

## METODA ROZWINIĘĆ ORTOGONALNYCH W IDENTYFIKACJI SYSTEMÓW I ROZPOZNAWANIU OBRAZÓW

Włodzimierz Greblicki\*, Przemysław Śliwiński<sup>†</sup>

Politechnika Wroclawska, Instytut Cybernetyki Technicznej

Wybrzeże Wyspiańskiego 27, 50-370 Wrocław, email: \*wgre(†slk)@ict.pwr.wroc.pl

*Streszczenie* - W pracy zastosowano metodę estymacji funkcji regresji opartą o rozwinięcie w szereg ortogonalny w algorytmach identyfikacji systemów i rozpoznawania obrazów. Wyznaczono warunki zbieżności algorytmów w odniesieniu do liczby pomiarów (w identyfikacji) lub par uczących (w rozpoznawaniu) i zastosowanego szeregu ortogonalnego.

### I. WSTĘP

Metoda rozwinięć ortogonalnych jest skutecznym narzędziem zarówno w uczeniu rozpoznawania jak i identyfikacji systemów, zob. m.in. monografię Devroye'a, Cyórfiego i Lugosiego [4] oraz prace Greblickiego i Pawłaka [7]-[9] (z dziedziny rozpoznawania) oraz prace Greblickiego [6], Greblickiego i Pawłaka [10]-[12], Krzyżaka [18], Hasiewiczza [14]-[16], Pawłaka [19], Pawłaka i Hasiewiczza [20] i Śliwińskiego [23] (z identyfikacji). W obydwu problemach pojawia się bowiem zadanie estymacji funkcji regresji, które można skutecznie rozwiązać stosując właśnie metodę rozwinięć ortogonalnych. Jej istotą jest umiejętność rozwinięcia odpowiedniej funkcji regresji w nieskończony szereg ortogonalny. Jako rodzinę takich funkcji ortogonalnych przyjmujemy funkcje trygonometryczne, Legendre'a, Laguerre'a, Hermite'a oraz falki Daubechies.

### II. PREZENTACJA METODY

**Ortogonalny estymator funkcji regresji.** Przedstawimy teraz jak estymuje się funkcję regresji przez rozwinięcie w szereg ortogonalny. Załóżmy w tym celu, że  $(X_1, Y_1), \dots, (X_n, Y_n)$  są niezależnymi obserwacjami pary zmiennych losowych  $(X, Y)$ . Na ich podstawie chcemy estymować funkcję regresji

$$\mu(x) = E\{Y|X=x\}.$$

Posłużymy się przy tym układem  $\varphi_k(x)$ ,  $k = 0, 1, 2, \dots$ , funkcji ortonormalnych na ustalonym zbiorze  $\Delta$ , tzn. takich, że  $\int_{\Delta} \varphi_i^2(x) dx = 1$  oraz  $\int_{\Delta} \varphi_i(x)\varphi_j(x) dx = 0$  dla  $i \neq j$ .

Jest oczywiste, że

$$\mu(x) = \frac{g(x)}{f(x)},$$

gdzie  $g(x) = \mu(x)f(x)$ . Przez  $f$  oznaczyliśmy gęstość prawdopodobieństwa zmiennej losowej  $x$ . Rozwijając  $g$  oraz  $f$  w przyjęty szereg ortogonalny, otrzymujemy

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

gdzie

$$a_k = E\{Y_n \varphi_k(X_n)\} \quad \text{oraz} \quad b_k = E\{\varphi_k(X_n)\}.$$

Jest rzeczą oczywistą, że można oczekiwać, że dla poprawnie wybranych układów funkcji ortogonalnych

$$\lim_{q \rightarrow \infty} \sum_{k=0}^q a_k \varphi_k(x) = g(x) = \mu(x)f(x)$$

oraz

$$\lim_{q \rightarrow \infty} \sum_{k=0}^q b_k \varphi_k(x) = f(x).$$

W sytuacji takiej

$$\lim_{q \rightarrow \infty} \frac{\sum_{k=0}^q a_k \varphi_k(x)}{\sum_{k=0}^q b_k \varphi_k(x)} = \mu(x).$$

Z uwagi na to, jako estymator  $\hat{\mu}(x)$  funkcji regresji przyjmujemy teraz

$$\hat{\mu}(x) = \frac{\sum_{k=0}^{q(n)} \hat{a}_k \varphi_k(x)}{\sum_{k=0}^{q(n)} \hat{b}_k \varphi_k(x)}, \quad (1)$$

gdzie

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

są estymatorami współczynników  $a_k$  oraz  $b_k$ . W estymatorze tym  $q(n)$  jest odpowiednio wybranym ciągiem liczb naturalnych.

**Zbieżność.** Aby wykazać zbieżność estymatora zauważmy, że

$$E \sum_{k=0}^{q(n)} \hat{a}_k \varphi_k(x) = \sum_{k=0}^{q(n)} a_k \varphi_k(x).$$

co wynika z tego, że  $E \hat{a}_k = a_k$ . Jeśli zatem

$$\lim_{n \rightarrow \infty} q(n) = \infty, \quad (2)$$

to

$$\lim_{n \rightarrow \infty} \sum_{k=0}^{q(n)} a_k \varphi_k(x) = g(x)$$

w każdym punkcie  $x \in \Delta$ , w którym

$$\lim_{n \rightarrow \infty} \sum_{k=0}^n a_k \varphi_k(x) = g(x). \quad (3)$$

Podobnie

$$\lim_{n \rightarrow \infty} \sum_{k=0}^{q(n)} b_k \varphi_k(x) = f(x)$$

w każdym punkcie  $x \in \Delta$ , w którym

$$\lim_{n \rightarrow \infty} \sum_{k=0}^n b_k \varphi_k(x) = f(x). \quad (4)$$

W analizie wariancji założymy, że poszczególne funkcje układu ortonormalnego są ograniczone w sposób określony przez nierówności jak poniżej:

$$|\varphi_k(x)| \leq c_k, \text{ dla wszystkich } x \in \Delta. \quad (5)$$

Zacznijmy od tego, że

$$\begin{aligned} \text{var} \left[ \sum_{k=0}^{q(n)} \hat{a}_k \varphi_k(x) \right] &= \frac{1}{n^2} \text{var} \left[ \sum_{i=1}^n \sum_{k=0}^{q(n)} Y_i \varphi_k(X_i) \varphi_k(x) \right] \\ &= \frac{1}{n} \text{var} \left[ Y \sum_{k=0}^{q(n)} \varphi_k(X) \varphi_k(x) \right] \\ &\leq \frac{1}{n} E \left\{ Y \sum_{k=0}^{q(n)} \varphi_k(X) \varphi_k(x) \right\}^2, \end{aligned}$$

Ponieważ

$$\left| \sum_{k=0}^{q(n)} \varphi_k(X) \varphi_k(x) \right| \leq \sum_{k=0}^{q(n)} c_k^2,$$

zatem badana wariancja jest ograniczona przez  $n^{-1} \left[ \sum_{k=0}^{q(n)} c_k^2 \right]^2 EY^2$ . Jeśli zatem

$$\lim_{n \rightarrow \infty} \frac{1}{n} \left[ \sum_{k=0}^{q(n)} c_k^2 \right]^2 = 0, \quad (6)$$

to

$$\lim_{n \rightarrow \infty} \text{var} \left[ \sum_{k=0}^{q(n)} \hat{a}_k \varphi_k(x) \right] = 0$$

i z podobnych powodów

$$\lim_{n \rightarrow \infty} \text{var} \left[ \sum_{k=0}^{q(n)} \hat{b}_k \varphi_k(x) \right] = 0.$$

Wykazaliśmy więc następujący lemat:

**Lemat 1** Jeśli ciąg  $\{q(n)\}$  spełnia warunki (2) oraz (6), to

$$\hat{\mu}(x) \xrightarrow{n \rightarrow \infty} \mu(x) \text{ w/g prawdopodobieństwa}$$

w każdym punkcie  $x \in \Delta$ , w którym zachodzą zbieżności (4) oraz (3).

Jest przy tym oczywiste, że – z uwagi na warunek (6) – wybór ciągu  $q(n)$  zależy od ustalonego układu funkcji ortogonalnych  $\varphi_k(x)$ . Z kolei, od funkcji ortogonalnych zależy to, w których punktach  $x$  zachodzą zbieżności (4) oraz (3). Poniżej podamy różne typy funkcji ortogonalnych i przedstawimy ich istotne dla nas własności.

### III. PRZYKŁADY ESTYMATORÓW

**Szereg trygonometryczny.** Załóżmy, że  $\Delta = (-\pi, \pi)$  oraz  $\varphi_k(x) = \phi_k(x)$ , gdzie

$$\phi_k(x) = e^{ikx}, k = 0, \pm 1, \pm 2, \dots, \quad (7)$$

co oznacza, że stosujemy zespoloną postać szeregu trygonometrycznego. Jest oczywiste, że  $\max_{-\pi \leq x \leq \pi} |\phi_k(x)| \leq 1$ , skąd wynika, że  $c_k = 1$  oraz  $n^{-1} \left[ \sum_{k=0}^{q(n)} c_k^2 \right]^2 = q^2(n)/n$ . Jak wiadomo, zbieżność

$$\lim_{n \rightarrow \infty} \sum_{|k|=0}^n a_k \phi_k(x) = g(x) = \mu(x) f(x).$$

a także

$$\lim_{n \rightarrow \infty} \sum_{|k|=0}^n b_k \phi_k(x) = f(x).$$

zachodzi w każdym punkcie  $x \in \Delta$ , w którym  $f$  oraz  $\mu$  są różniczkowalne, patrz [21]. Zatem lemat 1 przyjmuje teraz następującą postać:

**Lemat 2** Jeśli

$$\lim_{n \rightarrow \infty} q(n) = \infty \text{ oraz } \lim_{n \rightarrow \infty} \frac{q^2(n)}{n} = 0,$$

to, dla estymatora funkcji regresji (1) wykorzystującego zespolony szereg trygonometryczny (7),

$$\hat{\mu}(x) \xrightarrow{n \rightarrow \infty} \mu(x) \text{ w/g prawdopodobieństwa}$$

w każdym punkcie  $x \in (-\pi, \pi)$ , w którym gęstość  $f$  oraz regresja  $\mu$  są różniczkowalne oraz  $f(x) > 0$ .

**Szereg Legendre'a.** Załóżmy, że  $\Delta = (-1, 1)$  oraz  $\varphi_k(x) = p_k(x)$ , gdzie

$$p_k(x) = (\sqrt{2k+1}/\sqrt{2})P_k(x), \quad (8)$$

przy czym

$$P_k(x) = (1/2^k k!) \frac{d^k}{dx^k} (x^2 - 1)^2$$

jest  $k$ -tym wielomianem Legendre'a. Jak można sprawdzić  $P_0(x) = 1$ ,  $P_1(x) = x$ ,  $P_2(x) = (3/2)x^2 - 1/2$  itd. oraz  $\max_{-1 \leq x \leq 1} |p_k(x)| \leq 1$ , co oznacza, że  $c_k = 1$ . Zatem  $n^{-1} \left[ \sum_{k=0}^{q(n)} c_k^2 \right]^2 = q^2(n)/n$ . Ponadto zbieżności (3) oraz (4) zachodzą w każdym punkcie  $x \in (-1, 1)$  w którym  $f$  oraz  $\mu$  są różniczkowalne, patrz [21].

Z lematu 1 wynika teraz, że:

**Lemat 3** *Jeśli*

$$\lim_{n \rightarrow \infty} q(n) = \infty \text{ oraz } \lim_{n \rightarrow \infty} \frac{q^2(n)}{n} = 0,$$

to, dla estymatora funkcji regresji wykorzystującego szereg Legendre'a (8),

$$\hat{\mu}(x) \xrightarrow{n \rightarrow \infty} \mu(x) \text{ w/g prawdopodobieństwa}$$

w każdym punkcie  $x \in (-1, 1)$ , w którym gęstość  $f$  oraz regresja  $\mu$  są różniczkowalne oraz  $f(x) > 0$ .

**Szereg Laguerre'a.** Teraz  $\Delta = (0, \infty)$  oraz  $\varphi_k(x) = l_k(x)$ , gdzie

$$l_k(x) = e^{-x/2} L_k(x), \quad (9)$$

i gdzie

$$L_k^+(x) = (1/k!) e^x \frac{d^k}{dx^k} (x^k e^{-x})$$

jest  $k$ -tym wielomianem Laguerre'a. Nietrudno sprawdzić, że  $L_0(x) = 1$ ,  $L_1(x) = -x + 1$ ,  $L_2(x) = x^2/2 - 2x + 1$ , ... . Ponieważ, patrz [22],  $\max_{0 < x < \infty} |l_k(x)| \leq ck$ , zatem  $n^{-1} \left[ \sum_{k=0}^{q(n)} c_k^2 \right]^2 = q^6(n)/n$ . Ponadto, zbieżności (3) oraz (4) zachodzą w każdym punkcie  $x \in (0, \infty)$  w którym  $f$  oraz  $\mu$  są różniczkowalne, patrz [22].

Korzystając teraz z lematu 1 otrzymujemy następujący:

**Lemat 4** *Jeśli*

$$\lim_{n \rightarrow \infty} q(n) = \infty \text{ oraz } \lim_{n \rightarrow \infty} \frac{q^6(n)}{n} = 0,$$

to, dla estymatora funkcji regresji wykorzystującego szereg Laguerre'a (9),

$$\hat{\mu}(x) \xrightarrow{n \rightarrow \infty} \mu(x) \text{ w/g prawdopodobieństwa}$$

w każdym punkcie  $x \in (0, \infty)$ , w którym gęstość  $f$  oraz regresja  $\mu$  są różniczkowalne oraz  $f(x) > 0$ .

**Szereg Hermite'a.** Niech teraz  $\Delta = (-\infty, \infty)$  oraz  $\varphi_k(x) = h_k(x)$ , gdzie

$$h_k(x) = (1/\sqrt{2^k k! \sqrt{\pi}}) H_k(x), \quad (10)$$

przy czym

$$H_k(x) = (-1)^k e^{x^2} \frac{d^k}{dx^k} e^{-x^2}$$

jest  $k$ -tym wielomianem Hermite'a. Jak można sprawdzić,  $H_0(x) = 1$ ,  $H_1(x) = -2x$ ,  $H_2(x) = 4x^2 - 2$ , ... . Ponieważ  $\max_{-\infty < x < \infty} |h_k(x)| \leq ck^{-1/2}$ , patrz [22], zatem  $n^{-1} \left[ \sum_{k=0}^{q(n)} c_k^2 \right]^2 = q^{5/3}(n)/n$ . Ponadto, zbieżności (3) oraz (4) zachodzą w każdym punkcie  $x \in (-\infty, \infty)$  w którym  $f$  oraz  $\mu$  są różniczkowalne.

Dzięki lematowi 1 otrzymujemy następujący:

**Lemat 5** *Jeśli*

$$\lim_{n \rightarrow \infty} q(n) = \infty \text{ oraz } \lim_{n \rightarrow \infty} \frac{q^{5/3}(n)}{n} = 0, \quad (11)$$

to, dla estymatora funkcji regresji wykorzystującego szereg Hermite'a (10),

$$\hat{\mu}(x) \xrightarrow{n \rightarrow \infty} \mu(x) \text{ w/g prawdopodobieństwa}$$

w każdym punkcie  $x \in (-\infty, \infty)$ , w którym gęstość  $f$  oraz regresja  $\mu$  są różniczkowalne oraz  $f(x) > 0$ .

**Szeregi falkowe Daubechies.** Załóżmy obecnie, że  $\Delta = (-\infty, \infty)$  oraz przyjmijmy, że

$$\begin{aligned} \varphi_0(x) &= \sum_{n=n_{\min}(x;p)}^{n_{\max}(x;p)} D_{0,n}^p(x) \\ \varphi_k(x) &= \sum_{l=l_{\min}(x;k,p)}^{l_{\max}(x;k,p)} W_{kl}^p(x). \end{aligned} \quad (12)$$

dla  $k = 1, 2, \dots$ , przy czym funkcje

$$\begin{aligned} D_{0,n}^p(x) &= D^p(x-n) \\ W_{kl}^p(x) &= 2^{k/2} W^p(2^k x - l) \end{aligned}$$

są skalowanymi przesunięciami tzw. falki-ojca (nazywanej też funkcją skalującą)  $D^p(x)$  oraz falki-matki  $W^p(x)$  o numerze (indeksie) falkowym  $p$ , (zob. np. [3, p. 194], [25]). Indeksy  $n$  i  $l$  przebiegają, odpowiednio, od  $n_{\min}(x;p) = \lfloor x \rfloor - 2p + 2$  do  $n_{\max}(x;p) = \lfloor x \rfloor$  oraz od  $l_{\min}(x;k,p) = \lfloor 2^k x \rfloor - p + 1$  do  $l_{\max}(x;k,p) = \lfloor 2^k x \rfloor + p - 1$ .

**Uwaga 1** W algorytmach falkowych można też zastosować alternatywną postać rozwinięcia falkowego

$$\varphi_k(x) = \sum_{l=l_{\min}(x;k,p)}^{l_{\max}(x;k,p)} W_{kl}^p(x), k=0, \pm 1, \pm 2, \dots \quad (13)$$

Dla obu rozwinięć zachodzi nierówność  $\max_{-\infty < x < \infty} |\varphi_k(x)| \leq c2^{k/2}$  (por. (5)) skąd wynika, że  $n^{-1} \left[ \sum_{k=0}^{q(n)} c_k^2 \right]^2 = 4 \cdot 2^{q(n)}/n$ . A zatem na podstawie lematu 1 otrzymujemy następujący:

**Lemat 6** Jeśli

$$\lim_{n \rightarrow \infty} q(n) = \infty \text{ oraz } \lim_{n \rightarrow \infty} \frac{2^{q(n)}}{n} = 0,$$

to, dla estymatora funkcji regresji wykorzystującego szeregi falkowe Daubechies (12) albo (13),

$$\hat{\mu}(x) \xrightarrow{n \rightarrow \infty} \mu(x) \text{ w/g prawdopodobieństwa}$$

dla każdego  $p$ , w każdym punkcie  $x \in (-\infty, \infty)$ , w którym gęstość  $f$  oraz regresja  $\mu$  są ciągłe oraz  $f(x) > 0$  [17].

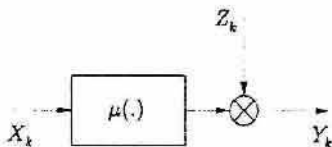
#### IV. IDENTYFIKACJA SYSTEMÓW

Pokażemy teraz w jaki sposób przedstawioną powyżej metodę rozwinięć ortogonalnych można zastosować do identyfikacji charakterystyki nieliniowej.

**System statyczny.** Stacyjny system pokazany na rys. 1, ma charakterystykę  $\mu$ . Pobudzany jest on stacjonarnym, nieskorelowanym sygnałem losowym. Wyjście jest zakłócone addytywnym białym szumem o zerowej średniej. Zadanie identyfikacji polega na estymacji nieznannej charakterystyki  $\mu$ . Ponieważ

$$E\{Y_n | X_n = x\} = \mu(x),$$

a zatem jest ono równoważne estymacji funkcji regresji na podstawie pomiarów  $(X_1, Y_1), \dots, (X_n, Y_n)$ .



Rys. 1. System statyczny.

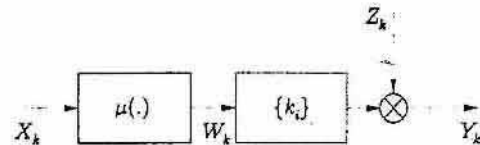
Jeśli sygnał wejściowy jest ograniczony i np.  $-1 \leq X \leq 1$ , to w celu estymacji charakterystyki  $\mu$  można

zastosować rozwinięcie w szereg Legendre'a, co prowadzi do poniższego algorytmu:

$$\hat{\mu}(x) = \frac{\sum_{k=0}^{q(n)} \hat{a}_k p_k(x)}{\sum_{k=0}^{q(n)} \hat{b}_k p_k(x)} \quad (14)$$

gdzie  $\hat{a}_k = n^{-1} \sum_{i=1}^n Y_i p_k(X_i)$  oraz  $\hat{b}_k = n^{-1} \sum_{i=1}^n p_k(X_i)$ ,

**System Hammersteina.** Bardziej skomplikowane jest wykrywanie nieliniowej charakterystyki w systemie Hammersteina, patrz rys. 2, na podstawie obserwacji  $(X_1, Y_1), \dots, (X_n, Y_n)$ . Sygnał  $W_n$  łączący podsystem nieliniowy o charakterystyce  $\mu$  i liniowy, dynamiczny o odpowiedzi skokowej  $\{k_i\}$  nie jest bowiem dostępny pomiarowi.



Rys. 2. System Hammersteina.

Zakładamy przy tym, że sygnał pobudzający system,  $\{X_n; n = \dots, -1, 0, 1, \dots\}$ , jest stacjonarnym, białym szumem. Podobny charakter ma nieskorelowane z nim zakłócenie  $Z_n$  o zerowej średniej. Jak można sprawdzić  $E\{Y_n | X_n = \hat{x}\} = \alpha\mu(x) + \beta$ , gdzie  $\alpha = k_0$ ,  $\beta = E\{\mu(X)\} \sum_{i=1}^{\infty} k_i$ . Zakładając dla prostoty, że  $E\{m(X_0)\} = 0$ , otrzymujemy  $E\{Y_n | X_n = x\} = \alpha\mu(x)$ . Oznacza to, że identyfikacja podsystemu nieliniowego jest równoznaczna estymacji funkcji regresji. To, że charakterystykę można wykryć jedynie z dokładnością do pewnej stałej  $\alpha$  jest oczywistym skutkiem szeregowej struktury systemu. W dalszych rozważaniach, aby nie komplikować oznaczeń przyjmujemy  $k_0 = 1$ , co oznacza, że

$$E\{Y_n | X_n = x\} = \mu(x).$$

Założymy, podobnie jak w przypadku systemu statycznego, że sygnał wejściowy jest ograniczony i  $-1 \leq U \leq 1$ . W celu estymacji charakterystyki  $\mu$  zastosujemy ponownie algorytm (14). Ze względu na to, że teraz obserwacje  $Y_i$  oraz  $Y_j$  są zależne, analiza wariancji jest nieco trudniejsza. Warunek zbieżności algorytmu jednak nie ulega zmianie, patrz np. [6]

## V. ROZPOZNAWANIE

Pokażemy teraz, że zadanie uczenia rozpoznawania polega na właściwej estymacji funkcji regresji. Niech zatem  $(\theta, X)$  będzie parą zmiennych losowych. Pierwsza z nich przyjmuje wartości w zbiorze  $\{1, 2, \dots, m\}$ , elementy którego nazywają się klasami, a druga na prostej  $R$ . Rozkład pary  $(\theta, X)$  opisuje się przez prawdopodobieństwa  $p_i = P\{\theta = i\}$  poszczególnych klas oraz warunkowe gęstości prawdopodobieństwa  $f_i$  zmiennej  $X$  w tych klasach,  $i = 1, 2, \dots, m$ . Rozpoznawanie polega na estymacji  $\theta$  na podstawie  $X$ .

Regułę rozpoznawania definiuje się jako funkcję, która każdemu punktowi  $x \in R$  przyporządkowuje element ze zbioru  $M$ . Jakością reguły  $\psi$  jest  $P\{\psi(X) \neq \theta\}$ , czyli prawdopodobieństwo błędnego rozpoznania. Najlepszą regułę oznaczymy przez  $\psi^*$ , a jej jakość przez  $R^*$ . Jak wiadomo,  $\psi^*(x) = \arg \max_{i \in M} P\{\theta = i | X = x\}$ . Zauważając, że

$$P\{\theta = i | X = x\} = E\{I_{\{\theta=i\}} | X = x\},$$

gdzie  $I_{\{\theta=i\}} = 1$  jeśli  $\theta = i$  oraz  $I_{\{\theta=i\}} = 0$  jeśli  $\theta \neq i$ , otrzymujemy bardziej użyteczną postać reguły optymalnej

$$\psi^*(x) = \arg \max_{i=1, \dots, m} \mu_i(x), \quad (1)$$

gdzie  $\mu_i(x) = E\{I_{\{\theta=i\}} | X = x\}$  jest funkcją regresji.

Zadanie uczenia pojawia się, gdy prawdopodobieństwa klas lub rozkłady w klasach nie są znane, a brak tej wstępnej informacji jest rekompensowany przez ciąg uczący  $V_n = \{(\theta_1, X_1), (\theta_2, X_2), \dots, (\theta_n, X_n)\}$ , tzn. ciąg niezależnych obserwacji pary  $(\theta, X)$ , czyli ciąg  $n$  prawidłowo rozpoznanych obserwacji  $X_1, X_2, \dots, X_n$ .

Istotą naszego podejścia jest to, że na podstawie ciągu uczącego estymujemy funkcje regresji  $\mu_1(x), \dots, \mu_m(x)$ . Estymatory te oznaczmy jako  $\hat{\mu}_1(x), \dots, \hat{\mu}_m(x)$ . Zatem

$$\hat{\mu}_i(x) = \frac{\sum_{k=0}^{q(n)} \hat{a}_{ik} \varphi_k(x)}{\sum_{k=0}^{q(n)} \hat{b}_k \varphi_k(x)},$$

gdzie

$$\hat{a}_{ik} = \frac{1}{n} \sum_{j=0}^n I_{\{\theta_j=i\}} \varphi_k(X_j) \quad \text{oraz} \quad \hat{b}_k = \frac{1}{n} \sum_{j=0}^n \varphi_k(X_j)$$

Jako algorytm uczenia przyjmujemy regułę, która  $x$  zalicza do klasy określonej wzorem  $\arg \max_{i \in M} \hat{\mu}_i(x)$ , czyli dla której

$$\arg \max_{i=1, \dots, m} \sum_{j=0}^n \sum_{k=0}^{q(n)} I_{\{\theta_j=i\}} \varphi_k(X_j) \varphi_k(x)$$

Jak wiadomo, patrz np. [4], jeśli  $\hat{\mu}_i(x) \rightarrow \mu_i(x)$  w/g prawdopodobieństwa, gdy  $n \rightarrow \infty$ , to

$$P\{\text{klasyfikacja jest poprawna} | V_n\} \xrightarrow{n \rightarrow \infty} 1 \quad (15)$$

w/g prawdopodobieństwa. Mówimy, że algorytm uczenia jest asymptotycznie optymalny.

Jeśli zatem zastosujemy estymator wykorzystujący szereg Hermite'a, to nasza reguła uczenia zalicza  $x \in (-\infty, \infty)$  do klasy, dla której

$$\arg \max_{i=1, \dots, m} \sum_{j=0}^n \sum_{k=0}^{q(n)} I_{\{\theta_j=i\}} h_k(X_j) h_k(x).$$

Jest oczywiste, że jeśli wszystkie gęstości  $f_1, \dots, f_m$  są ciągle, to z lematu 5 wynika, że jeśli ciąg  $\{q(n)\}$  spełnia warunek (11), to nasz algorytm uczenia rozpoznawania jest asymptotycznie optymalny.

## VI. ZAKOŃCZENIE

W pracy pokazano, że metoda estymacji funkcji regresji w oparciu o rozwinięcie w szereg ortogonalny może być zastosowana w identyfikacji systemów, np. do identyfikacji nieliniowych charakterystyk systemów statycznych i systemów Hammersteina, a także w zadaniu uczenia rozpoznawania.

Zastosowanie szeregów ortogonalnych pozwala algorytmom zbiegać do optymalnych rozwiązań przy bardzo małej, nieparametrycznej, wiedzy wstępnej o identyfikowanej nieliniowości i funkcji gęstości wejścia systemu (w przypadku zadania identyfikacji) oraz o funkcjach gęstości prawdopodobieństwa cech w klasach (w zadaniu rozpoznawania), co stanowi istotny wyróżnik tych algorytmów (por. [1]-[2], [4], [13], [24], [26]). Zbieżności poszczególnych algorytmów zachodzą w tych punktach, w których ma miejsce zbieżność zastosowanych w nich szeregach ortogonalnych.

## Bibliografia

- [1] Billings S., Fakhouri S., Non-linear system identification using the Hammerstein model. *International Journal of Systems Science*, 10(5):567-578, 1979.
- [2] Chang F. and Luus R., A noniterative method for identification using Hammerstein model. *IEEE Transactions on Automatic Control*, pages 465-468, October 1971.

- [3] Daubechies I., *Ten Lectures on Wavelets*. SIAM, Philadelphia, 1992.
- [4] Devroye, L., Györfi, L., Lugosi, G., *A Probabilistic Theory of Pattern Recognition*, Springer, New York, 1996.
- [5] Gallman P., A comparison of two Hammerstein model identification algorithms. *IEEE Transactions on Automatic Control*, AC-21:125-127, February 1976.
- [6] Greblicki W., Nonparametric orthogonal series identification of Hammerstein systems. *International Journal of Systems Science*, 20(12):2355-2367, 1989.
- [7] Greblicki W., Pawlak M., Classification using the Fourier series estimate of multivariate density functions, *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 11, No. 10, pp. 726-730, October 1981.
- [8] Greblicki W., Pawlak M., A classification procedure using the multiple Fourier series, *Information Sciences*, vol. 26, pp. 115-126, 1982.
- [9] Greblicki W., Pawlak M., Almost sure convergence of classification procedures using Hermite series density estimates, *Pattern Recognition Letters*, vol. 2, pp. 13-17, 1983.
- [10] Greblicki W., Pawlak M., Nonparametric identification of a cascade nonlinear time series system. *Signal Processing*, 22:61-75, 1991.
- [11] Greblicki W., Pawlak M., Nonparametric identification of a particular nonlinear time series system. *IEEE Transactions on Signal Processing*, 40(4):985-989, April 1992.
- [12] Greblicki W., Pawlak M., Nonparametric recovering nonlinearities in block-oriented systems with the help of Laguerre polynomials. *Control Theory and Advanced Technology*, 10(4, Part A):771-791, 1994.
- [13] Haist N., Chang F., Luus R., Nonlinear identification in the presence of correlated noise using a Hammerstein model. *IEEE Transactions on Automatic Control*, AC-15:553-555, October 1973.
- [14] Hasiewicz Z., Hammerstein system identification by the Haar multiresolution approximation. *International Journal of Adaptive Control and Signal Processing*, 13:697-717, 1999.
- [15] Hasiewicz Z., Modular neural networks for nonlinearity recovering by the Haar approximation, *Neural Networks*, 13:1107-1133, 2000.
- [16] Hasiewicz Z., Non-parametric estimation of nonlinearity in a cascade time series system by multiscale approximation, *Signal Processing*, 81:791-807, 2001.
- [17] Kelly S., Kon M., Raphael M., Pointwise convergence of wavelet expansions, *Bulletin of The American Mathematical Society*, 30:87-94, 1994.
- [18] Krzyżak A., On estimation of a class of nonlinear systems by the kernel regression estimate. *IEEE Transactions on Information Theory*, 36(1):141-152, February 1990.
- [19] Pawlak M., On the series expansion approach to the identification of Hammerstein system. *IEEE Transactions on Automatic Control*, 36:763-767, 1991.
- [20] Pawlak M., Hasiewicz Z., Nonlinear system identification by the Haar multiresolution analysis. *IEEE Transactions on Circuits and Systems, I-Fundamental Theory and Applications*, 45(9):945-961, September 1998.
- [21] Sansone G., *Orthogonal Functions*. Interscience, New York, 1959.
- [22] Szegő G., *Orthogonal Polynomials*. American Mathematical Society, Providence, R.I., third edition, 1974.
- [23] Śliwiński P., *Algorytmy identyfikacji systemów nieliniowych za pomocą fulek (rozprawa doktorska)*, Raport Instytutu Cybernetyki Technicznej, PRE 82/2000, Politechnika Wroclawska, Wrocław, 2000.
- [24] Thathachar M., Ramaswamy S., Identification of a class of non-linear systems. *International Journal of Control*, 18(4):741-752, 1973.
- [25] Wojtaszczyk P., *Teoria fulek, Państwowe Wydawnictwa Naukowe*, Warszawa, 2000.
- [26] Zi-Qiang L., A nonparametric polynomial identification algorithm for the Hammerstein system. *IEEE Transactions on Automatic Control*, 42(10):1435-1441, October 1997.