



dać się przedstawić w postaci parametrycznej, co oznacza, że jednoznaczne określenie funkcji należącej do niej jest równoznaczne ustaleniu wektora parametrów o znanym i ograniczonym wymiarze. Problem identyfikacji parametrycznej polega zatem na estymacji rzeczywistej wartości wektora parametrów. Typowa jest tutaj sytuacja, gdy o nieznannej charakterystyce wiadomo, że jest wielomianem znanego stopnia, a identyfikacja polega na estymacji jego współczynników. Wydaje się jednak, że są zadania praktyczne, w których nie dysponuje się niestety wiedzą o badanym obiekcie na tyle szeroką, że umożliwiłoby to przedstawienie problemu w postaci parametrycznej. Można wtedy stosować metody nieparametryczne, gdyż nie wymagają one tak znacznej informacji apriorycznej. Pozwalają one identyfikować nieznaną charakterystykę tzn. otrzymywać algorytmy zbieżne do niej nawet wtedy, gdy jej postać funkcyjna jest zupełnie nieznaną i wiadomo o niej zaledwie, że jest np. całkowalna z kwadratem.

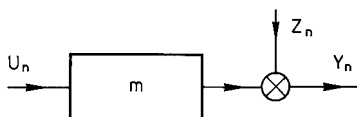
Algorytmy proponowane przez podejście nieparametryczne, czyli algorytmy nieparametryczne są zatem skutecznym narzędziem w sytuacjach, gdy wstępna informacja o systemie jest znikoma. Warto podkreślić także, że nie należy ich traktować jako konkurencyjne wobec tych, które otrzymuje się w ramach metod parametrycznych. W istotny jednak sposób poszerzają one nasze możliwości poznawcze. Pozwalają one bowiem identyfikować systemy nieliniowe przy bardzo ubogiej informacji apriorycznej, gdy klasyczne podejście parametryczne zawodzi.

Ze względu na brak publikacji na ten temat w języku polskim i jednocześnie spore zainteresowanie w literaturze obcojęzycznej, niniejsza praca ma charakter informacyjny i przeglądowy.

## 2. Identyfikacja systemu statycznego

### 2.1. Wprowadzenie

Rozpatrzmy teraz zadanie identyfikacji systemu statycznego, rys. 1.



Rys. 1. Identyfikowany system statyczny

W systemie tym:

$$Y_n = m(U_n) + Z_n, \quad (1)$$

gdzie  $U_n$  i  $Y_n$  są odpowiednio wejściem i wyjściem, a  $Z_n$  zakłóceniem pomiarowym.  $U_n$  i  $Z_n$  są zmiennymi losowymi. Ponieważ o funkcji  $m$  zakładamy, że jest mierzalna w sensie Borela, także  $Y_n$  jest zmienną losową. Ponadto sygnał wejściowy jest stacjonarnym białym szumem tzn. dla  $n \neq m$  zmienne losowe  $U_n$  i  $U_m$  są niezależne i mają jednakowe rozkłady. Podobnie zakłócenie jest stacjonarnym białym szumem, a jego wartość średnia jest równa zero. Przyjmujemy także, że sygnał wejściowy i szum pomiarowy są procesami stochastycznymi wzajemnie niezależnymi. Założenia powyższe obowiązują w całej pracy i nie będą powtarzane w podawanych w niej twierdzeniach.

Zadanie identyfikacji polega na wnioskowaniu o nieznaną charakterystyce  $m$  na podstawie obserwacji niezależnych par  $(U_1, Y_1), \dots, (U_n, Y_n)$ . Zauważając, że:

$$m(u) = E\{Y_n \mid U_n = u\}, \quad (2)$$

można stwierdzić, że z punktu widzenia statystyki matematycznej problem polega na estymacji funkcji regresji. Zakładając teraz istnienie gęstości prawdopodobieństwa sygnału wejściowego możemy stwierdzić, że:

$$m(u) = g(u)/f(u), \quad (3)$$

gdzie  $g(u) = m(u)f(u)$ . Przez  $f$  oznaczyliśmy gęstość prawdopodobieństwa zmiennej losowej  $U_n$ . Podając różne oszacowania wyrażeń znajdujących się w liczniku i mianowniku ułamka (3) otrzymuje się różne algorytmy identyfikacji nieparametrycznej.

## 2.2. Algorytmy identyfikacji

Omówimy teraz dwa często stosowane nieparametryczne algorytmy estymacji  $m(u)$  tzn. algorytm z jądrem i algorytm ortogonalny.

Zauważmy, że jako przybliżenia licznika i mianownika w ułamku (3) można przyjąć odpowiednio:

$$\frac{1}{2h} \int_{u-h}^{u+h} \left[ \int yf(v, y) dy \right] dv \quad \text{oraz} \quad \frac{1}{2h} \int_{u-h}^{u+h} f(v) dv, \quad (4)$$

gdzie  $h$  jest pewną dodatnią liczbą. Przez  $f(\cdot, \cdot)$  oznaczyliśmy łączną gęstość prawdopodobieństwa wejścia i wyjścia. Wyrażenia podane w (4) dają się łatwo oszacować na podstawie danych pomiarowych. Oznaczając bowiem przez  $I_u$  zbiór wszystkich  $i \in \{1, 2, \dots, n\}$ , dla których  $|U_i - u| \leq h$ , można jako nieobciążone oszacowania tych wyrażeń przyjąć odpowiednio:

$$\frac{1}{2nh} \sum_{i \in I_u} Y_i \quad \text{oraz} \quad \frac{1}{2nh} \sum_{i \in I_u} U_i. \quad (5)$$

Oznaczając:

$$K(u) = \begin{cases} \frac{1}{2} & \text{dla } |u| \leq 1 \\ 0 & \text{dla } |u| > 1, \end{cases} \quad (6)$$

można z kolei wyrażenia w (5) przepisać w następującej postaci:

$$\frac{1}{nh} \sum_{i=1}^n Y_i K\left(\frac{u - U_i}{h}\right) \quad \text{oraz} \quad \frac{1}{nh} \sum_{i=1}^n K\left(\frac{u - U_i}{h}\right). \quad (7)$$

Ponieważ powyższe wielkości przyjęliśmy jako oszacowania tych w (4), które są z kolei przybliżeniami  $g(u)$  i  $f(u)$ , to estymator nieznanego  $m(u)$  definiujemy następująco:

$$\frac{\frac{1}{nh} \sum_{i=1}^n Y_i K\left(\frac{u - U_i}{h}\right)}{\frac{1}{nh} \sum_{i=1}^n K\left(\frac{u - U_i}{h}\right)}.$$

Jest oczywiste, że aby zapewnić zbieżność powyższego wyrażenia do  $m(u)$  należy współczynnik  $h$  uzależnić od liczby pomiarów  $n$ . Prowadzi to do następującej ostatecznej definicji estymatora:

$$\hat{m}_n(u) = \frac{\frac{1}{nh(n)} \sum_{i=1}^n Y_i K\left(\frac{u - U_i}{h(n)}\right)}{\frac{1}{nh(n)} \sum_{i=1}^n K\left(\frac{u - U_i}{h(n)}\right)}. \quad (8)$$

gdzie  $K$  jest tzw. jądrem estymatora,  $\{h(n); n = 1, 2, \dots\}$  ciągiem liczb dodatnich. Dopuszczając stosowanie innych funkcji  $K$ , a nie tylko tej zdefiniowanej wzorem (6), otrzymuje się całą klasę algorytmów nazywaną w literaturze estymatorem z jądrem.

Algorytm, który przedstawimy teraz wykorzystuje szeregi ortogonalne i dlatego nazywany jest ortogonalnym. Załóżmy, że  $\{\varphi_k; k = 0, 1, 2, \dots\}$  jest układem funkcji ortonormalnych na całej prostej. Niech:

$$g(u) \sim \sum_{k=0}^{\infty} a_k \varphi_k(u) \quad \text{oraz} \quad f(u) \sim \sum_{k=0}^{\infty} b_k \varphi_k(u),$$

co oznacza, że:

$$a_k = \int m(u) f(u) \varphi_k(u) du \quad \text{oraz} \quad b_k = \int f(u) \varphi_k(u) du.$$

Jako przybliżenie  $g(u)$  i  $f(u)$  czyli wyrażen stojących odpowiednio w liczniku i mianowniku ułamka (3) przyjmiemy teraz

$$\sum_{k=0}^{N(n)} a_k \varphi_k(u) \quad \text{oraz} \quad \sum_{k=0}^{N(n)} b_k \varphi_k(u),$$

gdzie  $\{N(n); n = 1, 2, \dots\}$  jest pewnym ciągiem liczb naturalnych. Jest oczywiste, że współczynniki  $a_0, a_1, \dots$  oraz  $b_0, b_1, \dots$  rozwinięć ortogonalnych są nieznane. Zauważając jednak, że  $a_k = E\{Y_1 \varphi_k(U_1)\}$  oraz  $b_k = E\varphi_k(U_1)$ , można podać następujące nieobciążone ich estymatory:

$$\tilde{a}_k = \frac{1}{n} \sum_{i=1}^n Y_i \varphi_k(U_i) \quad \text{oraz} \quad \tilde{b}_k = \frac{1}{n} \sum_{i=1}^n \varphi_k(U_i).$$

W rezultacie otrzymujemy następujący algorytm identyfikacji:

$$\tilde{m}_n(u) = \frac{\sum_{k=0}^{N(n)} \tilde{a}_k \varphi_k(u)}{\sum_{k=0}^{N(n)} \tilde{b}_k \varphi_k(u)}. \quad (9)$$

Stosując różne układy funkcji ortogonalnych otrzymuje się różne wersje algorytmu (9). Jako układ ten można przyjąć np. rodzinę funkcji Hermite'a. Jeśli sygnał wejściowy jest ograniczony, to można stosować układy ortogonalne na odcinku, czyli np. układ trygonometryczny lub wielomiany Legendre'a.

Pomimo, że zasady konstrukcji estymatorów (8) i (9) są całkowicie różne, istnieje

między nimi pewne podobieństwo. Oznaczając bowiem:

$$k_N(u, v) = \sum_{k=0}^N \varphi_k(u) \varphi_k(v),$$

możemy łatwo sprawdzić, że:

$$\tilde{m}_n(u) = \frac{\sum_{i=1}^n Y_i k_{N(n)}(u, U_i)}{\sum_{i=1}^n k_{N(n)}(u, U_i)},$$

gdzie  $k_N$  jest tzw. jądrem szeregu ortogonalnego. Dla układu trygonometrycznego  $k_N$  jest jądrem Dirichleta, gdy natomiast system ortonormalny jest typu wielomianowego, jądro to można łatwo wyrazić w znacznie prostszej obliczeniowo postaci jeśli skorzysta się z wzoru Christoffela-Darboux.

### 2.3. Zbieżność algorytmów identyfikacji

Podstawowy problem jaki się teraz pojawia, to zbieżność przedstawionych powyżej algorytmów identyfikacji do nieznannej charakterystyki. W odniesieniu do algorytmu z jądrem należy odpowiedzieć na pytanie w jaki sposób wybrać ciąg  $\{h(n)\}$  oraz jądro  $K$ . Wobec algorytmu ortogonalnego wylania się zagadnienie wyboru układu funkcji ortogonalnych i ciągu  $\{N(n)\}$ . O zagadnieniach tych traktuje szereg prac. Zaczniemy od twierdzenia dotyczącego algorytmu z jądrem.

**Twierdzenie 1.** Załóżmy, że:

$$\int m^2(u) f(u) du < \infty \quad \text{i} \quad EZ_1^2 < \infty. \quad (10)$$

Niech  $K$  będzie nieujemną i ograniczoną funkcją mierzalną w sensie Borela taką, że:

$$\int K(u) du = 1 \quad \text{oraz} \quad |u| |K(u)| \rightarrow 0, \quad \text{gdy} \quad |u| \rightarrow \infty.$$

Jeżeli:

$$h(n) \xrightarrow{n} 0 \quad \text{oraz} \quad nh(n) \xrightarrow{n} \infty, \quad (11)$$

to:

$$\hat{m}_n(u) \xrightarrow{n} m(u) \quad \text{w/g prawdopodobieństwa}$$

w każdym punkcie  $u \in (-\infty, \infty)$ , w którym zarówno charakterystyka  $m$  jak i gęstość  $f$  są ciągłe oraz jednocześnie  $f(u) > 0$ .

**Dowód.** Oznaczmy przez  $\hat{g}_n(u)$  i  $\hat{f}_n(u)$  wyrażenia stojące odpowiednio w liczniku i mianowniku wyrażenia (8). Oczywiście

$$E\hat{g}_n(u) = \frac{1}{h(n)} \int K\left(\frac{u-v}{h(n)}\right) g(v) dv.$$

Zbieżność powyższego do  $g(u)$ , gdy  $h(n)$  zdąży do zera jest przedmiotem twierdzeń analizy matematycznej. Korzystając np. z Twierdzenia 9.9 z książki Wheedena i Zygmunda [33] wnioskujemy, że, ponieważ  $g$  jest funkcją całkowalną, wyrażenie powyższe zbiega się do

$g(u)$  w każdym punkcie ciągłości funkcji  $g$ . Z kolei:

$$\begin{aligned} \text{var } \hat{g}_n(u) = \frac{1}{nh^2(n)} \text{var}(Y_1 K\left(\frac{u-U_1}{h(n)}\right)) &\leq \frac{k}{nh(n)} \left[ \sigma^2 \frac{1}{h(n)} \int K\left(\frac{u-v}{h(n)}\right) f(v) dv \right. \\ &\left. + \frac{1}{h(n)} \int K\left(\frac{u-v}{h(n)}\right) m^2(v) f(v) dv \right] \end{aligned}$$

gdzie  $k = \sup K(u)$ ,  $\sigma^2 = \text{var } Z_1$ . Korzystając ponownie ze wspomnianego powyżej twierdzenia dochodzimy do wniosku, że wyrażenie w nawiasach kwadratowych zbiega się do  $\sigma^2 f(u) + m^2(u) f(u)$ , gdy  $h(n)$  dąży do zera. Zbieżność ta ma miejsce w każdym punkcie  $u$ , w którym zarówno  $m$  jak i  $f$  są ciągłe. Wykazaliśmy zatem, że:

$$E(\hat{g}_n(u) - g(u))^2 \xrightarrow{n} 0,$$

w każdym punkcie  $u$ , w którym zarówno  $m$  jak i  $f$  są ciągłe.

Powtarzając powyższe rozumowanie w odniesieniu do  $f(u)$  można łatwo sprawdzić, że:

$$E(\hat{f}_n(u) - f(u))^2 \xrightarrow{n} 0,$$

w każdym punkcie ciągłości funkcji  $f$ .

Z tego co pokazaliśmy oraz definicji estymatora wynika teza, co kończy dowód.

Istnieje wiele funkcji spełniających warunki Twierdzenia 1, i które można przyjąć jako jądro estymatora. Oto kilka przykładów:

$$\begin{array}{ll} \text{a) } \begin{cases} 1 - |u| & \text{dla } |u| \leq 1 \\ 0 & \text{dla } |u| > 1 \end{cases} & \text{c) } (1/2)e^{-|u|} \\ \text{b) } (2\pi)^{-1/2} e^{-u^2/2} & \text{d) } 1/\pi(1 + u^2). \end{array}$$

Jeśli jako ciąg liczbowy przyjąć  $h(n) = cn^{-\alpha}$ ,  $c > 0$ , to warunki (11) są spełnione dla  $0 < \alpha < 1$ .

Zauważmy, że jedynym założeniem Twierdzenia 1 odnoszącym się do informacji wstępnej o nieznanym systemie jest (10). Wymaga ono aby istniał drugi moment  $Em^2(U_1)$  sygnału pojawiającego się bezpośrednio za elementem o nieznannej charakterystyce oraz aby zakłócenie pomiarowe posiadało skończoną wariancję. Nie wymagana jest przy tym żadna znajomość funkcyjnego kształtu nieznannej charakterystyki. Odnośnie sygnału wejściowego, żąda się aby jego rozkład prawdopodobieństwa posiadał gęstość, aczkolwiek może być ona zupełnie dowolna.

A teraz uwaga dotycząca specyfikacji punktów, w których algorytm zbiega się do nieznannej charakterystyki. Zauważmy, że jeśli gęstość  $f$  jest ciągła, to algorytm jest zgodny w każdym punkcie, w którym charakterystyka  $m$  jest ciągła a gęstość dodatnia. Zwróćmy jednak uwagę, że charakterystyka nie musi być bynajmniej funkcją wszędzie ciągłą, może to być np. charakterystyka typu przekaźnikowego.

Twierdzenie 1 można łatwo wzmocnić, aczkolwiek wymagałoby to stosowania bardziej zaawansowanych narzędzi. Stosując np. technikę obcięć można wykazać, że założenie (10) można zastąpić przez mniej ograniczający warunek  $\int |m(u)|f(u) du < \infty$  i  $E|Z_1| < \infty$ , Devroye [1], Greblicki, Krzyżak i Pawlak [14].

W pracy tej zakładamy, że rozkład sygnału wejściowego jest na tyle regularny, że istnieje jego gęstość. Z założenia tego można zrezygnować, gdyż okazuje się, że algorytm zbiega się do nieznannej charakterystyki także przy dowolnym rozkładzie sygnału wejściowego nawet takim, który gęstości nie posiada, Devroye [1], Devroye i Wagner [3], Greblicki, Krzyżak i Pawlak [14], Stone [31].

Twierdzenie 1 daje sporą swobodę w wyborze jądra  $K$  i ciągu liczbowego  $\{h(n)\}$ . Jest oczywiste, że wybór ten ma wpływ na szybkość z jaką algorytm się zbiega. Krytyczny okazuje się tutaj wybór ciągu liczbowego, gdyż wpływ jądra na tę szybkość jest niewielki. Można wykazać, że jeśli zarówno  $m$  jak i  $f$  są w punkcie  $u$  dwukrotnie różniczkowalne oraz  $h(n) = c_1 n^{-1/5}$ ,  $c_1 > 0$ , to:

$$P\{|\hat{m}_n(u) - m(u)| > \varepsilon |m(u)|\} \leq d_1 n^{-4/5},$$

dla każdego  $\varepsilon > 0$ , gdzie  $d_1$  jest pewną liczbą niezależną od  $n$ , Greblicki i Krzyżak [13]. Rezultat ten może stanowić pewną rekomendację odnośnie wyboru ciągu liczbowego.

Jeśli chodzi o algorytm ortogonalny, to różne jego wersje otrzymywane przez wykorzystywanie różnych układów ortogonalnych wymagają różnych sposobów analizy. Mając zatem na uwadze ograniczoność miejsca poprzestaniemy na algorytmie z układem Hermite'a. Jak wiadomo, w układzie tym:

$$\varphi_k(u) = (2^k k! \pi^{1/2})^{-1/2} e^{-u^2/2} H_k(u),$$

gdzie:

$$H_k(u) = e^{u^2} \frac{d^k}{du^k} e^{-u^2}$$

jest  $k$ -tym wielomianem Hermite'a. Dla przykładu,  $H_0(u) = 1$ ,  $H_1(u) = -2u$ ,  $H_2(u) = 4u^2 - 2$ ,  $H_3(u) = -8u^3 + 12u$ , itd.

Twierdzenie 2. Załóżmy, że spełnione jest założenie (10). Jeżeli  $f, g \in L_p$ ,  $p > 1$ , oraz:

$$N(n) \xrightarrow{n} \infty \quad \text{oraz} \quad N(n)n^2 \xrightarrow{n} 0, \quad (12)$$

to dla algorytmu wykorzystującego układ Hermite'a

$$\tilde{m}_n(u) \xrightarrow{n} m(u) \quad \text{w/g prawdopodobieństwa}$$

w każdym punkcie  $u \in (-\infty, \infty)$ , w którym zarówno charakterystyka  $m$  jak i gęstość  $f$  są różniczkowalne i jednocześnie  $f(u) > 0$ .

Dowód powyższego twierdzenia można znaleźć w pracy Greblicki i Pawlak [15].

Trzeba dodać, że podobnie jak miało to miejsce wobec Twierdzenia 1 także i w Twierdzeniu 2 założenie (10) można osłabić, Greblicki i Pawlak [15].

Warto w tym miejscu zwrócić uwagę na to, że zbiory punktów określone przez Twierdzenie 1 i 2, w których algorytm z jądrem i algorytm ortogonalny z układem Hermite'a zbiegają się, są różne. Jeżeli np. sygnał wejściowy ma rozkład normalny, to zgodnie z Twierdzeniem 1 algorytm z jądrem jest zbieżny do  $m(u)$  w każdym punkcie ciągłości  $m$ . Twierdzenie 2 zapewnia natomiast w tej sytuacji zbieżność algorytmu ortogonalnego jedynie w punktach różniczkowalności  $m$ . Różnica ta nie jest przypadkowa i występuje także przy korzystaniu z innych układów ortogonalnych, gdyż ma głębsze przyczyny.

Odnośnie szybkości zbieżności rozważanego algorytmu ortogonalnego wiadomo, że jeśli  $m$  oraz  $f$  są dwukrotnie różniczkowalne, to przy spełnieniu pewnych dodatkowych założeń:

$$P\{|\tilde{m}_n(u) - m(u)| > \varepsilon m(u)\} \leq d_2 n^{-4/5},$$

$\varepsilon > 0$ ,  $d_2 > 0$ , gdy tylko  $N(n) = [c_2 n^{1/2}]$ , gdzie  $c_2 > 0$ , natomiast  $[\cdot]$  oznacza część całkowitą, Greblicki i Pawlak [15]. Zauważmy, że szybkość powyższa posiada rząd  $n^{-4/5}$ , czyli ten sam jaki ma algorytm z jądrem. Warto dodać, że obserwacja ta powtarza się także dla algorytmów korzystających z innych układów ortogonalnych.

### 2.4. Uwagi

Powyżej podano jedynie podstawowe własności prezentowanych algorytmów identyfikacji. Ograniczono się przy tym do jednowymiarowego wejścia. O ile modyfikacja algorytmu z jądrem na przypadek wielu wejść i dalsza analiza nie stanowią poważnej trudności, to nie można niestety tego powiedzieć o algorytmie ortogonalnym. Jak do tej pory, z wyjątkiem algorytmu z wielowymiarowym układem trygonometrycznym, niczego istotnego na ten temat nie wiadomo.

Duże zainteresowanie budzi błąd całkowity, a nie tylko punktowy, do którego ograniczyliśmy się tutaj, gdyż podaje on globalną ocenę jakości algorytmu. Do błędu całkowitego wrócimy omawiając rezultaty eksperymentu numerycznego. Badana jest także zbieżność z prawdopodobieństwem 1, oraz inne typy zbieżności probabilistycznej. Otrzymywane w literaturze rezultaty dotyczą także specyfikacji punktów, w których algorytmy są zbieżne. W tej pracy stwierdziliśmy np. zbieżność w punktach ciągłości, czy różniczkowalności nieznannej charakterystyki. Znane są wyniki o zbieżności w prawie wszystkich, według odpowiednich miar, punktach.

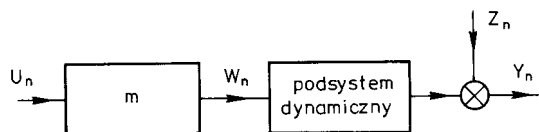
Stosuje się także różne pochodne wersje przedstawionych powyżej algorytmów. Odnośnie algorytmu z jądrem warto wspomnieć o jego wersjach rekurencyjnych, Devroye i Wagner [2], Greblicki i Krzyżak [13], Greblicki i Pawlak [18].

Warto odnotować próbę zastosowania algorytmów ortogonalnych do estymacji zmieniającej się w czasie funkcji regresji, czyli inaczej mówiąc do identyfikacji obiektu niestacjonarnego, Greblicki, Rutkowska i Rutkowski [21].

Obszerny i szczegółowy przegląd wyników uzyskanych w dziedzinie nieparametrycznej estymacji funkcji regresji oraz gęstości prawdopodobieństwa zawiera książka Prakasy Rao [24]. Można w niej znaleźć bogaty wykaz publikacji wydanych do 1980 roku dotyczących nie tylko estymacji jako takiej lecz także zastosowań.

### 3. Identyfikacja systemów dynamicznych

Podjęcie nieparametryczne zastosowano także do identyfikacji pewnej klasy nieliniowych systemów dynamicznych. Klasa ta obejmuje systemy określane jako systemy Hammersteina, a które nazwę zawdzięczają temu, że ich zachowanie opisuje się funkcjonalem Hammersteina.



Rys. 2. Identyfikowany system Hammersteina

W systemie takim, rys. 2,

$$W_n = m(U_n),$$

gdzie  $m$  jest nieznaną nieliniową charakterystyką. Podsystem liniowy opisywany jest następującym równaniem stanu:

$$\begin{cases} \mathbf{X}_{n+1} = \mathbf{A}\mathbf{X}_n + \mathbf{b}W_n \\ Y_n = \mathbf{c}^T \mathbf{X}_n + Z_n, \end{cases}$$



gdzie  $\mathbf{X}_n$  jest wektorem stanu,  $Z_n$  szumem pomiarowym,  $Y_n$  wyjściem systemu,  $T$  oznacza transpozycję.  $\mathbf{A}$ ,  $\mathbf{b}$  oraz  $\mathbf{c}$  są nieznane, lecz wiadomo, że  $\mathbf{A}$  jest macierzą, której wszystkie wartości własne leżą wewnątrz koła jednostkowego. Podobnie jak poprzednio sygnał wejściowy jest stacjonarnym białym szumem losowym określonym tym razem dla  $n = \dots, -1, 0, 1, \dots$ . Odnośnie zakłóceń pomiarowych zakładamy, że mają skończoną wariancję. Nieznana charakterystyka  $m$  spełnia następujące założenie:

$$|m(u)| \leq c_1 + c_2|u|,$$

dla pewnych  $c_1$  oraz  $c_2$ . Dzięki temu zadanie jest prawidłowo sformułowane w tym sensie, że  $\mathbf{X}_n$  będące sumą nieskończonego szeregu zmiennych losowych jest także zmienną losową. Jest oczywiste, że założenie to nie jest w żadnym wypadku związane z przedstawioną poniżej metodą identyfikacji.

Zadanie polega na identyfikacji obydwu podsystemów na podstawie obserwacji wejścia i wyjścia całego systemu tzn. na podstawie  $U_1, Y_1, U_2, Y_2, \dots$ . Ponieważ odpowiedź impulsową podsystemu liniowego można łatwo estymować metodami korelacyjnymi, ograniczymy się do identyfikacji podsystemu nieliniowego.

Zauważmy, że:

$$E\{Y_{n+1} | U_n = u\} = \alpha m(u) + \beta, \quad (13)$$

gdzie  $\alpha = \mathbf{c}^T \mathbf{b}$  oraz  $\beta = \mathbf{c}^T \mathbf{A} E \mathbf{X}_n$ .

Załóżmy chwilowo, że gęstość sygnału wejściowego jest funkcją parzystą, a nieznana charakterystyka nieparzystą. Wówczas  $E \mathbf{X}_n = 0$  i w rezultacie  $\beta = 0$ . Do przypadku  $\beta \neq 0$  wrócimy później. Estymacja funkcji regresji (11) prowadzi zatem do wykrycia  $\alpha m(u)$ , gdzie  $\alpha$  jest nieznaną liczbą. Nie jest to jednak specyficzną własnością prezentowanej metody, lecz cechą ogólną wynikającą z szeregowej struktury identyfikowanego systemu. Każdy bowiem algorytm konstruowany przy przyjętych w pracy założeniach wstępnych o systemie posiadać może jedynie zdolność estymacji nieznanej charakterystyki z dokładnością do nieznanego mnożnika, którym w naszym przypadku jest  $\alpha$ .

Algorytm z jądrem przyjmuje teraz następującą postać:

$$\hat{m}_n(u) = \frac{\sum_{i=1}^n Y_i K\left(\frac{u - U_{i-1}}{h(n)}\right)}{\sum_{i=1}^n K\left(\frac{u - U_{i-1}}{h(n)}\right)}, \quad (14)$$

natomiast algorytm ortogonalny zachowuje postać (9) z tą różnicą, że teraz:

$$\begin{aligned} \tilde{a}_k &= \frac{1}{n} \sum_{i=1}^n Y_i \varphi_k(U_{i-1}), \\ \tilde{b}_k &= \frac{1}{n} \sum_{i=1}^n \varphi_k(U_{i-1}) \end{aligned}$$

Trzeba tutaj zwrócić uwagę na zasadniczą różnicę pomiędzy sytuacją obecną, a tą, z którą mieliśmy do czynienia przy identyfikacji systemu statycznego. Wtedy wnioskowaliśmy na podstawie obserwacji stochastycznie niezależnych par  $(U_1, Y_1), (U_2, Y_2), \dots$ , natomiast teraz identyfikujemy mając obserwacje par zależnych  $(U_0, Y_1), (U_1, Y_2), \dots$ . Pary te są zależne, gdyż zmienne losowe  $Y_1, Y_2, \dots$  są zależne, ponieważ są wyjściem systemu dynamicznego.

Obecny problem sprowadza się do poprzedniego gdy  $A = 0$ , tzn. gdy liniowy podsystem dynamiczny redukuje się do opóźnienia o jedną chwilę.

Wspomniana powyżej zależność obserwowanych par powoduje pojawienie się dodatkowych trudności analitycznych i dlatego ograniczymy się jedynie do stwierdzenia, że rezultaty dotyczące algorytmu (8) podane w Twierdzeniu 1 można powtórzyć dla estymatora (14), Greblicki i Pawlak [16, 17]. Pewnym zaskoczeniem jest, że wystąpienie zależności pomiędzy obserwacjami nie powoduje zmniejszenia rzędu szybkości zbieżności.

Można wykazać, że teza Twierdzenia 2 dotycząca algorytmu ortogonalnego z systemem Hermite'a pozostaje słuszna także i teraz, Greblicki [11]. Szybkość zbieżności także nie ulega pogorszeniu.

Do tej pory procesy  $\{U_n\}$  i  $\{Y_n\}$  były określone dla  $n = \dots, -1, 0 - 1, \dots$  i były przy tym stacjonarne. Interesująca jest także sytuacja, gdy procesy te definiowane są dla  $n = 0, 1, 2, \dots$ , tzn. gdy nie mają one w chwili zerowej żadnej historii. Proces  $\{Y_n\}$  jest teraz niestacjonarny, pomimo, że proces wejściowy i system jako taki są stacjonarne. Łatwo spostrzec, że teraz

$$E\{Y_{n+1} | U_n = u\} = \alpha m(u) + \beta_n,$$

gdzie  $\beta_n = c^T A E X_n$ , przy czym  $E X_n = A^n x_0$ , gdzie  $x_0$  jest początkowym wektorem stanu. Zakładamy ponadto, że wektor ten nie ma charakteru losowego. Okazuje się, że algorytm z jądrem zachowuje się poprawnie także i w tej sytuacji (Greblicki i Pawlak [19]).

Ze względów obliczeniowych wygodnie jest niekiedy stosować rekurencyjne wersje algorytmu (14) (Greblicki i Pawlak [20]). Definiuje się je następująco:

$$m'_n(u) = \frac{\sum_{i=1}^n \frac{1}{h(i)} Y_i K\left(\frac{u - U_{i-1}}{h(i)}\right)}{\sum_{i=1}^n \frac{1}{h(i)} K\left(\frac{u - U_{i-1}}{h(i)}\right)}$$

oraz:

$$m''_n(u) = \frac{\sum_{i=1}^n Y_i K\left(\frac{u - U_{i-1}}{h(i)}\right)}{\sum_{i=1}^n K\left(\frac{u - U_{i-1}}{h(i)}\right)}$$

Łatwo sprawdzić, że  $m'_n(u) = L'_n(u)/M'_n(u)$ , gdzie:

$$L'_n(u) = L'_{n-1}(u) + \frac{1}{n} \left[ \frac{1}{h(n)} Y_n K\left(\frac{u - U_{i-1}}{h(n)}\right) - L'_{n-1}(u) \right]$$

oraz:

$$M'_n(u) = M'_{n-1}(u) + \frac{1}{n} \left[ \frac{1}{h(n)} K\left(\frac{u - U_{i-1}}{h(n)}\right) - M'_{n-1}(u) \right],$$

przy czym  $L'_0(u) = 0$  oraz  $M'_0(u) = 0$ . Alternatywna formuła rekurencyjna, którą podamy dla drugiego z przedstawionych powyżej algorytmów identyfikacji, jest następująca:

$$L''_n(u) = L''_{n-1}(u) + Y_n K\left(\frac{u - U_{i-1}}{h(n)}\right)$$

oraz

$$M''_n(u) = M''_{n-1}(u) + K\left(\frac{u - U_{i-1}}{h(n)}\right),$$

przy czym  $L_0''(u) = 0$  oraz  $M_0''(u) = 0$ . Przez  $L_n''(u)$  i  $M_n''(u)$  oznaczyliśmy odpowiednio licznik i mianownik we wzorze definiującym algorytm.

Wróćmy teraz do założenia dotyczącego  $\beta$ . Jeżeli  $\beta \neq 0$ , to można jednak w pewnych sytuacjach estymować  $m(u)$ . Jeśli bowiem  $m(0) = 0$ , co jest jak się wydaje częstym przypadkiem, to jako algorytm identyfikacji można przyjąć  $\hat{m}_n(u) - \hat{m}_n(0)$ . Jest oczywiste, że proponowany algorytm zbiega się do  $\alpha m(u)$ , jeśli tylko  $\hat{m}_n(0)$  zdąży do  $m(0)$ .

#### 4. Przykład numeryczny

Podane w pracy rezultaty analizy teoretycznej dotyczą własności asymptotycznych. Nie wiele, jeśli cokolwiek, wiadomo o ich zachowaniu się dla niewielkiej liczby obserwacji. W tym obszarze musimy się jedynie zadowolić wynikami badań symulacyjnych. Poniżej podamy wyniki eksperymentu numerycznego, w którym zbadano zachowanie się dwóch algorytmów, przy pomocy których estymowano charakterystykę nieliniową w systemie Hammersteina. W systemie tym charakterystyka ta jest funkcją nieciągłą i jest zdefiniowana następująco:

$$m(u) = \begin{cases} 0,4 & \text{dla } 0,5 < u \\ u & \text{dla } -0,5 < u < 0,5 \\ -0,4 & \text{dla } u < -0,5. \end{cases}$$

Ponadto  $X_n$  jest skalarem, przy czym  $X_{n+1} = 0,2 \cdot X_n + Z_n$ . Sygnał wejściowy ma rozkład równomierny na odcinku  $(-1;1)$ , a zakłócenie na odcinku  $(-0,5;0,5)$ . Pierwszy z estymatorów, to algorytm z jądrem, przy czym jako jądro przyjęto (6). Jako drugi zastosowano algorytm ortogonalny z układem trygonometrycznym.

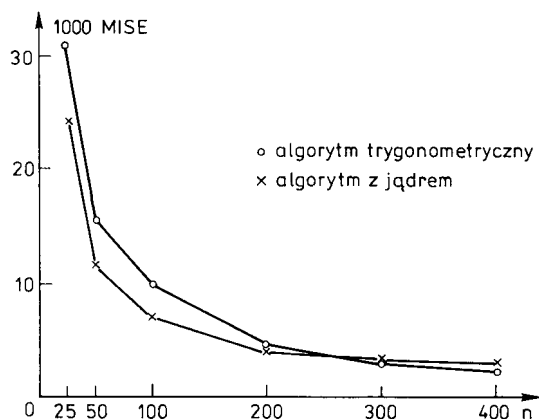
Jakość każdego z tych algorytmów oceniano biorąc pod uwagę średni, całkowity błąd kwadratowy (MISE):

$$E \int_{-0,8}^{0,8} (m_n(u) - m(u))^2 du,$$

przy czym jako  $m_n(u)$  przyjmowano kolejno każdy z dwóch algorytmów. Jakość oceniano na odcinku będącym częścią wyjściowego przedziału  $(-1;1)$ . Uczyniono tak, aby nie dyskwalifikować z góry estymatora trygonometrycznego. Jak bowiem wiadomo, rozwinięcia trygonometryczne nie zachowują się poprawnie na końcach odcinka, na którym dokonuje się tego rozwinięcia.

Błąd szacowano empirycznie na podstawie wielu niezależnych realizacji. Jako liczbę obserwacji przyjmowano kolejno 25, 50, 100, 200, 300 i 400. Jeśli chodzi o estymator z jądrem, to dla każdego z powyższych  $n$  wybrano eksperymentalnie takie  $h(n)$ , które minimalizuje nasz błąd całkowity. Podobnie postąpiono z algorytmem trygonometrycznym z tą różnicą, że tym razem wybierano optymalne  $N(n)$ . W ten sposób dla każdego z estymatorów otrzymano, dla podanych powyżej  $n$ , najmniejszy, możliwy do uzyskania błąd całkowity, a wyniki przedstawiono na rys. 3.

Narzucającym się spostrzeżeniem jest to, że otrzymane wykresy są bardzo podobne, gdyż zaobserwowane różnice w znacznym stopniu mieszczą się w ramach błędu spowodowanego empirycznym charakterem przykładu. Wydaje się więc, że podane wyżej wyniki analizy teoretycznej i przedstawione rezultaty przykładu symulacyjnego skłaniają do wnio-



Rys. 3. Zależność błędów całkowitych od liczby obserwacji

sku o bardzo podobnym zachowaniu się obydwu algorytmów zarówno dla małych jak i dużych  $n$ .

### 5. Uwagi końcowe

W niniejszej pracy ograniczyliśmy się do dwóch typów algorytmów nieparametrycznych. Trzeba jednak stwierdzić, że bada się ich znacznie więcej, a najpopularniejsze z nich wydają się splajnowy, Silverman [30], i najbliższy sąsiad, Stone [31]. Czynione są także próby powiązania podejścia nieparametrycznego z metodą najmniejszych kwadratów, Rafajłowicz [26].

Tutaj omawialiśmy sytuację, w której system pobudzany jest sygnałem losowym. Badane są także algorytmy identyfikujące nieznaną charakterystykę obiektu statycznego w przypadku, gdy wejście jest natury deterministycznej, Gałkowski i Rutkowski [4], Georgiev [5, 6], Georgiev i Greblicki [7], Priestley i Chao [23], Schuster i Yakowitz [29]. Czynione są próby zastosowania ich do identyfikacji systemów opisywanych równaniami o pochodnych cząstkowych, Rafajłowicz [25], a także niestacjonarnych Rutkowski [27, 28].

Na zakończenie należałoby jeszcze raz stwierdzić, że metody nieparametryczne stwarzają możliwości identyfikacji systemów w warunkach daleko posuniętego braku informacji wstępnej o badanym systemie. Jest przy tym oczywiste, że zapłacić trzeba za to pewną cenę, którą jest wolniejsza ich zbieżność. Jak podaliśmy, szybkość, którą zapewniają one przy dwukrotnie różniczkowalnej charakterystyce, jest rzędu  $n^{-4/5}$ . Metody parametryczne mają z reguły rząd  $n^{-1}$ . Z drugiej jednak strony algorytmy nieparametryczne są pod względem obliczeniowym nieporównanie szybsze od parametrycznych otrzymywanych np. metodą najmniejszych kwadratów, która wymaga czasochłonnego odwracania macierzy. Różnica ta staje się tym bardziej istotna im większy jest wymiar nieznanego wektora parametrów.

Wydaje się, że na koniec można stwierdzić, że nieparametryczne metody estymacji są istotną propozycją ze strony statystyki matematycznej, którą warto wykorzystać i rozwijać dla potrzeb identyfikacji systemów nieliniowych.

## Wykaz literatury

- [1] L. DEVROYE: On the almost everywhere convergence of non-parametric regression function estimate. *Annals of Statistics*, vol. 9 (1981), pp. 1310–1319.
- [2] L. DEVROYE, T.J. WAGNER: On the L1 convergence of kernel regression function estimators with application in discrimination. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, vol. 51 (1980), pp. 15–25.
- [3] L. DEVROYE, T.J. WAGNER: Distribution-free consistency results in nonparametric discrimination and regression function estimation. *Annals of Statistics*, vol. 8 (1980), pp. 231–239.
- [4] T. GALKOWSKI, L. RUTKOWSKI: Nonparametric recovery of multivariate function with applications to system identification. *Proc. IEEE*, vol. 73 (1985), pp. 942–943.
- [5] A. GEORGIEV: On the recovery of functions and their derivatives from imperfect measurements. *IEEE Trans. Systems, Man, and Cybernetics*, vol. SMC-14 (1984), pp. 900–904.
- [6] A. GEORGIEV: Asymptotic properties of the multivariate Nadaraya-Watson regression function estimate: The fixed design case. *Statistics and Probability Letters*, vol. 7 (1989), pp. 35–40.
- [7] A. GEORGIEV, W. GREBLICKI: Nonparametric function recovering from noisy observations. *J. Statistical Planning and Inference*, vol. 13 (1986), pp. 1–14.
- [8] W. GREBLICKI: Nieparametryczna identyfikacja charakterystyk statycznych. *Podstawy Sterowania*, t. 3 (1973), ss. 11–18.
- [9] W. GREBLICKI: Identyfikacja statyczna metodą szeregów ortogonalnych. *Podstawy Sterowania*, t. 4 (1974), ss. 3–12.
- [10] W. GREBLICKI: Nonparametric system identification by orthogonal series. *Problems of Control and Information Theory*, vol. 8 (1979), pp. 67–73.
- [11] W. GREBLICKI: Nonparametric orthogonal series identification of Hammerstein systems. *Int. J. Systems Science*, vol. 20 (1989), pp. 2355–2367.
- [12] W. GREBLICKI, A. KRZYŻAK: Nonparametric identification of a memoryless system with a cascade structure. *Int. J. Systems Science*, vol. 10 (1979), pp. 1301–1310.
- [13] W. GREBLICKI, A. KRZYŻAK: Asymptotic properties of kernel estimates of a regression function. *J. Statistical Planning and Inference*, vol. 4 (1980), pp. 81–90.
- [14] W. GREBLICKI, A. KRZYŻAK, M. PAWLAK: Distribution-free pointwise consistency of kernel regression estimate. *Annals of Statistics*, vol. 12 (1985), pp. 1570–1575.
- [15] W. GREBLICKI, M. PAWLAK: Fourier and Hermite series estimates of regression functions. *Annals of the Institute of Statistical Mathematics*, vol. 37 (1985), pp. 443–454.
- [16] W. GREBLICKI, M. PAWLAK: Identification of discrete Hammerstein systems using kernel regression estimate. *IEEE Trans. Automatic Control*, vol. AC-31 (1986), pp. 74–77.
- [17] W. GREBLICKI, M. PAWLAK: Hammerstein system identification by non-parametric regression estimation. *Int. J. Control*, vol. 45 (1987), pp. 343–354.
- [18] W. GREBLICKI, M. PAWLAK: Necessary and sufficient consistency conditions for a recursive kernel regression estimate. *J. Multivariate Analysis*, vol. 23 (1987), pp. 67–76.
- [19] W. GREBLICKI, M. PAWLAK: Nonparametric identification of Hammerstein systems. *IEEE Trans. Information Theory*, vol. 35 (1989), pp. 409–418.
- [20] W. GREBLICKI, M. PAWLAK: Recursive nonparametric identification of Hammerstein systems. *J. Franklin Institute*, vol. 326 (1989), pp. 461–481.
- [21] W. GREBLICKI, D. RUTKOWSKA, L. RUTKOWSKI: An orthogonal series estimate of time-varying regression. *Annals of the Institute of Statistical Mathematics*, vol. 35 (1983), pp. 215–228.
- [22] E. NADARAYA: On estimating regression. *Theory of Probability and its Application*, vol. 9 (1964), pp. 141–142.
- [23] M.B. PRIESTLEY, M.T. CHAO: Nonparametric function fitting. *J. Royal Statistical Society*, vol. 34 (1972), pp. 385–392.
- [24] B.L.S. PRAKASA RAO: Nonparametric Functional Estimation. Academic Press, Orlando, FL, 1983.
- [25] E. RAFAJŁOWICZ: Nonparametric algorithm for identification of weakly nonlinear static distributed-parameter systems. *Systems and Control Letters*, vol. 4 (1984), pp. 91–96.
- [26] E. RAFAJŁOWICZ: Nonparametric orthogonal series estimators of regression: A class attaining the optimal convergence rate in  $L_2$ . *Statistics and Probability Letters*, vol. 5 (1987), pp. 219–224.
- [27] L. RUTKOWSKI: On-line identification of time-varying systems by nonparametric technique. *IEEE Trans. Automatic Control*, vol. AC-27 (1982), pp. 228–230.

- [28] L. RUTKOWSKI: Nonparametric identification of quasi-stationary systems. *Systems and Control Letters*, vol. 6 (1985), pp. 33–35.
- [29] E. SCHUSTER, S. YAKOWITZ: Contribution to the theory of nonparametric regression with application to system identification. *Annals of Statistics*, vol. 7 (1979), pp. 139–149.
- [30] B.W. SILVERMAN: Some aspects of the spline smoothing approach to non-parametric regression curve fitting. *J. Royal Statistical Society*, vol. 47 (1985), pp. 1–52.
- [31] C.J. STONE: Consistent nonparametric regression. *Annals of Statistics*, vol. 5 (1977), pp. 595–645.
- [32] G.S. WATSON: Smooth regression analysis. *Sankhya*, vol. 26 (1964), pp. 359–372.
- [33] R.L. WHEEDEN, A. ZYGMUND: *Measure and Integral*. M. Dekker, New York, 1977.

*Instytut Cybernetyki Technicznej  
Politechniki Wrocławskiej*

В. ГРЕБЛИЦКИ

### НЕПАРАМЕТРИЧЕСКАЯ ИДЕНТИФИКАЦИЯ СИСТЕМ УПРАВЛЕНИЯ

Резюме

В работе представлены непараметрические алгоритмы оценки функции регрессии, а также применение их к идентификации систем управления. Даются их основные свойства, такие как: сходимость и скорость этой сходимости. Затем рассматриваются непараметрические алгоритмы идентификации некоторого класса нелинейных динамических систем и представлены результаты, касающиеся их сходимости. Работа является прежде всего представлением непараметрических методов с точки зрения применения их в идентификации и рассматривает результаты, полученные до настоящего времени в этой области.

W. GREBLICKI

### NON-PARAMETRIC IDENTIFICATION OF CONTROL SYSTEMS

Summary

The paper presents non-parametric algorithms for estimation of a regression function. They are applied for identification of control systems. Basic properties of these algorithms, such as convergence and speed of convergence, are determined. Non-parametric algorithms to be used for identification of a certain class of nonlinear dynamic systems are also examined. Results concerning their convergence are given. The paper is aimed, first of all, at presenting non-parametric methods from the standpoint of their application for identification purposes. Results in this field obtained so far are discussed.