Problems of Control and Information Theory, Vol. 8 (1), pp. 67–73 (1979)

NONPARAMETRIC SYSTEM IDENTIFICATION BY ORTHOGONAL SERIES

W. GREBLICKI

(Wrocław)

(Received February 20, 1977)

A nonparametric statistical method for memoryless system identification is presented. The method, based on the expansion in an infinite orthogonal series is asymptotically optimal. This means that a sequence of models converges to that one, which is optimal among all measurable models, i.e. converges to the regression of the output on the input. The square quality index converges to that of the optimal model.

1. Introduction

Almost all papers concerning statistical identification of memoryless systems assume that a model is described by finite number of parameters, see Aizerman [1], Bubnicki [2], Saridis [6], Tsypkin [8]. A special attention is given to the stochastic approximation, see [3], [6], [8]. The nonparametric approach is presented by Braverman [2], and the problem is treated by the method of potential functions. Another nonparametric method for identification by orthogonal series is given in this lecture and its basic asymptotic properties are shown.

2. Statement of the problem

The input of a memoryless system is a k-dimentional vector $\mathbf{x}$ which belongs to a set $\mathfrak{X} \subset \mathfrak{R}^k$, the output is a scalar $y \in \mathfrak{Y} = \mathfrak{R}^1$. Let μ be a σ-finite measure on the space $\mathfrak{R}^k$, and let $\mathfrak{X}$ be μ-measurable. Both the input and the output are random. We denote the input-output pair of random variables by (X, Y) and assume that $EY^2 < \infty$. Let $f(x, y)$ be its probability density function $f(y/x)$ is the conditional output density and $f(x)$ is the input density. Obviously, the conditional output density describes the system's properties.

Any μ-measurable function $\Phi(x)$ mapping $\mathfrak{X}$ into $\mathfrak{Y}$ will be called a model. It is known that the quality index

$$R(\Phi) = \iint (y - \Phi(x))^2 f(x, y)\, dx\, dy \tag{1}$$

is minimized by the optimal model

$$\Phi_0(x) = \int y f(y/x)\, dy\,. \tag{2}$$

Throughout this paper all integrals are taken over either the space $\mathcal{X}$ or the space $\mathcal{Y}$ or both of them, respectively.

Let us assume that the input density $f(x)$ is known, but the conditional output density $f(y/x)$ is completely unknown. The identification problem is to estimate the optimal model having a learning sequence, i.e. a sample of independent observations

$$(X_1, Y_1), \ldots, (X_n, Y_n)$$

of the pair (X, Y).

In this paper the problem is treated by orthogonal series. It is shown that the method is asymptotically optimal, i.e. when a number of observations tends to infinity, the sequence of models converges to the optimal model and the quality index converges to the minimal one.

3. Identification procedures

Let L_2 be the space of all μ-measurable and square integrable models, and let $\{\varphi_i(x)\}$ $(i = 0, 1, 2, \ldots)$ be a complete system of orthonormal functions of L_2. Of course, the optimal model is not necessarily square integrable. If, however, it is an element of L_2,

$$\Phi_0(x) = \sum_{i=0}^{\infty} a_i \varphi_i(x)\,,$$

where

$$a_i = \int \Phi_0(x)\, \varphi_i(x)\, dx = E\{Y\varphi_i(X)/f(X)\}\,.$$

All the coefficients of this expansion are estimated from the learning sequence according to the following formula

$$\hat{a}_{in} = n^{-1} \sum_{j=1}^{n} [Y_j\, \varphi_i(X_j)/f(X_j)]$$

and as a model we take

$$\hat{\Phi}_n(x) = \sum_{i=o}^{N(n)} \hat{a}_{in}\, \varphi_i(x)\,,$$

where $\{N(n)\}$ is a sequence of numbers.

We shall now give a theorem on the asymptotic optimality of this model.

Theorem 1. If $\Phi_0(x) \in L_2$,

$$EY^2 < \infty, \tag{4}$$

$$f(x) \geq \alpha > 0 \tag{5}$$

(α is independent of x), all the functions of the orthonormal system are jointly bounded, i.e.

$$|\varphi_i(x)| \leq c \tag{6}$$

(c is independent of i), and

$$\lim_{n\to\infty} N(n) = \infty, \quad \lim_{n\to\infty} N(n)/n = 0, \tag{7}$$

then

$$\lim_{n\to\infty} E \int (\Phi_0(x) - \hat{\Phi}_n(x))^2 \, dx = 0. \tag{8}$$

Proof. It is clear that the estimator (3) is unbiased and its variance is bounded by

$$E(a_i - \hat{a}_{in})^2 \leq n^{-1} E\{Y\varphi_i(X)/f(X)\}^2 \leq n^{-1}\alpha^{-2}c^2 EY^2 = dn^{-1}.$$

By Fubini's theorem and Parseval's formula we have

$$E \int (\Phi_0(x) - \hat{\Phi}_n(x))^2 \, dx = \sum_{i=0}^{N(n)} E(a_i - \hat{a}_{in})^2 + \sum_{i=N(n)+1}^{\infty} a_i^2 \leq$$

$$\leq dn^{-1} N(n) + \sum_{i=N(n)+1}^{\infty} a_i^2.$$

Convergence (8) is a consequence of the above inequality and (7). Thus, the theorem is proved.

As it has been shown, in order that the optimal model may be expanded, its square integrability must be assumed. This condition can be neglected while expanding $\Phi_0(x)f^{1/2}(x)$. It is square integrable, because of $EY^2 < \infty$. Thus,

$$\Phi_0(x)f^{1/2}(x) \sim \sum_{i=0} b_i \varphi_i(x),$$

where

$$b_i = \int \Phi_0(x)\, \varphi_i(x)\, f^{1/2}(x)\, dx = E\{Y\varphi_i(X)\, f^{-1/2}(X)\}.$$

Each b_i will be estimated by

$$\hat{b}_{in} = n^{-1} \sum_{j=1}^{n} [Y_j \varphi_i(X_j)\, f^{-1/2}(X_j)].$$

We assume that

$$\hat{\Phi}'_n(x) = f^{-1/2}(x) \sum_{i=0}^{N(n)} \hat{b}_{in}\, \varphi_i(x)$$

is a model.

Theorem 2. If conditions (4), (5), (6) and (7) are fulfilled, then

$$\lim_{n\to\infty} E \int (\Phi_0(x) - \hat{\Phi}'_n(x))^2 f(x)\, dx = 0. \tag{9}$$

The proof is similar to that of Theorem 1.

It is easy to see that for any model

$$R(\Phi) = R(\Phi_0) + \int (\Phi_0(x) - \Phi(x))^2 f(x)\, dx.$$

Therefore by (9) we get

Corollary. Under all the assumptions of Theorem 2

$$\lim_{n\to\infty} ER(\hat{\Phi}'_n) = R(\Phi_0).$$

From a practical viewpoint, (5) is the only essential assumption in Theorems 1 and 2. It confines us to input signals having a density bounded from zero. It is clear that it can be satisfied only if the space $\mathscr{X}$ has a finite measure. If, however $\Phi_0(x) f(x)$ is square integrable, (5) can be neglected. In this case

$$\Phi_0(x)f(x) \sim \sum_{i=0}^{\infty} c_i \varphi_i(x),$$

where

$$c_i = E\{Y\varphi_i(X)\}.$$

Let

$$\hat{\Phi}''_n(x) = f^{-1}(x) \sum_{i=0}^{N(n)} \hat{c}_{in}\, \varphi_i(x),$$

where

$$\hat{c}_{in} = n^{-1} \sum_{j=1}^{n} Y_j \varphi_i(X_j).$$

Theorem 3. If $\Phi_0(x)f(x) \in L_2$, and (4), (6), (7) hold, then

$$\lim_{n\to\infty} E \int ((\Phi_0(x) - \hat{\Phi}''_n(x))^2 f^2(x)\, dx = 0.$$

The theorem can be established in the same way as Theorem 1.

4. Examples

The method presented in this paper requires that all the functions of an orthonormal system be jointly bounded. Therefore not every complete orthonormal system is useful for our purposes.

If $\mathcal{X}$ is a real line, i.e. $\mathcal{X} = (-\infty, \infty)$, Hermite's system

$$\varphi_i(x) = 2^{-1/2}(i\,!)^{-1/2}\Pi^{-1/4}H_i(x),$$

where

$$H_i(x) = (-1)^i\, e^{x^2} \frac{d^i}{dx^i} e^{-x^2}$$

is the ith Hermite's polynomial, may be applied, since all of its functions are jointly bounded, see [5].

If $\mathcal{X} = [0, \infty)$ Laguerre's system

$$\varphi_i(x) = e^{-x/2}\, L_i(x),$$

where

$$L_i(x) = e^x\,(i\,!)^{-1} \frac{d^i}{dx^i}(e^{-x}x^i)$$

is the ith Laguerre's polynomial, is useful, since

$$|L_i(x)| \leq e^{x/2},$$

see [7], and consequently all its functions are bounded by 1.

If $\mathcal{X} = [-\Pi, \Pi]$, the trigonometric system

$$\frac{1}{\sqrt{\pi}},\ \frac{1}{\sqrt{2\pi}}\sin x,\ \frac{1}{\sqrt{2\pi}}\cos x,\ \ldots$$

may be employed.

On the other hand, a system

$$y = \psi(x) + z,$$

where $\psi(x)$ is a Borel function defined on $\mathcal{X} = \mathcal{R}^1$ is an example of those considered in this paper. The additive, random noise is independent of the input and has zero mean and the variance σ^2. It is clear that $\psi(x)$ is the optimal model, i.e. $\psi(x) = \Phi_0(x)$ and condition (4) is fulfilled iff

$$\int \psi^2(x)\, f(x)\, dx < \infty.$$

Thus, if the input density is bounded, $\psi(x)f(x)$ is square integrable and by Theorem 2 we get

$$\lim E \int (\psi(x) - \widehat{\Phi}'_n(x))^2 f(x)\, dx = 0,$$

and by the Corollary we have

$$\lim_{n\to\infty} E \iint (y - \widehat{\Phi}'_n(x))^2 f(x, y)\, dxdy = \sigma^2 .$$

5. Final remarks

It should be noticed that there exists an essential difference between parametric and nonparametric problems. As a model in the former one, one chooses e.g. a function of the type

$$\gamma_1\eta_1(x) + \ldots + \gamma_m\eta_m(x),$$

where $\eta_1(x), \ldots, \eta_m(x)$ are fixed functions, and m is fixed and finite. Such a model may be interpreted as a point in a finite dimensional space, whereas the models $\Phi_0(x)$, $\widehat{\Phi}'_n(x)$, $\widehat{\Phi}''_n(x)$ may be considered as points in infinite dimensional vector spaces. Finally, we would like to emphasize that the method can be generalized to the case, when the input density is also unknown, see [4].

References

1. *Aizerman, M. A., Braverman, E. E., Rozonoer, L. I.*, The method of potential functions for the problem of restoring the characteristic of a function converter for randomly observed points. Avtomat. i Telemekh. vol. **25**, No. 12, (1964), pp. 1705–1714 (in Russian).
2. *Braverman, E. M.*, On the potential functions method. Avtomat. i Telemekh., vol. **26**, No. 12, (1966), pp. 2205–2213 (in Russian).
3. *Bubnicki, Z.*, System identification via estimation of probability distribution and moments. 2nd IFAC Symposium on Identification, Prague 1970.
4. *Greblicki, W.*, Asymptotically optimal probabilistic algorithms for pattern recognition and identification. Scientific Papers of the Institute of Technical Cybernetics of the Wrocław Technical University No. 18, Monographs No. 3, Wrocław, 1974 (in Polish).
5. *Sansone, G.*, Orthogonal functions, Interescience, New York, 1959.
6. *Saridis, G. N., Nikolic, Z. J., Fu, K. S.*, Stochastic approximation algorithms for system identification, estimation, and decomposition of mixtures. IEEE Trans. SSC, vol. **5**, pp. 8–15, January 1969.
7. *Szegő, G.*, Orthogonal polynomials, Providence 1974 (third edition).
8. *Tsypkin, Ya. Z.*, Establishing characteristics of a function transformer from randomly observed points. Avtomat. i Telemekh. vol. **26**, No. 11, (1965), pp. 1945–1950 (in Russian).

Непараметрическая идентификация методом ортогональных рядов

В. ГРЕБЛИЦКИ

(Вроцлав)

В работе представляется непараметрический метод идентификации систем без памяти. Метод использует некоторые свойства ортогональных рядов. Доказано, что последовательность модели сходится к оптимальной модели, которая является функцией регрессии выхода по входу.

W. Greblicki
Institute of Engineering Cybernetics
Wrocław Technical University
ul. Janiszewskiego 11/17
Wrocław
Poland

AKADÉMIAI KIADÓ, BUDAPEST
PUBLISHING HOUSE OF THE HUNGARIAN ACADEMY OF SCIENCES
VERLAG DER UNGARISCHEN AKADEMIE DER WISSENSCHAFTEN
MAISON D'EDITIONS DE L'ACADEMIE DES SCIENCES DE HONGRIE
ИЗДАТЕЛЬСТВО АКАДЕМИИ НАУК ВЕНГРИИ