Proceedings of the
47th IEEE Conference on Decision and Control
Cancun, Mexico, Dec. 9-11, 2008

Kernel density estimation 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 multivariate ARMAX models in adaptive tracking. We establish 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.**

I. 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 [3], [4], [5] [8], [12], [13], [14], [15]. However, only few references may be found on nonparametric estimation in adaptive tracking [19], [20], [21],[24]. 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

$$A(R)X_n = B(R)U_n + C(R)\varepsilon_n, \tag{1}$$

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) can be rewritten as

$$X_{n+1} = \theta^{\,t}\Psi_n + U_n + \varepsilon_{n+1}, \tag{2}$$

where $\Psi_n = (X_n^p, U_n^q, \varepsilon_n^r)^t$ with

$$\begin{aligned} X_n^p &= (X_n^t, \ldots, X_{n-p+1}^t), \\ U_n^q &= (U_{n-1}^t, \ldots, U_{n-q+1}^t), \\ \varepsilon_n^r &= (\varepsilon_n^t, \ldots, \varepsilon_{n-r+1}^t). \end{aligned}$$

B. Bercu is with Université Bordeaux 1, Institut de Mathématiques de Bordeaux, UMR 5251, 351 cours de la libération, 33405 Talence cedex, France Bernard.Bercu@math.u-bordeaux1.fr

B.Portier is with INSA de Rouen, Département Génie Mathématiques Place Emile Blondel, 76131 Mont-Saint-Aignan cedex, France Bruno.Portier@insa-rouen.fr

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

$$\widehat{\theta}_{n+1} = \widehat{\theta}_n + S_n^{-1}\Phi_n(X_{n+1} - U_n - \widehat{\theta}_n^{\,t}\Phi_n)^t,$$
$$\widehat{\varepsilon}_{n+1} = X_{n+1} - U_n - \widehat{\theta}_n^{\,t}\Phi_n,$$

where the initial value $\widehat{\theta}_0$ is arbitrarily chosen, the vector $\Phi_n = (X_n^p, U_n^q, \widehat{\varepsilon}_n^r)^t$ with $\widehat{\varepsilon}_n^r = (\widehat{\varepsilon}_n^t, \ldots, \widehat{\varepsilon}_{n-r+1}^t)$ and

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

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

$$U_n = x_{n+1}^* - \widehat{\theta}_n^{\,t}\,\Phi_n. \tag{3}$$

By substituting (3) into (2), we obtain the closed-loop system

$$X_{n+1} - x_{n+1}^* = \pi_n + \varepsilon_{n+1}, \tag{4}$$

where

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

In all the sequel, 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. Our purpose is to study the asymptotic properties of a recursive kernel density estimator (RKDE) of f given, for all $x \in \mathbb{R}^d$ and $n \geq 1$, by

$$\widehat{f}_n(x) = \frac{1}{n}\sum_{i=1}^{n} \frac{1}{h_i^d}\, K\left(\frac{X_i - x_i^* - x}{h_i}\right), \tag{5}$$

where the kernel K is a chosen density function and the bandwidth (h_n) is a sequence of positive real numbers decreasing to zero. Since the pioneer works of Parzen [17] and Rosenblatt [22], 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 [9], [10], [23] 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 [8], [11], one can realize that kernel density estimation results are not available in adaptive tracking.

978-1-4244-3124-3/08/$25.00 ©2008 IEEE

Our purpose is to 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). A goodness-of-fit test for f based on $\widehat{f}_n$ is also provided together with some simulation experiments.

II. ON THE KERNEL

We shall now propose three different choices for the kernel function K. The kernel K is a nonnegative function, bounded with compact support, satisfying

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

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,

$$K(t) = a_s \, \mathbb{I}_{(\|t\| \leq s)},$$

the Epanechnikov kernel with scaling factor s,

$$K(t) = b_s \left(1 - \| t \|^2/s^2\right) \mathbb{I}_{(\|t\| \leq s)},$$

and the Gaussian kernel with truncation level s,

$$K(t) = c_s \exp\left(- \| t \|^2/2\right) \mathbb{I}_{(\|t\| \leq s)}.$$

III. MAIN RESULTS

First of all, we shall make use of the classical assumptions of causality and passivity and 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 present several asymptotic results for the RKDE $\widehat{f}_n$ of f recently obtained by Bercu and Portier [6], the first one dealing with the almost sure convergence properties of $\widehat{f}_n$.

Theorem 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 $(\log n)^2$ and

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

Then, for any $x \in \mathbb{R}^d$, $\widehat{f}_n(x)$ converges a.s. to $f(x)$. If 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

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

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 $c = \max(b, \alpha d)$, we have the almost sure uniform rates of convergence

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

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 satisfied if $\alpha \in]\delta, 1/d[$ with $\delta = \max(1/(d + 2), b/d)$.

Remark 2: In the particular case of controlled autoregressive process

$$X_{n+1} = A_1 X_n + \cdots + A_p X_{n-p+1} + U_n + \varepsilon_{n+1}, \quad (6)$$

the assumptions [*A1*] and [*A2*] are clearly useless and the associated prediction errors sequence (π_n) satisfies

$$\sum_{i=0}^{n} \| \pi_i \|^2 = O(\log n) \quad \text{a.s.} \quad (7)$$

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

$$X_{n+1} = \theta \, \varphi(X_n, \ldots, X_{n-p+1}) + U_n + \varepsilon_{n+1} \quad (8)$$

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, see [5]. One more time, we are able to deduce such results because the associated prediction errors sequence (π_n) satisfies (7).

The second result is related to the CLT for $\widehat{f}_n$.

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

$$\lim_{n\to\infty} \frac{h_n^d}{n} \sum_{i=1}^{n} h_i^{-d} = \ell_h \quad (9)$$

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

$$\sqrt{n\, h_n^d}\Big(\widehat{f}_n(x) - f(x)\Big) \xrightarrow{\mathcal{L}} \mathcal{N}\Big(0, \tau^2 \ell_h f(x)\Big). \quad (10)$$

In addition, let $x_1, \ldots, x_N$ be N distinct points of $\mathbb{R}^d$ and denote $G_n(x_i) = \sqrt{n\, h_n^d}(\widehat{f}_n(x_i) - f(x_i))$ and its limit $G(x_i) = \mathcal{N}(0, \tau^2 \ell_h f(x_i))$. Then, we also have

$$(G_n(x_1), \ldots, G_n(x_N)) \xrightarrow{\mathcal{L}} (G(x_1), \ldots, G(x_N)) \quad (11)$$

where $G(x_1), \ldots, G(x_N)$ are independent.

Remark 3: Convergence (10) is identical to the one obtained by Duflo [11] for stationary processes. Besides, it is worthless to require the bandwidth condition (9) for the nonrecursive KDE of f, and ℓ_h has to be replaced by 1 in (10). 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 (6) or (8), 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 $\max(nh_n^{d+4}, n^b h_n^d) = o(1)$.

IV. Application to a goodness-of-fit test

We shall now propose a statistical test associated with the probability density function f based on our convergence results for $\widehat{f}_n$. We wish to test

$$\mathcal{H}_0 \,:\, \text{«} f = f_0 \text{»} \qquad \text{vs} \qquad \mathcal{H}_1 \,:\, \text{«} f \neq f_0 \text{»}$$

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. 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 an independant 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 [16] 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 $\widehat{f}_n$ 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

$$\sigma^2 = \frac{1}{\tau^2 \ell_h} \sum_{j=1}^{N} \frac{(f(x_j) - f_0(x_j))^2}{f(x_j)},$$

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

Theorem 3: 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 Theorems 1 and 2. Then, under $\mathcal{H}_0$,

$$nh_n^d\, T_n(N) \quad \xrightarrow{\mathcal{L}} \quad \chi^2(N). \tag{12}$$

Moreover, under $\mathcal{H}_1$ and if one can find some point x of $\mathbb{R}^d$ 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

$$\sqrt{n\, h_n^d}\Big(T_n(N) - \sigma^2\Big) \quad \xrightarrow{\mathcal{L}} \quad \mathcal{N}\Big(0, \lambda^2\Big). \tag{13}$$

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 (13) 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 holds with

$$\sigma^2 = \frac{1}{\tau^2 \ell_h} \sum_{j=1}^{N} \frac{(f(x_j) - f_0(x_j))^2}{f_0(x_j)},$$

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

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.

V. SIMULATIONS

The goal of this section is illustrate our asymptotic results by simulations. We shall investigate the finite sample properties of our statistical test under both hypothesis $\mathcal{H}_0$ and $\mathcal{H}_1$ without some bootstrap procedure as it is usual in this context of nonparametric tests. Since it has never been experimented, we shall not restrict ourselves to models of form (2), 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.

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

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

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

where (ε_n) is a sequence of centered independent and identically distributed random variables with probability density function f. We have arbitrarily chosen the values $\theta = 7/10$, $\theta = 2$, and $\theta = 1/2$ for the AR, ARX and NARX models, respectively. We consider three choices of noise distributions, given in Figure 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

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

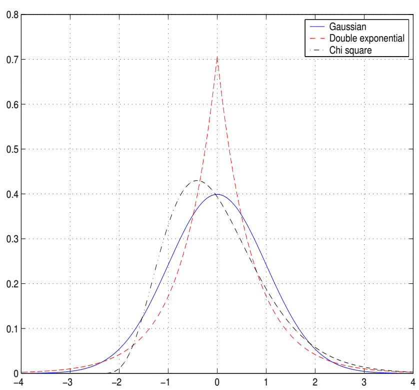

Fig. 1. Noise distributions

For AR, ARX, and NARX models, we estimate the unknown parameter θ by use of the standard least squares estimator $\widehat{\theta}_n$. For AR model, the probability density function f is estimated using the RKDE given by (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 $n = 200$ and $n = 1000$. 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

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

TABLE I

AR MODEL. RESULTS UNDER $\mathcal{H}_0$ AND $\mathcal{H}_1$ WITH TEST LEVEL 5%. EMPIRICAL LEVEL IN BOLD AND PERCENTAGE OF CORRECT DECISIONS.

	$n=200$, $N=8$			$n=1000$, $N=22$		
	$\mathcal{H}f_0$	$\mathcal{H}f_1$	$\mathcal{H}f_2$	$\mathcal{H}f_0$	$\mathcal{H}f_1$	$\mathcal{H}f_2$
$\mathcal{G}f_0$	**4.5%** *0.045*	31.2%	25.6%	**3.7%** *0.023*	99.8%	98.8%
$\mathcal{G}f_1$	49.7%	**5.7%** *0.032*	73.1%	100%	**4.8%** *0.019*	100%
$\mathcal{G}f_2$	19.3%	54.6%	**3.7%** *0.045*	96.6%	100%	**3.8%** *0.013*

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 [18]. 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 asymptotically holds, remains true for small to moderate sample sizes. We take $N = 8$ and $N = 22$ equidistant points for sample sizes $n = 200$ and $n = 1000$, respectively. It should be noted that only a few number of points is needed to make a decision.

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, while $\mathcal{H}f_0$, $\mathcal{H}f_1$, and $\mathcal{H}f_2$ mean that we are testing the assumptions $\mathcal{H}_0$: « $f = f_0$ », $\mathcal{H}_0$: « $f = f_1$ », and $\mathcal{H}_0$: « $f = f_2$ », 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 value 0.048.

We shall now comment on the test results contained in Tables I to III. 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

TABLE II

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.

	$n=200$, $N=8$			$n=1000$, $N=22$		
	$\mathcal{H}f_0$	$\mathcal{H}f_1$	$\mathcal{H}f_2$	$\mathcal{H}f_0$	$\mathcal{H}f_1$	$\mathcal{H}f_2$
$\mathcal{G}f_0$	**3.8%** *0.042*	35.7%	28%	**3.7%** *0.018*	99.7%	98.2%
$\mathcal{G}f_1$	45.8%	**5.5%** *0.053*	71.5%	100%	**5%** *0.022*	100%
$\mathcal{G}f_2$	21.2%	54.5%	**3.2%** *0.029*	96.7%	100%	**5.1%** *0.029*

TABLE III

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.

	$n=200$, $N=8$			$n=1000$, $N=22$		
	$\mathcal{H}f_0$	$\mathcal{H}f_1$	$\mathcal{H}f_2$	$\mathcal{H}f_0$	$\mathcal{H}f_1$	$\mathcal{H}f_2$
$\mathcal{G}f_0$	**3%** *0.037*	37.1%	28.5%	**4.3%** *0.037*	99.5%	98.6%
$\mathcal{G}f_1$	44.6%	**5.2%** *0.021*	72%	100%	**5.1%** *0.017*	100%
$\mathcal{G}f_2$	19.8%	58.3%	**3.7%** *0.021*	97.2%	100%	**5%** *0.039*

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 sizes and for a large class of models.

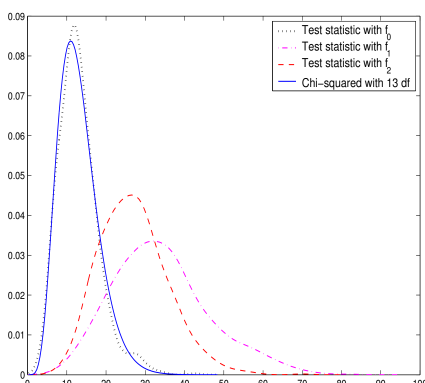

Fig. 2. Power of the goodness of fit test

Figure 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. Finally, the power of separation of our statistical test is clearly significant.

VI. CONCLUSION

We have carried out a sharp analysis of the convergence of a RKDE associated with the driven noise of multivariate ARMAX models in adaptive tracking. A goodness-of-fit test based on our RKDE was also provided. Surprisingly, one can realize that only few papers deal with nonparametric estimation in adaptive tracking. We hope that this contribution will be a starting point for further investigation on nonparametric estimation by the control community.

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] B. Bercu, Weighted estimation and tracking for ARMAX models, *SIAM J. Control Optim.*, 33, (1995), pp. 89-106.

[4] 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.

[5] 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.

[6] B. Bercu and B. Portier, Kernel density estimation and goodness-of-fit test in adaptive tracking, To appear in *SIAM J. Control Optim.*, 47, (2008).

[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] H. F. Chen and L. Guo, Identification and Stochastic Adaptive Control, *Birkhäuser*, Boston, 1991.

[9] L. Devroye, A Course in Density Estimation, *Birkhäuser*, Boston, 1987.

[10] L. Devroye and G. Lugosi, Combinatorial Methods in Density Estimation, *Springer Verlag*, New York, 2001.

[11] M. Duflo, Random Iterative Models, *Springer Verlag*, Berlin, 1997.

[12] 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.

[13] L. Guo, Further results on least squares based adaptive minimum variance control, *SIAM J. Control Optim.*, 32, (1994), pp. 187-212.

[14] L. Guo, Self convergence of weighted least squares with applications to stochastic adaptive control, *IEEE Trans. Automat. Control*, 41 (1996), pp. 79-89.

[15] 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.

[16] S. Lee and S. Na, On the Bickel-Rosenblatt test for first-order autoregressive models, *Stat. Prob. Letters*, 56, (2002), pp. 23-35.

[17] E. Parzen, On estimation of a probability density function and mode, *Annals of Math. Stat.*, 33, (1962), pp. 1065-1076.

[18] J.M. Poggi and B. Portier, A test of linearity for functional autoregressive models, *Journ. of Time Ser. Ana.*, 18, (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, (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, (2000), pp. 411-432.

[22] M. Rosenblatt, Remarks on some nonparametric estimates of a density function, *Annals of Math. Stat.*, 27, (1956), pp. 832-837.

[23] B. W. Silverman Density estimation for statistics and data analysis, *Chapman & Hall*, New York, 1986.

[24] 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.