SIAM J. CONTROL OPTIM.
Vol. 47, No. 5, pp. 2440–2457

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 of the driven noise of multivariate ARMAX models 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 carry out 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

DOI. 10.1137/070694739

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 a few references may be found on nonparametric estimation in adaptive tracking [20], [21], [22], [25]. 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 carry out a goodness-of-fit test. Consider the multivariate ARMAX model of order (p, q, r) given, for all $n \geq 0$, by

$$A(R)X_n = B(R)U_n + C(R)\varepsilon_n, \tag{1.1}$$

where X_n, U_n, and ε_n are the d-dimensional system output, input, and driven noise, respectively. Denote by R the shift-back operator and set

$$\begin{aligned} A(R) &= I_d - A_1R - \cdots - A_pR^p, \\ B(R) &= B_1R + \cdots + B_qR^q, \\ C(R) &= I_d + C_1R + \cdots + C_rR^r, \end{aligned}$$

where A_i, B_j, and C_k are unknown matrices and I_d is the identity matrice of order d. For the sake of simplicity, we shall assume that the high frequency gain matrix B_1 is known with $B_1 = I_d$. Hence, the unknown parameter of the model is given by

$$\theta^{\,t} = (A_1, \ldots, A_p, B_2, \ldots, B_q, C_1, \ldots, C_r).$$

Relation (1.1) can be rewritten as

$$X_{n+1} = \theta^{\,t}\Psi_n + U_n + \varepsilon_{n+1}, \tag{1.2}$$

*Received by the editors June 18, 2007; accepted for publication (in revised form) April 28, 2008; published electronically September 8, 2008.
http://www.siam.org/journals/sicon/47-5/69473.html
†Institut de Mathématiques de Bordeaux, Université Bordeaux 1, UMR 5251 and INRIA Bordeaux Sud-Ouest, Team CQFD, 351 cours de la libération, 33405 Talence cedex, France (Bernard.Bercu@math.u-bordeaux1.fr).
‡Départment de Génie Mathématiques, Laboratoire de Mathématique, INSA de Rouen, LMI-EA 3226, place Emile Blondel, BP 08, 76131 Mont-Saint-Aignan cedex, France (Bruno.Portier@insa-rouen.fr).

where

$$\Psi_n^t = \left(X_n^t, \ldots, X_{n-p+1}^t, U_{n-1}^t, \ldots, U_{n-q+1}^t, \varepsilon_n^t, \ldots, \varepsilon_{n-r+1}^t\right).$$

The most common way for estimating θ is to make use of the extended least-squares (ELS) algorithm given, for all $n \geq 0$, by

$$\widehat{\theta}_{n+1} = \widehat{\theta}_n + S_n^{-1}\Phi_n \left(X_{n+1} - U_n - \widehat{\theta}_n^t \Phi_n\right)^t,$$
$$\widehat{\varepsilon}_{n+1} = X_{n+1} - U_n - \widehat{\theta}_n^t \Phi_n,$$
$$\Phi_n^t = \left(X_n^t, \ldots, X_{n-p+1}^t, U_{n-1}^t, \ldots, U_{n-q+1}^t, \widehat{\varepsilon}_n^t, \ldots, \widehat{\varepsilon}_{n-r+1}^t\right),$$

where the initial value $\widehat{\theta}_0$ may be arbitrarily chosen. Moreover,

$$S_n = \sum_{i=0}^{n} \Phi_i \Phi_i^t + S,$$

where S is a positive definite and deterministic matrix introduced in order to avoid useless invertibility assumption. The crucial role played by the control 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], the adaptive tracking control U_n is given, for all $n \geq 0$, by

$$U_n = x_{n+1}^* - \widehat{\theta}_n^{\,t}\, \Phi_n. \tag{1.3}$$

By substituting (1.3) into (1.2), we obtain the closed-loop system

$$X_{n+1} - x_{n+1}^* = \pi_n + \varepsilon_{n+1}, \tag{1.4}$$

where

$$\pi_n = \theta^{\,t} \Psi_n - \widehat{\theta}_n^t \Phi_n$$

is the prediction error at time n. In the following, we shall assume that the driven noise (ε_n) is a sequence of centered independent and identically distributed random vectors with positive definite covariance matrix Γ and unknown probability density function denoted by f.

The purpose of this paper is to study the asymptotic properties of a kernel density estimator (KDE) of f. Since the pioneer works of Parzen [18] and Rosenblatt [23], the asymptotic properties of such a kernel estimator have been widely investigated in the context of independent and identically distributed random variables as well as for mixing random variables. We refer the reader to [10], [11], [24] for some excellent books on density estimation for stationary processes. Although the stability of ARMAX models in adaptive tracking has been deeply investigated in the literature [9], [12], one can realize that kernel density estimation results are not available in adaptive tracking.

Let us now define our KDE of f associated with model (1.2). When the sequence (ε_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 (ε_n) is of course unobservable. However, when the tracking objective is fullfilled, the prediction error π_n is as close as possible to zero. Consequently, via (1.4), we can choose $X_n - x_n^*$ as a predictor of ε_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

$$\widehat{f}_n(x) = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{h_i^d}K\left(\frac{X_i - x_i^* - x}{h_i}\right). \tag{1.5}$$

Our purpose is first to show that $\widehat{f}_n$ behaves pretty well as a RKDE of f in adaptive tracking and second to carry out a goodness-of-fit test for f based on $\widehat{f}_n$. Such a goodness-of-fit test is very popular in time series, in particular, for testing the normality hypothesis. For independent and identically distributed samples, we can mention the well-known and very popular Kolmogorov–Smirnov and Cramér–Von Mises statistical tests based on the empirical distribution function as well as the Bickel and Rosenblatt test [7] based on a KDE. Recently, for stationary autoregressive processes, several authors have proposed goodness-of-fit tests based on KDE [2], [17]. However, to the best of our knowledge, no work is concerned with asymptotic properties of KDE in adaptive tracking.

The paper is organized as follows. Section 2 is devoted to the asymptotic behavior of $\widehat{f}_n$. We establish the almost sure pointwise and uniform convergence of $\widehat{f}_n$ 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 the following, we shall assume that the kernel K is a non-negative function, bounded with compact support, such that

$$\int_{\mathbb{R}^d} K(t)\mathrm{d}t = 1 \qquad \text{and} \qquad \int_{\mathbb{R}^d} K^2(t)\mathrm{d}t = \tau^2.$$

For example, for some $s > 0$ and some known positive constants a_s, b_s, c_s, one can make use of the uniform kernel on the sphere of $\mathbb{R}^d$ with radius s, $K(t) = a_s \mathbb{1}_{(\|t\| \leq s)}$, the Epanechnikov kernel with scaling factor s, $K(t) = b_s\left(1 - \|t\|^2/s^2\right)\mathbb{1}_{(\|t\| \leq s)}$, and the Gaussian kernel with truncation level s, $K(t) = c_s \exp(-\|t\|^2/2)\mathbb{1}_{(\|t\| \leq s)}$.

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

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

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

Furthermore, we shall also make use of the classical assumptions of causality and passivity as well as the traditional smoothness hypothesis on the probability density function f.

Causality [A1]. *For all* $z \in \mathbb{C}$ *with* $|z| \leq 1$, $\det(z^{-1}B(z)) \neq 0$.
Passivity [A2]. *For all* $z \in \mathbb{C}$ *with* $|z| = 1$, $\det(C(z)) \neq 0$ *and* $C^{-1}(z) > \frac{1}{2}I_d$.
Density [A3]. *The function* f *is positive and differentiable with bounded gradient.*
We shall now propose 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] *to* [A3] *hold and suppose that* (ε_n) *has finite moment of order* $a > 2$. *In addition, assume 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* $f(x)$. *As soon as the bandwidth* (h_n) *satisfies* $\max(nh_n^{d+2}, n^b h_n^d) = o(\log \log n)$ *for some* $b \in]2/a, 1[$, *we also have*

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

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* $\beta \in](1 + c)/2, 1[$ *with* $c = \max(b, \alpha d)$,

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

Proof. The proof is given in Appendix A. □

Remark 1. The bandwidth condition associated with the almost sure pointwise convergence is clearly not restrictive and it is satisfied when $h_n = n^{-\alpha}$ with $\alpha \in]0, 1/d[$. In this particular case, the bandwidth condition required for the LIL is obviously satisfied as soon as $\alpha \in]\delta, 1/d[$ with $\delta = \max(1/(d + 2), b/d)$.

Remark 2. In the particular case of controlled autoregressive process

$$X_{n+1} = A_1 X_n + \cdots + A_p X_{n-p+1} + U_n + \varepsilon_{n+1}, \tag{2.3}$$

the assumptions [A1] and [A2] are clearly useless and the associated prediction errors sequence (π_n) satisfies (see, e.g., [5])

$$\sum_{i=0}^{n} \| \pi_i \|^2 = O(\log n) \qquad \text{a.s.} \tag{2.4}$$

Thanks to this sharp result on the sequence (π_n), we only have to assume that $\max(nh_n^{d+2}, h_n^d \log n) = o(\log\log n)$ for the LIL. This bandwidth condition is immediately satisfied when $h_n = n^{-\alpha}$ with $\alpha \in]1/(d + 2), 1/d[$. Moreover, for the uniform convergence, it is only necessary to assume that $\beta \in](1 + \alpha d)/2, 1[$. All of the above is also true for the scalar nonlinear controlled autoregressive process

$$X_{n+1} = \theta \, \varphi(X_n, \ldots, X_{n-p+1}) + U_n + \varepsilon_{n+1} \tag{2.5}$$

under suitable moment assumption on (ε_n) and as soon as the function $\varphi : \mathbb{R}^p \to \mathbb{R}$ does not increase to infinity faster than a polynomial of degree < 4 [6]. We are again able to deduce such results because the associated prediction errors sequence (π_n) satisfies (2.4). Finally, Theorem 2.1 holds for the RKDE associated with nonlinear controlled autoregressive processes as soon as the associated prediction errors sequence (π_n) satisfies a stability property such as (2.4).

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

THEOREM 2.2. *Assume that* [A1] *to* [A3] *hold and suppose that* (ε_n) *has finite moment of order* $a > 2$. *Moreover, assume that the bandwidth* (h_n) *satisfies* $\max(nh_n^{d+2}, n^b h_n^d) = o(1)$ *for some* $b \in]2/a, 1[$, *together with*

$$\lim_{n\to\infty} \frac{h_n^d}{n} \sum_{i=1}^{n} h_i^{-d} = \ell_h \tag{2.6}$$

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

$$G_n(x) = \sqrt{n\, h_n^d}\left(\widehat{f}_n(x) - f(x)\right) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, \tau^2 \ell_h f(x)\right) = G(x). \tag{2.7}$$

In addition, for N *distinct points* $x_1, \ldots, x_N$ *of* $\mathbb{R}^d$, *we also have*

$$(G_n(x_1), \ldots, G_n(x_N)) \xrightarrow{\mathcal{L}} (G(x_1), \ldots, G(x_N)), \tag{2.8}$$

where $G(x_1), \ldots, G(x_N)$ *are independent Gaussian random variables.*

Proof. The proof is given in Appendix B. □

Remark 3. Convergence (2.7) is identical to the one obtained by Duflo [12] for stationary processes. Besides, it is worthless to require the bandwidth condition (2.6) for the nonrecursive KDE of f, and ℓ_h has to be replaced by 1 in (2.7). Finally, if $h_n = n^{-\alpha}$, it is necessary to assume that $\alpha \in]\delta, 1/d[$ with $\delta = \max(1/(d+2), b/d)$ and we obviously have $\ell_h = (1+\alpha d)^{-1}$. In addition, for the controlled autoregressive processes given by (2.3) or (2.5), we only have to assume that $\alpha \in]1/(d+2), 1/d[$.

Remark 4. When the density function f belongs to $C^2(\mathbb{R}^d)$ with a bounded second derivative and for symmetric kernel K, we can relax the bandwidth condition by $\max(nh_n^{d+4}, n^b h_n^d) = o(1)$.

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

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

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 since the pioneer works of Kolmogorov–Smirnov and Cramér–Von Mises. Indeed, many statistical procedures require the assumption of normality for the driven white noise (see, e.g., [3] or [8]). Consequently, a goodness-of-fit test for the white noise density is of particular interest. 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 to provide a goodness-of-fit test for f based on the RKDE $\widehat{f}_n$. Such an approach has been already used by Bickel and Rosenblatt [7]. Indeed, for the independent and identically distributed sample, they 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 scalar autoregressive framework by Lee and Na [17] and more recently by Bachmann and Dette [2]. However, due to some technical reasons, it seems impossible to extend this approach to our adaptive tracking context. Therefore, we propose a new strategy and we carry out a goodness-of-fit test for f based on the multivariate CLT for $\widetilde{f}_n$ together with the LIL. Our statistical test consists of a suitably normalized sum of the quadratic deviation between the true density and the RKDE $\widehat{f}_n$ evaluated on N distinct points of $\mathbb{R}^d$. More precisely, it is defined by

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

where $x_1, \ldots, x_N$ are N distinct points of $\mathbb{R}^d$. We shall make use of

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

THEOREM 3.1. *Assume that* [A1] *to* [A3] *hold and suppose that* (ε_n) *has finite moment of order* $a > 2$. *Moreover, assume 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* $\mathcal{H}_0$,

$$nh_n^d \, T_n(N) \xrightarrow{\mathcal{L}} \chi^2(N). \tag{3.1}$$

Moreover, under $\mathcal{H}_1$ *and if one can find* $x \in \{x_1, x_2, \ldots, x_N\}$ *such that* $f(x) \neq f_0(x)$, *then* $T_n(N)$ *converges a.s. towards* σ^2. *In addition, we also have*

$$\sqrt{n\, h_n^d}\, \left(T_n(N) - \sigma^2\right) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, \lambda^2\right). \tag{3.2}$$

Remark 5. 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 $nh_n^d \, T_n(N)$ by a $\chi^2(N)$ one. On the other hand, under the alternative hypothesis $\mathcal{H}_1$, if σ^2 is positive, $nh_n^d \, T_n(N)$ 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 δ whenever $nh_n^d \, T_n(N) > a_\delta$ where a_δ stands for the $(1-\delta)$ quantile of the $\chi^2(N)$ 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 6. It is also possible to make use of the test statistic $Z_n(N)$ defined by

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

In that case, Theorem 3.1 holds with

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

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 $T_n(N)$ because it allows one 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

$$T_n(N) - \sigma^2 = A_n + B_n, \tag{3.3}$$

where

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

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

We can deduce from (2.8) and the pointwise almost sure convergence of $\widehat{f}_n$ to f that

$$\sqrt{\frac{n\, h_n^d}{\tau^2 \ell_h}} \left(\frac{\widehat{f}_n(x_1) - f(x_1)}{\sqrt{\widehat{f}_n(x_1)}}, \dots, \frac{\widehat{f}_n(x_N) - f(x_N)}{\sqrt{\widehat{f}_n(x_N)}} \right) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, I_N\right), \tag{3.4}$$

where I_N stands for the identity matrix of order N. Hence, it immediately follows from (3.4) that

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

Consequently, we clearly obtain (3.1) from (3.3) together with (3.5) since, under the null hypothesis $\mathcal{H}_0$, σ^2 and B_n vanish. Under the alternative hypothesis $\mathcal{H}_1$, it is straighforward to see that $T_n(N)$ converges a.s. towards σ^2 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

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

which implies that

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

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

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

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. ∎

4. Simulation experiments. In this section, we investigate the finite sample properties of our statistical test under both hypotheses $\mathcal{H}_0$ and $\mathcal{H}_1$ without some bootstrap procedure as is usual in this context of nonparametric tests. Since it has never been experimented, we shall not restrict ourselves to models of form (1.2), but

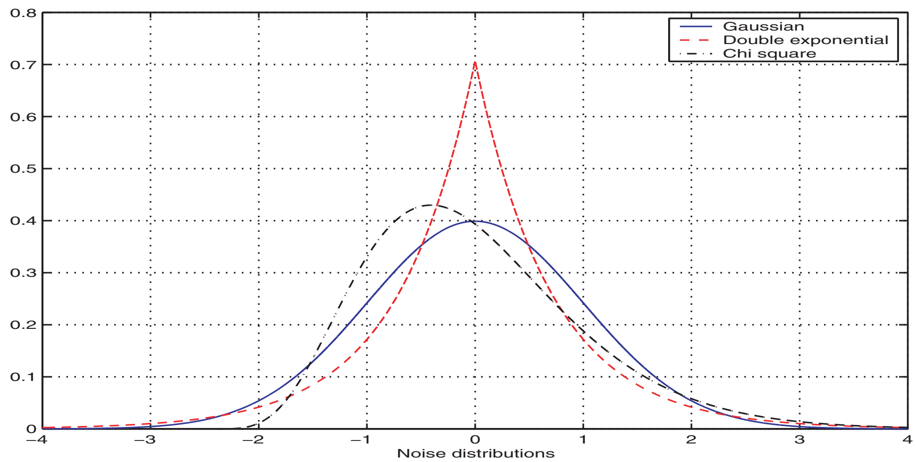

FIG. 4.1.

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.

$$\text{(WN)} \qquad X_n = \varepsilon_n,$$

$$\text{(AR)} \qquad X_{n+1} = \theta X_n + \varepsilon_{n+1},$$

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

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

where (ε_n) is a sequence of centered independent and identically distributed random variables 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, given in Figure 4.1, that we combine two-by-two in order to study the performances of our statistical test under both $\mathcal{H}_0$ 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

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

For AR, ARX, and NARX models, we estimate the unknown parameter θ by use of the standard least-squares estimator $\widehat{\theta}_n$. For the AR model, the probability density function f is estimated using the RKDE given by (1.5) where $X_n - x_n^*$ is replaced by $X_n - \widehat{\theta}_n X_{n-1}$. For ARX and NARX models, the adaptive control U_n is given by $U_n = -\widehat{\theta}_n X_n$ and $U_n = -\widehat{\theta}_n X_n^2$, respectively.

For each model and each test of $\mathcal{H}_0$ against $\mathcal{H}_1$, we base our estimations on 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 $T_n(N)$ requires the choice of design points together with the specification of a bandwidth and a kernel for the RKDE $\widehat{f}_n$. The RKDE $\widehat{f}_n$ is constructed by use of the Epanechnikov kernel

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

and the bandwidth $h_n = n^{-1/3}$. For the denominator of $T_n(N)$, we use the Gaussian kernel and the usual bandwidth $h_n = n^{-1/5}$. 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.

For the choice of N and the points $x_1, \cdots, x_N$, we use the design points selection rule proposed by Poggi and Portier and fully described in [19]. More precisely, we proceed as follows. Starting from an estimate of the distribution of the driven noise, we choose N equidistant points $x_1 \cdots, x_N$ 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. Typically, we choose points $x_1, \cdots, x_N$ such that the distance between two neighboring points is $4n^{-1/3}$. This last condition allows us to make sure that the independence property in the multivariate CLT, which holds asymptotically, remains true for small to moderate sample sizes. We take $N = 8, 13$, and 22 equidistant points for sample sizes $n = 200$, 500, and 1000, respectively. It should be noted that only a few number of points is needed to make a decision.

In the sequel, the abbreviations $\mathcal{G}f_0$, $\mathcal{G}f_1$, and $\mathcal{G}f_2$ mean that the driven noise (ε_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 $\mathcal{H}_0 : \langle\langle\, f = f_0 \,\rangle\rangle$, $\mathcal{H}_0 : \langle\langle\, f = f_1 \,\rangle\rangle$, and $\mathcal{H}_0 : \langle\langle\, f = f_2 \,\rangle\rangle$, respectively. Finally, as we have chosen a test level $\alpha = 5\%$ and we have generated 800 trials, the Kolmogorov–Smirnov fitting statistic in italic has to be compared with the critical value 0.048.

We shall now comment on the test results contained in Tables 4.1–4.4. 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 $n^{2/3}T_n(N)$ is close to the $\chi^2(N)$ 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 superimposes 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 sizes and for a large class of models.

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.*

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

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.*

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

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.*

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

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

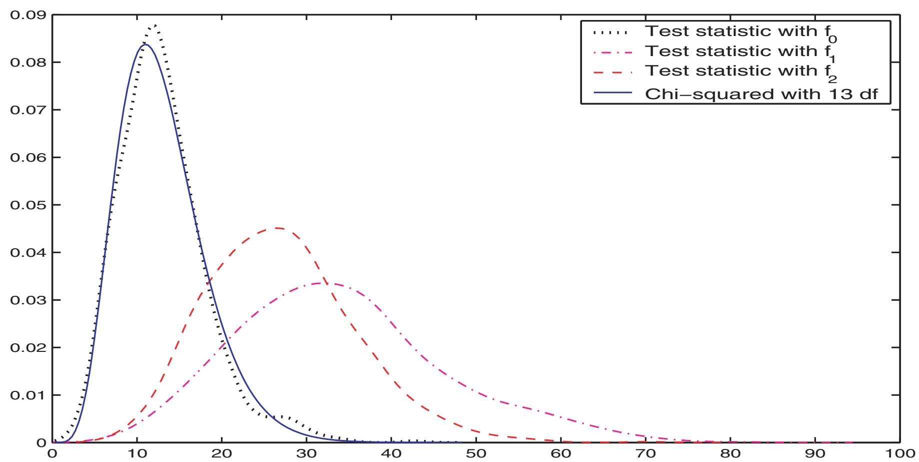

Fig. 4.2.

Figure 4.2 illustrates the empirical level and power of our test for the NARX model. We base our estimation on 800 trials of sample size $n = 500$ with $N = 13$ equidistant points. The driven noise (ε_n) is generated with the normal distribution f_0, and we are successively testing the assumptions $\mathcal{H}f_0$, $\mathcal{H}f_1$, and $\mathcal{H}f_2$. On the one hand, when we test the hypothesis $\mathcal{H}f_0$, we can observe that the distribution of our statistical test $n^{2/3}T_n(N)$ is superimposed with the $\chi^2(N)$ one. It clearly illustrates the good approximation of the distribution of $n^{2/3}T_n(N)$ by a $\chi^2(N)$ one under $\mathcal{H}f_0$ for moderate sample size. On the other hand, when we test the hypothesis $\mathcal{H}f_1$ as well as $\mathcal{H}f_2$, we can effectively see that the distribution of our statistical test $n^{2/3}T_n(N)$ is totally different from the $\chi^2(N)$ one. Consequently, the power of 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 infer from (1.5) that for all $x \in \mathbb{R}^d$ and $n \geq 1$,

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

with

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

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

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 σ-algebra of the events occuring 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,*

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

Proof. For any $x \in \mathbb{R}^d$, $(M_n(x))$ is a 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 $x \in \mathbb{R}^d$,

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

Consequently, we deduce from (1.4) that for all $x \in \mathbb{R}^d$

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

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

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

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 $\gamma > 0$,

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

which ensures that $M_n(x) = o(n)$ a.s. since nh_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). ∎

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*

$$\sup_{\|x\| \leq An^\gamma} |M_n(x)| = o\left(n^\beta\right) \qquad a.s., \tag{A.6}$$

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

Proof. Result (A.6) follows from 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_δ such that, for any $x, y \in \mathbb{R}^d$

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

Hence, for any $x, y \in \mathbb{R}^d$, we can derive that

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

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

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

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

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

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

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

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

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

Since the power δ can be chosen as small as one wishes, all four conditions of Theorem 6.4.34 of [12] are fullfilled which leads to Lemma A.2. ∎

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

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

via the change of variables $t = h_i^{-1}(\pi_{i-1} + s - x)$. As the density function f is differentiable with a bounded gradient, we obtain by a Taylor expansion that

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

Moreover, since (ε_n) has a finite moment of order $a > 2$, we deduce from [A1] and [A2] together with Theorem 1 of [13] that

$$\sum_{i=1}^{n} \| \pi_{i-1} \|^2 = O\left(n^b\right) \qquad \text{a.s.} \tag{A.9}$$

for all $b \in]2/a, 1[$. Hence, it follows from (A.9) together with the Cauchy–Schwarz inequality that

$$\sup_{x \in \mathbb{R}^d} |R_n(x)| = O(nh_n) + O\left(\sqrt{n^{1+b}}\right) \quad \text{a.s.} \tag{A.10}$$

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.10). The uniform, almost sure convergence on $\mathbb{R}^d$ still remains to be proven. Hereafter, we take $h_n = n^{-\alpha}$ with $\alpha \in]0, 1/d[$. On the one hand, we find from Lemma A.2 with $A = 2$ and $\gamma = 1/2$, that

$$\sup_{\|x\| \leq 2\sqrt{n}} |M_n(x)| = o\left(n^{\beta}\right) \quad \text{a.s.,} \tag{A.11}$$

where $\beta \in](1+\alpha d)/2, 1[$. From now on, we choose $\beta \in](1+c)/2, 1[$ with $c = \max(b, \alpha d)$. Since $\beta > (1+b)/2$, it implies that $n^{1+b} = o(n^{2\beta})$. Hence, it follows from the conjunction of (A.10) and (A.11) that

$$\sup_{\|x\| \leq 2\sqrt{n}} \left|\widehat{f}_n(x) - f(x)\right| = O\left(n^{-\alpha}\right) + o\left(n^{\beta-1}\right) \quad \text{a.s.} \tag{A.12}$$

On the other hand, we claim that

$$\sup_{\|x\| > 2\sqrt{n}} \left|\widehat{f}_n(x) - f(x)\right| = O\left(\frac{1}{n}\right) \quad \text{a.s.} \tag{A.13}$$

As a matter of fact, since (ε_n) has a finite moment of order $a > 2$, we infer from Lemma 2 of [13] that $\| X_n \|^2 = O\left(n^b\right)$ a.s. for some $b \in]2/a, 1[$, which implies that

$$\sup_{i \leq n} \| X_i - x_i^* \|^2 = o(n) \quad \text{a.s.}$$

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

$$\sup_{\|x\| > 2\sqrt{n}} \left|n\widehat{f}_n(x)\right| = \sup_{\|x\| > 2\sqrt{n}} \left|\sum_{i=1}^{n} K_i\left(X_i - x_i^* - x\right)\right| = O(1) \quad \text{a.s.} \tag{A.14}$$

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

$$\sup_{\|x\| > 2\sqrt{n}} f(x) = O\left(\frac{1}{n}\right). \tag{A.15}$$

Consequently, we obtain (A.13) from (A.14) and (A.15). Finally, we deduce (2.2) from (A.12) and (A.13), 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] *to* [A3] *hold and suppose that* (ε_n) *has a finite moment of order* $a > 2$. *Moreover, assume that the bandwidth* (h_n) *shares the same assumptions as in Theorem* 2.2. *Then, for any* $x \in \mathbb{R}^d$,

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

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

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

which implies that

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

Moreover, we also have

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

Consequently, we obtain the decomposition

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

where

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

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

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

Hence, it follows from (A.9) that

$$|B_n| = O\Big(h_n^{-d}\sqrt{n^{1+b}}\Big) \qquad \text{a.s.}$$

for all $b \in]2/a, 1[$. Furthermore, we immediately get from (2.6) that $n^{-1}h_n^d C_n$ converges to ℓ_h as n goes to infinity. Putting together those three contributions, we find that

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

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 $a > 0$ and $x \in \mathbb{R}^d$, let

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

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 $i \leq n$

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

Consequently, we find that for all $a > 0$ and $x \in \mathbb{R}^d$,

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

Therefore, as nh_n^d tends to infinity, we can deduce that, for all $a > 0$ and $x \in \mathbb{R}^d$, $\Lambda_n(a,x)$ tends to zero a.s. Finally, Lindeberg's condition is satisfied, which achieves the proof of Lemma B.1. □

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

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

Consequently, (2.7) immediately follows from (A.10) together with (B.1) and (B.4) as soon as $\max(nh_n^{d+2}, n^b h_n^d) = o(1)$. The multivariate CLT remains to be proven. Taking the previous results into account, it is enough 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 two independent Gaussian random variables. We can easily show this convergence by remarking that for two distinct points $x, y \in \mathbb{R}^d$

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

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

$$\mathbb{E}\left[\Delta M_i(x)\Delta M_i(y)|\mathcal{F}_{i-1}\right] \leq \mathbb{E}\left[K_i\left(X_i - x_i - x\right)K_i\left(X_i - x_i - y\right)|\mathcal{F}_{i-1}\right],$$

$$\leq \mathbb{E}\left[K_i\left(\pi_{i-1} + \varepsilon_i - x\right)K_i\left(\pi_{i-1} + \varepsilon_i - y\right)|\mathcal{F}_{i-1}\right],$$

which implies that

$$\mathbb{E}\left[\Delta M_i(x)\Delta M_i(y)|\mathcal{F}_{i-1}\right] \leq h_i^{-d}\int_{\mathbb{R}^d} K(t)\,K\left(t + h_i^{-1}(x-y)\right)f\left(h_i t + x - \pi_{i-1}\right)\mathrm{d}t.$$

Therefore, as the gradient of f is bounded, we obtain from (A.9) that

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

for all $b \in]2/a, 1[$, where

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

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. ∎

Remark 7. Result (B.5) ensures the asymptotic independence of the random variables $G_n(x_1), \ldots, G_n(x_N)$ in the multivariate CLT. Since the kernel K is compactly supported, for finite values of n, the left-hand side of (B.5) can be very small if we choose two points x and y sufficiently distant. This last point clarifies the design points selection rule described in section 4.

REFERENCES

[1] K. J. Åström and B. Wittenmark, *Adaptive Control*, 2nd ed., Addison-Wesley, New York, 1995.

[2] D. Bachmann and H. Dette, *A note on the Bickel-Rosenblatt test in autoregressive time series*, Statist. Probab. Lett., 74 (2005), pp. 221–234.

[3] M. Basseville and I. V. Nikiforov, *Detection of Abrupt Changes—Theory and Application*, Prentice Hall, Englewood Cliffs, NJ, 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*, Statist. Probab. Lett., 56 (2002), pp. 23–35.
[18] E. Parzen, *On estimation of a probability density function and mode*, Ann. Math. Stat., 33 (1962), pp. 1065–1076.
[19] J. M. Poggi and B. Portier, *A test of linearity for functional autoregressive models*, J. Time Ser. Anal., 18 (1997), pp. 615–640.
[20] J.-M. Poggi and B. Portier, *Nonlinear adaptive tracking using kernel estimators: estimation and test for linearity*, SIAM J. Control Optim., 39 (2000), pp. 707–727.
[21] B. Portier, *Adaptive control of discrete-time nonlinear systems combining nonparametric and parametric estimators*, Commun. Inf. Syst., 2 (2002), pp. 69–90.
[22] B. Portier and A. Oulidi, *Nonparametric estimation and adaptive control of functional autoregressive models*, SIAM J. Control Optim., 39 (2000), pp. 411–432.
[23] M. Rosenblatt, *Remarks on some nonparametric estimates of a density function*, Ann. Math. Stat., 27 (1956), pp. 832–837.
[24] B. W. Silverman, *Density Estimation for Statistics and Data Analysis*, Chapman & Hall, New York, 1986.
[25] L-L. Xie and L. Guo, *How much uncertainty can be dealt with by feedback?*, IEEE Trans. Automat. Control, 45 (2000), pp. 2203–2217.