2$	$\mathcal{H}f_0$	$\mathcal{H}f_1$	$\mathcal{H}f_2$
$\mathcal{G}f_0$	<b>4.2%</b> <i>0.035</i>	35.7%	26.2%	<b>5.3%</b> <i>0.029</i>	84.1%	70%	<b>5.2%</b> <i>0.024</i>	99.8%	98.6%
$\mathcal{G}f_1$	49%	<b>5.3%</b> <i>0.047</i>	74.1%	91.2%	<b>5.1%</b> <i>0.041</i>	99.3%	100%	<b>4.2%</b> <i>0.030</i>	100%
$\mathcal{G}f_2$	19.2%	53.5%	<b>4.2%</b> <i>0.047</i>	60%	97.3%	<b>4.7%</b> <i>0.031</i>	96.7%	100%	<b>4.5%</b> <i>0.009</i>

TABLE 4.2

**AR model.** Results under  $\mathcal{H}_0$  and  $\mathcal{H}_1$  with test level 5%. Empirical level in bold and percentage of correct decisions.

	$n = 200, q = 8$			$n = 500, q = 13$			$n = 1000, q = 22$		
	$\mathcal{H}f_0$	$\mathcal{H}f_1$	$\mathcal{H}f_2$	$\mathcal{H}f_0$	$\mathcal{H}f_1$	$\mathcal{H}f_2$	$\mathcal{H}f_0$	$\mathcal{H}f_1$	$\mathcal{H}f_2$
$\mathcal{G}f_0$	<b>4.5%</b> <i>0.045</i>	31.2%	25.6%	<b>4.8%</b> <i>0.014</i>	82%	65%	<b>3.7%</b> <i>0.023</i>	99.8%	98.8%
$\mathcal{G}f_1$	49.7%	<b>5.7%</b> <i>0.032</i>	73.1%	90.5%	<b>5%</b> <i>0.014</i>	99.1%	100%	<b>4.8%</b> <i>0.019</i>	100%
$\mathcal{G}f_2$	19.3%	54.6%	<b>3.7%</b> <i>0.045</i>	62%	96.6%	<b>3.5%</b> <i>0.022</i>	96.6%	100%	<b>3.8%</b> <i>0.013</i>

[Place Table 4.3 about here]

[Place Table 4.4 about here]

We shall now comment the test results contained in Tables 4.1 to 4.4. First of all, one can verify that our statistical test behaves pretty well under  $\mathcal{H}_0$ . Indeed, for each model and each noise distribution, the empirical level is close to the 5% theoretical value level as one can realize with the values in bold. In addition, the simulated distribution of  $T_n(q)$  is close to the  $\chi^2(q)$  distribution as one can observe with the values in italic of the Kolmogorov-Smirnov fitting statistic to be compared with the critical value at 5% equal to 0.048. Next, one can verify that the empirical power increases with the sample size, from 20% to 40% for  $n = 200$ , to 96% to 100% for  $n = 1000$ : it is more difficult to decide between  $f_0$  and  $f_2$  than between  $f_1$  and  $f_2$ , which is the easier situation. Finally, if one superimpose the four tables, one can observe that the results for the different models are almost the same. In conclusion, our statistical test behaves pretty well for small to moderate sample size and for a large class of models.

[Place Figure 4.2 about here]

Figure 4.2 illustrates the empirical power of our test for the NARX model. We base our estimation on  $N = 800$  trials of sample size  $n = 500$  with  $q = 13$  equidistant points. The driven noise  $(\varepsilon_n)$  is generated with the normal distribution  $f_0$  and we are testing the assumption  $\mathcal{H}_0: \langle f = f_1 \rangle$ . One can observe that the distribution of our statistical test  $T_n(q)$  is different from the  $\chi^2(q)$  one. Consequently, the power of

TABLE 4.3

**ARX model.** Results under  $\mathcal{H}_0$  and  $\mathcal{H}_1$  with test level 5% and learning period  $\tau = 100$ . Empirical level in bold and percentage of correct decisions.

	$n = 200, q = 8$			$n = 500, q = 13$			$n = 1000, q = 22$		
	$\mathcal{H}f_0$	$\mathcal{H}f_1$	$\mathcal{H}f_2$	$\mathcal{H}f_0$	$\mathcal{H}f_1$	$\mathcal{H}f_2$	$\mathcal{H}f_0$	$\mathcal{H}f_1$	$\mathcal{H}f_2$
$\mathcal{G}f_0$	<b>3.8%</b> 0.042	35.7%	28%	<b>4.2%</b> 0.029	81.5%	66%	<b>3.7%</b> 0.018	99.7%	98.2%
$\mathcal{G}f_1$	45.8%	<b>5.5%</b> 0.053	71.5%	87.5%	<b>4.7%</b> 0.021	99.3%	100%	<b>5%</b> 0.022	100%
$\mathcal{G}f_2$	21.2%	54.5%	<b>3.2%</b> 0.029	62%	95.6%	<b>2.5%</b> 0.040	96.7%	100%	<b>5.1%</b> 0.029

TABLE 4.4

**NARX model.** Results under  $\mathcal{H}_0$  and  $\mathcal{H}_1$  with test level 5% and learning period  $\tau = 100$ . Empirical level in bold and percentage of correct decisions.

	$n = 200, q = 8$			$n = 500, q = 13$			$n = 1000, q = 22$		
	$\mathcal{H}f_0$	$\mathcal{H}f_1$	$\mathcal{H}f_2$	$\mathcal{H}f_0$	$\mathcal{H}f_1$	$\mathcal{H}f_2$	$\mathcal{H}f_0$	$\mathcal{H}f_1$	$\mathcal{H}f_2$
$\mathcal{G}f_0$	<b>3%</b> 0.037	37.1%	28.5%	<b>4.8%</b> 0.029	83.5%	68.2%	<b>4.3%</b> 0.037	99.5%	98.6%
$\mathcal{G}f_1$	44.6%	<b>5.2%</b> 0.021	72%	89.8%	<b>4.5%</b> 0.022	99.2%	100%	<b>5.1%</b> 0.017	100%
$\mathcal{G}f_2$	19.8%	58.3%	<b>3.7%</b> 0.021	63.2%	95.5%	<b>4.7%</b> 0.05	97.2%	100%	<b>5%</b> 0.039

separation of our statistical test is clearly significant.

**Appendix A.** This appendix is devoted to the proof of Theorem 2.1. In order to prove the asymptotic properties of our RKDE  $\widehat{f}_n$  of  $f$ , we are led to introduce the martingale  $(M_n)$  associated with the sequence  $(\widehat{f}_n)$ . To be more precise, we immediately infer from (1.4) that for all  $x \in \mathbb{R}^d$  and  $n \geq 1$ ,

$$n(\widehat{f}_n(x) - f(x)) = M_n(x) + R_n(x) \quad (\text{A.1})$$

with

$$M_n(x) = \sum_{i=1}^n \left( K_i(X_i - x_i - x) - \mathbb{E}[K_i(X_i - x_i - x) | \mathcal{F}_{i-1}] \right) \quad (\text{A.2})$$

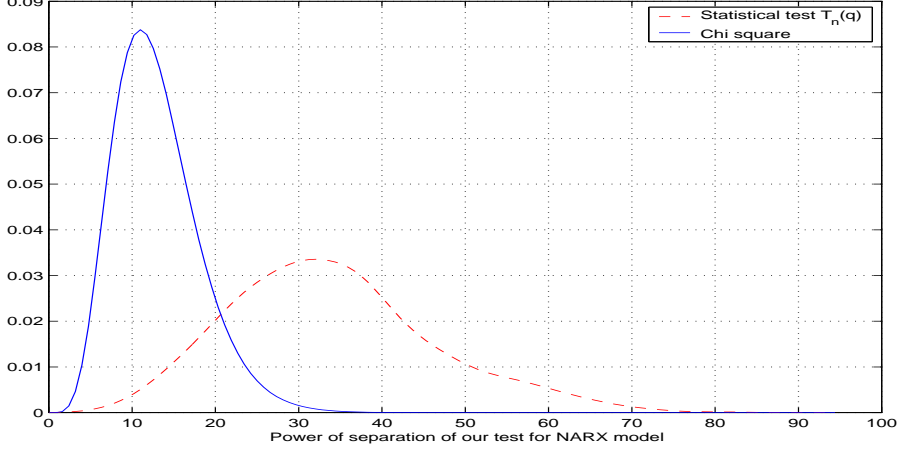
$$R_n(x) = \sum_{i=1}^n \mathbb{E}[K_i(X_i - x_i - x) | \mathcal{F}_{i-1}] - n f(x) \quad (\text{A.3})$$

where, for all  $y \in \mathbb{R}^d$ ,  $K_n(y) = h_n^{-d} K(h_n^{-1}y)$  and  $\mathcal{F}_n$  denotes the  $\sigma$ -algebra of the events occurring up to time  $n$ . The almost sure properties of  $(M_n)$  are given by the two following lemmas.

**LEMMA A.1.** Assume that  $nh_n^d$  tends to infinity faster than  $(\log n)^2$ . Then, for any  $x \in \mathbb{R}^d$ , we have  $M_n(x) = o(n)$  a.s. More precisely,

$$\limsup_{n \rightarrow \infty} \frac{|M_n(x)|}{\sqrt{2\tau^2 \|f\|_\infty n h_n^{-d} \log \log n}} \leq 1 \quad \text{a.s.} \quad (\text{A.4})$$

FIG. 4.2.



*Proof.* For any  $x \in \mathbb{R}^d$ ,  $(M_n(x))$  is a locally square integrable real martingale. In addition, its increasing process  $\langle M(x) \rangle_n$  satisfies  $\langle M(x) \rangle_n = O(nh_n^{-d})$ . As a matter of fact, for all  $x \in \mathbb{R}^d$ ,

$$\langle M(x) \rangle_n = \sum_{i=1}^n \mathbb{E} \left[ K_i^2(X_i - x_i - x) | \mathcal{F}_{i-1} \right] - \sum_{i=1}^n \left( \mathbb{E} [K_i(X_i - x_i - x) | \mathcal{F}_{i-1}] \right)^2.$$

Consequently, we deduce from (1.3) that for all  $x \in \mathbb{R}^d$

$$\langle M(x) \rangle_n \leq \sum_{i=1}^n h_i^{-2d} \int_{\mathbb{R}^d} K^2(h_i^{-1}(\pi_{i-1} + s - x)) f(s) ds. \quad (\text{A.5})$$

Via the change of variables  $t = h_i^{-1}(\pi_{i-1} + s - x)$  into (A.5), we find that

$$\langle M(x) \rangle_n \leq \sum_{i=1}^n h_i^{-d} \int_{\mathbb{R}^d} K^2(t) f(h_i t + x - \pi_{i-1}) dt \leq \tau^2 \|f\|_\infty \sum_{i=1}^n h_i^{-d}.$$

Therefore, as  $(h_n)$  is decreasing,  $\langle M(x) \rangle_n = O(nh_n^{-d})$ . Hence, it follows from the strong law of large numbers for martingales (see e.g. [12], Theorem 1.3.15, p. 20) that for all  $\gamma > 0$ ,

$$|M_n(x)|^2 = o\left(nh_n^{-d}(\log n)^{1+\gamma}\right) \quad \text{a.s.}$$

which ensures that  $M_n(x) = o(n)$  a.s. since  $n h_n^{-d}$  tends to infinity faster than  $(\log n)^2$ . Furthermore, for any  $x \in \mathbb{R}^d$ ,  $|M_n(x) - M_{n-1}(x)| \leq 2h_n^{-d} \|K\|_\infty$  which clearly implies that

$$|M_n(x) - M_{n-1}(x)| \leq C_n \sqrt{\frac{nh_n^{-d}}{\log \log n}}$$

where  $(C_n)$  is a deterministic sequence which tends to zero. Finally, we immediately obtain (A.4) from the upper bound in the law of iterated logarithm for martingales (see e.g. [12], Theorem 6.4.24, p. 209).  $\square$

LEMMA A.2. *Assume that the kernel  $K$  is Lipschitz and that the bandwidth  $(h_n)$  is given by  $h_n = n^{-\alpha}$  with  $\alpha \in ]0, 1/d[$ . Then, for any constants  $A > 0$  and  $\gamma > 0$ , we have the expanded uniform strong law*

$$\sup_{\|x\| \leq An^\gamma} |M_n(x)| = o(n^\beta) \quad a.s. \quad (\text{A.6})$$

where  $\beta \in ](1 + \alpha d)/2, 1[$ .

*Proof.* Formula (A.6) is a direct application of the expanded uniform strong law for martingales given by Theorem 6.4.34, p. 220 of [12]. First of all, for all  $x \in \mathbb{R}^d$ , set  $\Delta M_n(x) = M_n(x) - M_{n-1}(x)$ . We already saw in the proof of Lemma A.1 that there exists two positive constants  $a, b$  such that, for all  $n \geq 1$ ,  $\langle M(0) \rangle_n \leq an^{1+\alpha d}$  and  $|\Delta M_n(0)| \leq bn^{\alpha d}$ . In addition, since the kernel  $K$  is bounded and Lipschitz, for all  $\delta \in ]0, 1[$ , one can find some positive constant  $C_\delta$  such that, for any  $x, y \in \mathbb{R}^d$

$$|K(x) - K(y)| \leq C_\delta \|x - y\|^\delta. \quad (\text{A.7})$$

Hence, it is not hard to see that, for any  $x, y \in \mathbb{R}^d$

$$|\Delta M_n(x) - \Delta M_n(y)| \leq 2C_\delta \|x - y\|^\delta n^{\alpha(d+\delta)}.$$

Furthermore, similarly to (A.5), we have for any  $x, y \in \mathbb{R}^d$

$$\langle M(x) - M(y) \rangle_n \leq \sum_{i=1}^n i^{2\alpha d} \int_{\mathbb{R}^d} \left( K(i^\alpha(\pi_{i-1} + s - x)) - K(i^\alpha(\pi_{i-1} + s - y)) \right)^2 f(s) ds$$

which, by the change of variables  $t = i^\alpha(\pi_{i-1} + s - x)$ , leads to

$$\langle M(x) - M(y) \rangle_n \leq \|f\|_\infty \sum_{i=1}^n i^{\alpha d} \int_{\mathbb{R}^d} \left( K(t) - K(t + i^\alpha(x - y)) \right)^2 dt. \quad (\text{A.8})$$

In addition, as  $K$  is a density function, it follows using (A.7) that

$$\int_{\mathbb{R}^d} \left( K(t) - K(t + i^\alpha(x - y)) \right)^2 dt \leq 2C_{2\delta} \|x - y\|^{2\delta} i^{2\alpha\delta}.$$

Therefore, we deduce from (A.8) that for any  $x, y \in \mathbb{R}^d$

$$\langle M(x) - M(y) \rangle_n \leq 2C_{2\delta} \|x - y\|^{2\delta} n^{1+\alpha d+2\alpha\delta}.$$

Since the power  $\delta$  can be chosen as small as one wishes, all the four conditions of Theorem 6.4.34 of [12] are fulfilled which immediately leads to Lemma A.2.  $\square$

**Proof of Theorem 2.1.** In order to prove Theorem 2.1, it remains to study the almost sure asymptotic behavior of the remainder  $(R_n)$  in (A.1). It follows from (A.3) that

$$\begin{aligned} R_n(x) &= \sum_{i=1}^n h_i^{-d} \int_{\mathbb{R}^d} K(h_i^{-1}(\pi_{i-1} + s - x)) f(s) ds - nf(x), \\ &= \sum_{i=1}^n \int_{\mathbb{R}^d} K(t) \left( f(h_i t + x - \pi_{i-1}) - f(x) \right) dt, \end{aligned}$$

via the usual change of variables  $t = h_i^{-1}(\pi_{i-1} + s - x)$ . As the probability density function  $f$  belongs to  $C^2(\mathbb{R}^d)$  with bounded derivatives of order 2, we obtain by a Taylor expansion that

$$\sup_{x \in \mathbb{R}^d} |R_n(x)| = O\left(\sum_{i=1}^n h_i^2\right) + O\left(\sum_{i=1}^n \|\pi_{i-1}\|\right) \quad \text{a.s.}$$

Hence, it follows from the Cauchy-Schwarz inequality together with [A2] that

$$\sup_{x \in \mathbb{R}^d} |R_n(x)| = O(nh_n^2) + O(\sqrt{n \log n}) \quad \text{a.s.} \quad (\text{A.9})$$

Consequently,  $R_n(x) = o(n)$  a.s. which ensures that  $\widehat{f}_n(x)$  converges a.s. to  $f(x)$ . Moreover, we obtain (2.1) from the conjunction of Lemma A.1 and result (A.9). The uniform almost sure convergence on  $\mathbb{R}^d$  still remains to be proven. Recall that in that case,  $h_n = n^{-\alpha}$  with  $\alpha \in ]0, 1/d[$ . On the one hand, we find from Lemma A.2 with  $A = 2$  and  $\gamma = 1/2$ , that

$$\sup_{\|x\| \leq 2\sqrt{n}} |M_n(x)| = o(n^\beta) \quad \text{a.s.} \quad (\text{A.10})$$

where  $\beta \in ](1 + \alpha d)/2, 1[$ . One can observe that  $\beta > 1/2$  which implies that  $\sqrt{n \log n} = o(n^\beta)$ . Hence, taking it into account and combining (A.9) with (A.10), we deduce that

$$\sup_{\|x\| \leq 2\sqrt{n}} \left| \widehat{f}_n(x) - f(x) \right| = o(n^{\beta-1}) + O(n^{-2\alpha}) \quad \text{a.s.} \quad (\text{A.11})$$

On the other hand, we claim that

$$\sup_{\|x\| > 2\sqrt{n}} \left| \widehat{f}_n(x) - f(x) \right| = O\left(\frac{1}{n}\right) \quad \text{a.s.} \quad (\text{A.12})$$

As a matter of fact, since  $(\varepsilon_n)$  has a finite moment of order 2, we infer from (1.3) together with [A2] that

$$\sup_{k \leq n} \|X_k - x_k\|^2 = o(n) \quad \text{a.s.}$$

Hence, for  $k$  large enough,  $\|X_k - x_k\| < \sqrt{n}$  a.s. which ensures that, for  $x$  such that  $\|x\| > 2\sqrt{n}$ ,  $\|X_k - x_k - x\| > \sqrt{n}$  a.s. Therefore, since  $K$  is compactly supported, it clearly leads to

$$\sup_{\|x\| > 2\sqrt{n}} \left| n \widehat{f}_n(x) \right| = \sup_{\|x\| > 2\sqrt{n}} \left| \sum_{i=1}^n K_i(X_i - x_i - x) \right| = O(1) \quad \text{a.s.} \quad (\text{A.13})$$

In addition, since  $(\varepsilon_n)$  has a finite moment of order 2 and  $f$  is positive, it follows that  $f(x) = O(\|x\|^{-3})$  for large values of  $x$ , leading to

$$\sup_{\|x\| > 2\sqrt{n}} f(x) = O\left(\frac{1}{n}\right). \quad (\text{A.14})$$

Consequently, we immediately obtain (A.12) from (A.13) and (A.14). Finally, we deduce (2.2) from (A.11) and (A.12) which completes the proof of Theorem 2.1.

**Appendix B.** This appendix is concerned with the proof of Theorem 2.2. We first propose a CLT for the martingale  $(M_n)$ .

LEMMA B.1. *Assume that [A1] and [A2] hold and that the bandwidth  $(h_n)$  shares the same assumptions as in Theorem 2.2. Then, for any  $x \in \mathbb{R}^d$ ,*

$$\frac{M_n(x)}{\sqrt{nh_n^{-d}}} \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, \tau^2 \ell_h f(x)\right) \quad (\text{B.1})$$

*Proof.* In order to prove Lemma B.1, it is necessary to study the asymptotic behavior of the increasing process  $(\langle M(x) \rangle_n)$  properly normalized. For all  $i \geq 1$  and  $x \in \mathbb{R}^d$ , we have

$$\begin{aligned} \mathbb{E}[K_i(X_i - x_i - x) | \mathcal{F}_{i-1}] &= h_i^{-d} \int_{\mathbb{R}^d} K(h_i^{-1}(\pi_{i-1} + s - x)) f(s) ds \\ &= \int_{\mathbb{R}^d} K(t) f(h_i t + x - \pi_{i-1}) dt \leq \|f\|_\infty \end{aligned}$$

which implies that

$$\sum_{i=1}^n \left( \mathbb{E}[K_i(X_i - x_i - x) | \mathcal{F}_{i-1}] \right)^2 = O(n) \quad \text{a.s.} \quad (\text{B.2})$$

Moreover, we also have

$$\begin{aligned} \mathbb{E}[K_i^2(X_i - x_i - x) | \mathcal{F}_{i-1}] &= h_i^{-2d} \int_{\mathbb{R}^d} K^2(h_i^{-1}(\pi_{i-1} + s - x)) f(s) ds, \\ &= h_i^{-d} \int_{\mathbb{R}^d} K^2(t) f(h_i t + x - \pi_{i-1}) dt. \end{aligned}$$

Consequently, we obtain the decomposition

$$\sum_{i=1}^n \mathbb{E}[K_i^2(X_i - x_i - x) | \mathcal{F}_{i-1}] = A_n + \tau^2 B_n + \tau^2 f(x) C_n$$

where

$$\begin{aligned} A_n &= \sum_{i=1}^n h_i^{-d} \int_{\mathbb{R}^d} K^2(t) \left( f(h_i t + x - \pi_{i-1}) - f(x - \pi_{i-1}) \right) dt, \\ B_n &= \sum_{i=1}^n h_i^{-d} \left( f(x - \pi_{i-1}) - f(x) \right), \\ C_n &= \sum_{i=1}^n h_i^{-d}. \end{aligned}$$

As the gradient of  $f$  is bounded, we clearly have  $|A_n| = O(nh_n^{1-d})$  a.s. and

$$|B_n| = O\left( \sum_{i=1}^n h_i^{-d} \|\pi_{i-1}\| \right) \quad \text{a.s.}$$

Hence, it follows from [A2] that  $|B_n| = O(h_n^{-d} \sqrt{n \log n})$  a.s. Furthermore, we immediately get from (2.3) that  $n^{-1} h_n^d C_n$  converges to  $\ell_h$  as  $n$  goes to infinity. Putting together those three contributions, we find that

$$\lim_{n \rightarrow \infty} \frac{h_n^d}{n} \langle M(x) \rangle_n = \tau^2 \ell_h f(x) \quad \text{a.s.} \quad (\text{B.3})$$

In order to make use of the CLT for martingales (see e.g. [12], Corollary 2.1.10, p.46), it remains to check that Lindeberg's condition is satisfied. For all  $a > 0$ , let

$$\Lambda_n(a) = \frac{h_n^d}{n} \sum_{i=1}^n \mathbb{E} \left[ |\Delta M_i(x)|^2 \mathbb{1}_{(|\Delta M_i(x)| \geq a \sqrt{nh_n^{-d}})} | \mathcal{F}_{i-1} \right].$$

We already saw that for all  $i \leq n$ ,  $|\Delta M_i(x)| \leq 2 h_n^{-d} \|K\|_\infty$ . Hence, we clearly have for all  $i \leq n$

$$\mathbb{1}_{(|\Delta M_i(x)| \geq a \sqrt{nh_n^{-d}})} \leq \mathbb{1}_{(2\|K\|_\infty \geq a \sqrt{nh_n^{-d}})}.$$

Consequently, we find that for all  $a > 0$

$$\begin{aligned} \Lambda_n(a) &\leq \frac{h_n^d}{n} \mathbb{1}_{(2\|K\|_\infty \geq a \sqrt{nh_n^{-d}})} \sum_{i=1}^n \mathbb{E} \left[ |\Delta M_i|^2(x) | \mathcal{F}_{i-1} \right], \\ &\leq \frac{h_n^d}{n} \mathbb{1}_{(2\|K\|_\infty \geq a \sqrt{nh_n^{-d}})} \|K\|_\infty \|f\|_\infty \sum_{i=1}^n h_i^{-d}, \\ &\leq \|K\|_\infty \|f\|_\infty \mathbb{1}_{(2\|K\|_\infty \geq a \sqrt{nh_n^{-d}})}. \end{aligned}$$

Therefore, as  $nh_n^d$  tends to infinity, we obviously deduce that, for all  $a > 0$ ,  $\Lambda_n(a)$  tends to zero a.s. Finally, Lindeberg's condition is satisfied which achieves the proof of Lemma B.1.  $\square$

**Proof of Theorem 2.2.** We are now in position to prove Theorem 2.2. It follows from (A.1) that for any  $x \in \mathbb{R}^d$

$$\sqrt{nh_n^d} \left( \widehat{f}_n(x) - f(x) \right) = \frac{M_n(x) + R_n(x)}{\sqrt{nh_n^{-d}}}. \quad (\text{B.4})$$

Consequently, (2.4) immediately follows from (A.9) together with (B.1) and (B.4) as soon as  $nh_n^{d+4} = o(1)$ . The multivariate CLT remains to be proven. Taking the previous results into account, it suffices to prove that for two distinct points  $x, y \in \mathbb{R}^d$ , the random vector

$$\frac{1}{\sqrt{nh_n^{-d}}} \begin{pmatrix} M_n(x) \\ M_n(y) \end{pmatrix} \xrightarrow{\mathcal{L}} \begin{pmatrix} G(x) \\ G(y) \end{pmatrix}$$

where  $G(x)$  and  $G(y)$  are independent. We can easily show this convergence by remarking that for two distinct points  $x, y \in \mathbb{R}^d$

$$\lim_{n \rightarrow \infty} \frac{h_n^d}{n} \sum_{i=1}^n \mathbb{E} [\Delta M_i(x) \Delta M_i(y) | \mathcal{F}_{i-1}] = 0 \quad \text{a.s.} \quad (\text{B.5})$$



Indeed, for all  $i \geq 1$ , we have

$$\begin{aligned} \mathbb{E} [\Delta M_i(x) \Delta M_i(y) | \mathcal{F}_{i-1}] &\leq \mathbb{E} [K_i(X_i - x_i - x) K_i(X_i - x_i - y) | \mathcal{F}_{i-1}], \\ &\leq h_i^{-2d} \int_{\mathbb{R}^d} K(h_i^{-1}(\pi_{i-1} + s - x)) K(h_i^{-1}(\pi_{i-1} + s - y)) f(s) ds, \\ &\leq h_i^{-d} \int_{\mathbb{R}^d} K(t) K(t + h_i^{-1}(x - y)) f(h_i t + x - \pi_{i-1}) dt. \end{aligned}$$

Therefore, as the gradient of  $f$  is bounded, we obtain that

$$\sum_{i=1}^n \mathbb{E} [\Delta M_i(x) \Delta M_i(y) | \mathcal{F}_{i-1}] \leq H_n(x, y) + O\left(n h_n^{1-d}\right) + O\left(h_n^{-d} \sqrt{n \log n}\right) \quad \text{a.s.}$$

where

$$H_n(x, y) = \sum_{i=1}^n h_i^{-d} f(x) \int_{\mathbb{R}^d} K(t) K(t + h_i^{-1}(x - y)) dt.$$

However, using the fact that  $K$  is compactly supported, we can deduce that for  $i$  large enough, the integral at the right hand side of  $H_n(x, y)$  is zero. Finally, we obtain that convergence (B.5) is satisfied, which completes the proof of Theorem 2.2.

#### REFERENCES

- [1] K.J. ASTRÖM AND B. WITTENMARK, *Adaptive Control*, 2nd edition, Addison-Wesley, New York, 1995.
- [2] D. BACHMANN AND H. DETTE, *A note on the Bickel-Rosenblatt test in autoregressive time series*, Stat. & Prob. Letters, 74, (2005), pp. 221-234.
- [3] M. BASSEVILLE AND I.V. NIKIFOROV, *Detection of abrupt changes - Theory and application*, Prentice Hall, 1993.
- [4] B. BERCU, *Weighted estimation and tracking for ARMAX models*, SIAM J. Control Optim., 33 (1995), pp. 89-106.
- [5] B. BERCU, *Central limit theorem and law of iterated logarithm for least squares algorithms in adaptive tracking*, SIAM J. Control Optim., 36 (1998), pp. 910-928.
- [6] B. BERCU AND B. PORTIER, *Adaptive control of parametric nonlinear autoregressive models via a new martingale approach*, IEEE Trans. Automat. Control, 47, (2002), pp. 1524-1528.
- [7] P. BICKEL AND P. ROSENBLATT, *On some global measures of the deviation of density function estimators*, Ann. Statist. 1, (1973), pp. 1071-1095.
- [8] P.J. BROCKWELL AND R.A. DAVIES, *Time Series : Theory and Methods*, 2nd ed., Series in Statistics, Springer, New York, 1991.
- [9] H. F. CHEN AND L. GUO, *Identification and Stochastic Adaptive Control*, Birkhäuser, Boston, 1991.
- [10] L. DEVROYE, *A Course in Density Estimation*, Birkhäuser, Boston, 1987.
- [11] L. DEVROYE AND G. LUGOSI, *Combinatorial Methods in Density Estimation*, Springer Verlag, New York, 2001.
- [12] M. DUFLO, *Random Iterative Models*, Springer Verlag, Berlin, 1997.
- [13] L. GUO AND H. F. CHEN, *The Aström Wittenmark self-tuning regulator revisited and ELS-based adaptive trackers*, IEEE Trans. Automat. Control, 36 (1991), pp. 802-812.
- [14] L. GUO, *Further results on least squares based adaptive minimum variance control*, SIAM J. Control Optim., 32 (1994), pp. 187-212.
- [15] L. GUO, *Self convergence of weighted least squares with applications to stochastic adaptive control*, IEEE Trans. Automat. Control, 41 (1996), pp. 79-89.
- [16] T. L. LAI AND C. Z. WEI, *Extended least squares and their applications to adaptive control and prediction in linear systems*, IEEE Trans. Automat. Control, 31 (1986), pp. 898-906.
- [17] S. LEE AND S. NA, *On the Bickel-Rosenblatt test for first-order autoregressive models*, Stat. & Prob. Letters, 56, 1, (2002), pp. 23-35.

- [18] J.M. POGGI AND B. PORTIER, *A test of linearity for functional autoregressive models*, Journ. of Time Ser. Ana., Vol 18, 6, (1997), pp. 615–640.
- [19] J.-M. POGGI AND B. PORTIER, *Nonlinear adaptive tracking using kernel estimators : estimation and test for linearity*, SIAM J. Control Optim., 39, 3, (2000), pp. 707-727.
- [20] B. PORTIER, *Adaptive control of discrete-time nonlinear systems combining nonparametric and parametric estimators*, Communications in Information and Systems, 2 (2002), pp. 69-90.
- [21] B. PORTIER AND A. OULIDI, *Nonparametric estimation and adaptive control of functional autoregressive models*, SIAM J. Control Optim., 39, 2, (2000), pp. 411-432.
- [22] B. W. SILVERMAN *Density estimation for statistics and data analysis*, Chapman & Hall, New York, 1986.
- [23] L-L. XIE AND L. GUO, *How much uncertainty can be dealt with by feedback ?*. IEEE Trans. on Aut. Cont., 45 (2000), pp. 2203-2217.